

Univerzita Karlova v Praze
Fakulta sociálních věd

DISERTAČNÍ PRÁCE

2017

Johana Chylíková

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut sociologických studií

Analytické metody odhadu chyb měření v datech z výběrových
šetření

vypracovala: Mgr. Johana Chylíková
rok: 2017

školitel: PhDr. Jiří Vinopal, Ph.D.

Čestně prohlašuji, že práce byla vypracována samostatně a s použitím uvedených pramenů a literatury.

V Praze dne: 5. 3. 2017

Johana Chylíková

Poděkování:

Mgr. Jindřichovi Krejčímu, Ph.D.

Lukášovi a rodině za podporu

Abstrakt:

Dizertační práce se zaměřuje na oblast výzkumu chyb měření v datech ze sociálněvědních výběrových šetření. Chyby měření definuje v analyticko – teoretickém rámci, který vyplývá z analytické metody modelování strukturálních rovnic (SEM) a Klasické testové teorie (KTT), kterou obohacuje o složku systematické chyby měření. Tato práce má dva cíle, kterými se snaží přispět k rozvoji a rozšíření poznatků české sociálněvědní metodologie. Prvním cílem je představit a ilustrovat v Česku dosud nerozšířené metody pro odhad reliability a systematické chyby měření a upozornit na některé jejich problematické aspekty. Druhým cílem je s využitím těchto metod přinést originální poznatky o kvalitě dat z českých výběrových šetření.

Práce obsahuje tři empirické studie, z nichž každá používá jednu z metod odhadu chyby měření definovanou v rámci používaného teoreticko – analytického přístupu. První studie prezentuje analýzu reliability měření s využitím kvazisimplexového modelu (QSM), ilustruje jeho použití a přináší optimistické údaje o reliabilitě českých dat z panelového šetření EU SILC. Ve druhé studii je použit model konfirmativní faktorové analýzy, pracovně nazývaný CF-RS model, s jehož využitím lze v datech identifikovat systematickou chybu, jež vznikla v důsledku tzv. souhlasného stylu odpovídání.

Z výsledků analýzy českých dat ISSP vyplynulo, že český souhlasný styl odpovídání se zřejmě liší od souhlasného stylu odpovídání nalezeného v datech v zemích západní Evropy. CF-RS model v českých datech totiž nepotvrdil teoreticky očekávanou korelaci souhlasného stylu odpovídání s věkem a se vzděláním, která byla doložena v datech ve vybraných západoevropských zemích. Absence těchto korelací je v této práci interpretována jako české kulturní specifikum. Šestá kapitola se zaměřuje na problematiku experimentálního designu 2 split – ballot MTMM a analýzu těchto dat v tzv. True Score MTMM modelu. Text ukazuje, jak problematické je získat z těchto dat odhady reliability a validity, resp. systematické chyby měření vzniklé v důsledku použité metody měření, a vysvětluje, proč tyto problémy v analýze vznikají. Zájemce o analýzu 2 SB MTMM dat pak odkazuje na výsledky výpočtů vědeckého týmu, který připravuje 2 SB MTMM experimenty pro šetření European Social Survey. V následující kapitole je ilustrováno, jak lze využít údaje o reliabilitě a validitě dat v substanciální analýze. Práce povzbuzuje kvantitativně zaměřené sociology a socioložky, aby se zajímali o reliabilitu a validitu měření v datech, která používají, a snažili se v analýze používat položky o známé a relativně vysoké kvalitě. Závěrečná kapitola dizertace je věnována úvaze nad budoucností

výběrových šetření v éře Velkých dat a internetu a zhodnocení současného stavu, kdy Velká data ještě zcela nepronikla do akademické výzkumné praxe. Text dále upozorňuje na chyby a zkreslení, která obsahují různé typy Velkých dat či data z internetových anket a opt – in panelů.

Klíčová slova: výběrové šetření, metodologie, chyba měření, validita, reliabilita, kvazisimplexový model, souhlasný styl odpovídání, multitrait – multimethod, Velká data

Obsah

| | |
|---|----|
| 1. Úvod..... | 10 |
| 2. Chyby v datech ze sociálněvědních výběrových šetření | 14 |
| 2.1 Druhy chyb v sociálněvědních výběrových šetřeních..... | 14 |
| 2.2 Chyby měření | 15 |
| 2.3 Psychometrie jako základ pro výzkum chyb měření..... | 16 |
| 2.4 Klasická testová teorie | 17 |
| 2.5 Typy měření | 18 |
| 2.6 Reliabilita | 19 |
| 2.7 Metody odhadu reliability | 20 |
| 2.7.1 Metody odhadu reliability používající opakovaná měření | 21 |
| 2.7.2 Metody odhadu reliability používající více indikátorů měřeného konceptu | 23 |
| 2.8 Validita..... | 24 |
| 2.8.1 Druhy validity, resp. metody určení či odhadu validity | 24 |
| 3. Teoreticko – analytický rámec dizertační práce..... | 28 |
| 3. 1 Rozšíření Klasické testové teorie o systematickou složku | 28 |
| 3. 2 Teoreticko – analytický rámec práce..... | 29 |
| 3. 3 Analytická metoda modelování strukturálních rovnic (SEM)..... | 30 |
| 3. 3. 1 Model v analýze kvantitativních dat | 30 |
| 3. 3. 2 Modelování strukturálních rovnic s latentními proměnnými | 31 |
| 3. 3. 3 Odhad parametrů v modelu | 31 |
| 3. 3. 4 Shoda modelu s daty | 33 |
| 3. 3. 5 Ukazatele shody modelu s daty a jejich doporučené hodnoty..... | 36 |
| 3. 3. 6 Specifikace SEM modelu | 37 |
| 3. 2. 7 Problematické aspekty analýzy SEM | 37 |
| 3. 3. 8 Software pro strukturální modelování..... | 38 |
| 4. Kvazisimplexový model pro odhad reliability měření: Odhad reliability vybraných položek z českých šetření EU SILC..... | 40 |
| 4. 1 Kvazisimplexový model (QSM) a srovnání s metodou test – retest | 41 |
| 4. 2 Použití kvazisimplexového modelu | 42 |
| 4. 3 Popis kvazisimplexového modelu | 44 |
| 4. 4 Kritika kvazisimplexového modelu | 45 |
| 4. 5 Empirická část: Odhad reliability českých dat EU SILC kvazisimplexovým modelem | 49 |
| 4. 5. 1 Zdroje českých dat k analýze kvazisimplexovým modelem | 49 |
| 4. 5. 2 Šetření European Union Statistics on Income and Living Conditions (EU SILC) | 49 |
| 4. 5. 3 Položky z dotazníků EU SILC vhodné ke QSM analýze | 51 |

| | |
|--|-----|
| 4. 5. 4 Výběr případů ke QSM analýze | 52 |
| 4. 5. 5 Analýza kvazisimplexovým modelem a odhady reliability | 54 |
| 4. 5. 6 Výsledky | 59 |
| 4. 6 Diskuze..... | 60 |
| 5. Modelování stylu odpovídání v modelu konfirmativní faktorové analýzy: Souhlasný styl odpovídání v měření postojů k přistěhovalcům | 66 |
| 5. 1 Styl odpovídání a druhy stylů odpovídání..... | 68 |
| 5. 2 Příčiny vzniku stylů odpovídání..... | 69 |
| 5. 3 Vztah ARS k proměnným na individuální úrovni a na úrovni zemí | 71 |
| 5. 4 Vliv odchylky RS na kvalitu dat | 73 |
| 5. 5 Modelování RS v modelu CFA..... | 75 |
| 5. 6 Další metody identifikace odchylky RS..... | 78 |
| 5. 7 Empirická část: Modelování ARS jako latentní proměnné v modelu konfirmativní faktorové analýzy | 80 |
| 5. 7. 1 Popis Modelu CF-RS | 80 |
| 5. 7. 2 Použitá data | 83 |
| 5. 7. 3 Výsledky analýzy | 85 |
| 5. 7. 4 Analýza dat z Belgie a Francie a srovnání s výsledky z České republiky..... | 90 |
| 5. 8 Diskuze..... | 92 |
| 6. True Score MTMM metoda odhadu validity a reliability a analýza českých 2 SB MTMM dat z šetření European Social Survey | 98 |
| 6. 1 Multitrait multimethod (MTMM) experiment | 99 |
| 6. 2 Design sběru MTMM dat..... | 101 |
| 6. 3 Design sběru 2 split – ballot MTMM..... | 101 |
| 6. 4 Modelování MTMM dat..... | 102 |
| 6. 4. 1 True Score MTMM Model..... | 103 |
| 6. 4. 2 Specifikace True Score MTMM modelu..... | 104 |
| 6. 4. 3 Odhady koeficientů validity a reliability z TS MTMM modelu | 105 |
| 6. 4. 4 Využití odhadů kvality z TS MTMM modelu..... | 106 |
| 6. 5 Problematické aspekty analýzy 2 SB MTMM dat | 107 |
| 6. 5. 1 Nekonvergence TS MTMM modelů na „malých“ souborech..... | 108 |
| 6. 5. 2 Řešení pro úspěšnou analýzu 2 SB MTMM dat..... | 109 |
| 6. 5. 3 Nemožnost získat odhady pro některé 2 SB MTMM experimenty..... | 110 |
| 6. 6 Empirická část: Aplikace TS MTMM modelu na česká data šetření ESS | 110 |
| 6. 6. 1 Data European Social Survey..... | 111 |
| 6. 6. 2 Výsledky analýzy | 111 |

| | |
|--|-----|
| 6. 7 Diskuze..... | 113 |
| 7. Využití odhadů chyb měření pro substanciální analýzu..... | 114 |
| 7. 1 Vliv snížené validity a reliability na substanciální analýzu | 115 |
| 7. 2 Očištění dat o chybový rozptyl..... | 118 |
| 7. 3 Program pro odhad kvality dotazníkových položek Survey Quality Prediction | 119 |
| 7. 3. 1 Popis programu SQP | 120 |
| 7. 4 Diskuze..... | 122 |
| 8. Velká data a internet jako zdroj dat pro výzkum: Exkurz do současnosti a budoucnosti kvantitativního sociálněvědního výzkumu..... | 126 |
| 8. 1 Co jsou Velká data?..... | 130 |
| 8. 2 Epistemologie Velkých dat | 134 |
| 8. 2. 1 Velká data v epistemologii sociálních věd..... | 136 |
| 8. 3 Velká data jako zdroj informací pro sociálněvědní výzkum | 137 |
| 8. 3. 1 Náročnost analýzy Velkých dat..... | 138 |
| 8. 3. 2 Dostupnost Velkých dat pro sociálněvědní výzkum | 139 |
| 8. 3. 3 Nereprezentativita Velkých dat..... | 141 |
| 8. 3. 4 Výběr jednotek z Velkých dat a odchylka výběru | 142 |
| 8. 3. 5 Množství proměnných ve Velkých datech | 144 |
| 8. 3. 6 Nejistota ohledně motivací k akci | 145 |
| 8. 3. 7 Vliv sociálního okolí a závislost pozorování | 146 |
| 8. 3. 8 Online vs. Offline život..... | 149 |
| 8. 4 Internetové ankety | 151 |
| 8. 4. 1 Internetové ankety a neanonymní výzkumník..... | 152 |
| 8. 5 Internetové opt – in panely | 153 |
| 8. 6 Společná budoucnost Velkých dat, internetu a výběrových šetření | 154 |
| 9. Resumé..... | 159 |
| 10. Bibliografie: | 161 |
| 11. Přílohy | 176 |
| 12. Použitá data | 181 |
| 13. Seznam zkratk: | 182 |

1. Úvod

Tato dizertační práce se zabývá chybami měření v datech ze sociálněvědních výběrových šetření a používá teoreticko – analytický rámec daný rozšířenou Klasickou testovou teorií (KTT) a analytickou metodou modelování strukturálních rovnic (SEM). V tomto rámci lze definovat chyby měření v datech z výběrových šetření a stanovit metody pro identifikaci těchto chyb. Aplikovaný teoreticko – analytický rámec používá známé koncepty¹ reliability a validity. Pojetí reliability plně přebírá z Klasické testové teorie a validitu chápe v poměrně užším smyslu, než je obvyklé. Toto chápání vyplývá z původního konceptu tzv. konvergentní a diskriminantní validity a jeho dalšího rozšíření; validita je tak míra, do které měřená proměnná odpovídá měřenému konstruktovi vyjádřenému v modelu strukturálních rovnic při současné kontrole pro systematickou chybu měření a náhodnou chybu měření. Princip metod pro odhad chyby měření, které jsou prezentovány v této práci, pak spočívá v rozdělení rozptylu měřené proměnné na část odpovídající měřenému konstruktovi, náhodné chybě měření a systematické chybě, vzniklé vlivem použité metody.

Tato dizertační práce má dva hlavní cíle, kterými se snaží přispět k rozvoji a rozšíření poznatků české sociálněvědní metodologie. Prvním cílem je ilustrovat v Česku dosud nerozšířené metody pro odhad reliability a systematické chyby měření a upozornit na některé jejich problematické aspekty. Druhým cílem je s využitím těchto metod přinést originální poznatky o kvalitě dat z českých výběrových šetření. Tyto poznatky je možné využít jako samostatné informace o kvalitě konkrétních dat nebo jako vodítko pro další metodologický výzkum.

Metody, které jsou předmětem výkladu a nástrojem analýz v této práci, nejsou náhodně vybrány. Kromě toho, že jsou všechny řešitelné v rámci rozšířené KTT a jejich použití vyžaduje aplikaci SEM přístupu, zastupuje každá z nich metodu pro zjišťování buď jednoho ze dvou aspektů kvality dat – validity či reliability – nebo obou najednou. První použitá metoda poskytuje číselný odhad reliability, druhá odhaduje vliv systematické

¹ V sociálněvědní literatuře jsou často používány termíny koncept a konstrukt, ovšem neexistuje úplná shoda o významu těchto termínů; v některých případech se ani neodlišují a používá se pouze termín koncept nebo pouze termín konstrukt. V tomto textu používám termín koncept pro teoretický, příp. analytický ustálený nástroj vědeckého výzkumu, mluvím tedy např. o konceptu validity, konceptu reliability atp. Termín konstrukt používám pro vědecké uchopení empiricky existujícího a přímo nepozorovatelného, tj. abstraktního společenského fenoménu, mluvím tedy např. o konstruktovi postojů k přistěhovalcům, konstruktovi postavení mužů a žen ve společnosti atp.

chyby a třetí dokáže odhadnout reliabilitu i validitu zároveň. Prezentované metody tak pokrývají celé spektrum výzkumu chyb měření v aplikovaném teoreticko – analytickém rámci.

Práce obsahuje tři empirické studie. První ze studií je prezentována ve čtvrté kapitole, která se věnuje popisu a aplikaci metody pro odhad reliability nazvané kvazisimplexový model (QSM). V následující kapitole je aplikován model konfirmativní faktorové analýzy, pracovníčně nazývaný CF-RS model, k identifikaci systematické chyby, jež vznikla v důsledku tzv. souhlasného stylu odpovídání. Šestá kapitola je zaměřena na problematiku experimentálního designu 2 split – ballot MTMM dat a analýzu těchto dat v tzv. True Score MTMM modelu. Sedmá kapitola prezentuje praktické využití odhadů reliability a validity v substanciální analýze a informuje čtenářky a čtenáře o existenci online programu Survey Quality Prediction (SQP). Závěr práce je věnován úvaze nad budoucností výběrových šetření v éře internetu a Velkých dat; přináší základní informace o využívání Velkých dat a dat pocházejících z internetových anket a opt – in panelů v sociálních vědách a upozorňuje na obtíže, kterým sociální věda při implementaci těchto zdrojů dat do výzkumné praxe čelí.

Díky zaměření na celý teoreticko – analytický rámec, nikoliv pouze na jednu metodu, tento text pokrývá široké spektrum současného výzkumu chyb měření. Analytičky a analytici v oboru metodologie výběrových šetření především na západ od České republiky se již od začátku 70. let minulého století věnují výzkumu reliability prostřednictvím QSM, s rozvojem počítačů a statistického softwaru se pak objevil zájem o odhady efektu použité metody měření z MTMM dat. V 90. letech formulovali výzkumníci inovovaný TS MTMM model, který dokáže z MTMM dat odhadnout koeficienty validity i reliability. Zhruba ve stejné době se metodologové začali zabývat modelováním souhlasného stylu odpovídání, přičemž zájem o tento analytický přístup znovu ožil s používáním analýzy invariance mezinárodních dat a v současné době tvoří mainstream v metodologické analýze dat z šetření.

Kromě zprostředkování informací o metodách odhadu chyb měření a prezentace konkrétních výsledků analýz je cílem tohoto textu i připoutání pozornosti k výzkumu chyb měření. V západní Evropě a v USA jsou chyby měření tématem, kterému se v metodologii věnuje poměrně hodně prostoru, výzkumníci v Česku se však analýzou chyb měření běžně nezabývají. Od devadesátých let do současnosti vzniklo pouze několik studií, zabývajících

se kvalitou měření či přímo odhadováním reliability či validity dotazníkových položek. Na konci devadesátých let publikoval Sociologický časopis sérii článků Jana Řeháka a spolupracovník (Řehák, 1998a, b; Řehák, Bártová, Hamanová, 1998), který informoval o vybraných metodách analýzy reliability, byť poměrně úspornou formou. Celkový zájem o kvalitu dat z výběrových šetření prezentoval výklad a analýza v publikaci Krejčího (2008), validitu z kvalitativního hlediska zkoumal Vinopal (2008). Konstruktová validita v modelu kauzálních vztahů MIMIC byla předmětem studie Kreidla (2005) a vybranými metodami odhadu reliability se ve své práci zabýval Schubert (2010). Některé z těchto prací přinesly informace o kvalitě konkrétních dat, jejich dalším a neméně důležitým účelem bylo ale i seznámení české odborné veřejnosti s možnostmi analýzy chyb měření. Tato práce se chce zařadit po bok těchto studií, má však ambici přinést najednou větší množství informací o kvalitě českých dat z výběrových šetření než uvedené publikace a srozumitelnou formou představit hned několik metod, které jsou větší částí české odborné veřejnosti dosud neznámé.

Další z motivací pro zpracování tématu, prezentovaného v této práci, je přesvědčení, že chyby měření si pozornost výzkumníků a výzkumnice v sociálních vědách zaslouží. V sedmé kapitole této práce ukazují, jak silně mohou náhodné a systematické chyby měření ovlivnit lineární vztahy mezi měřenými proměnnými, a upozorňují, že systematické chyby navíc ovlivňují i průměry proměnných. Skutečnost, že po implementaci hlediska chyb měření do analýzy se v některých případech mohou zásadním způsobem změnit výsledky analýzy, by neměla zůstat nepovšimnuta. Pokud nám v kvantitativní analýze různých společenských témat jde o více či méně přesné informace o celé zkoumané populaci, tak bychom měli rovněž uvažovat nad tím, jak mohou být závěry celé studie ovlivněny chybami měření a jaké skutečnosti z toho vyplývají. Přestože tato práce ilustruje použití metod k identifikaci chyb měření a poskytuje konkrétní číselné odhady validity a reliability, případně kvantifikuje vliv systematické chyby stylu odpovídání, není jejím cílem práce přesvědčit výzkumníky a výzkumnice, aby brali chyby měření v potaz v běžných analýzách substantiálních témat. Vzhledem k tomu, že tzv. čištění dat o chybové rozptyly není v současné době v analýzách dat z výběrových šetření mainstreamem, a že prestižní zahraniční sociálněvědní časopisy kontrolu pro chyby měření nevyžadují, není pro takovou analytickou praxi důvod. Nicméně zájem o kvalitu dat ve světové vědě stoupá a různé aspekty kvality kvantitativních sociálněvědních dat se stávají

středem zájmu stále častěji. K tomuto trendu se hlásí i moje dizertační práce, jejímž cílem je ukázat možnou alternativu chápání dat z výběrových šetření.

V analýzách v této práci jsou použity tři velké zdroje dat; ve studii reliability jsou to data ze všech vln českých šetření Životní podmínky (EU SILC), v analýze s využitím CF-RS modelu byla aplikována data ze dvou vln šetření International Social Survey Programme (ISSP), a k ilustraci metody TS MTMM v šesté kapitole posloužila experimentální data ze dvou vln šetření European Social Survey (ESS). Získání orientace v těchto velkých a často komplexních datových zdrojích a jejich zpracování v analýze bylo poměrně časově náročné, studie v této práci proto vznikaly v období zhruba tří let. Části této práce již byly publikovány v odborných časopisech. Výsledky analýz českých MTMM dat z šetření European Social Survey jsou součástí textu „*Měření validity a reliability otázek v šetření European Social Survey a jeho využití*“ v recenzovaném časopise Data a výzkum – SDA Info (Chylíková, 2015), výstupy analýzy kvazisimplexovým modelem na datech z českých šetření EU SILC byly publikovány v článku „*Odhad reliability vybraných položek z českých šetření EU SILC kvazisimplexovým modelem*“ v Sociologickém časopise/Czech Sociological Review (Chylíková, 2016). Analýza stylu odpovídání v českých datech ISSP, prezentovaná v páté kapitole této práce dosud nebyla publikována. Jedná se o studii, která bude rozšířena v rámci metodologického projektu CSDA Research, financovaného ze zdrojů OP VVV, který bude probíhat v letech 2017 až 2019. V tomto projektu je naplánováno další měření českého souhlasného stylu odpovídání, jež přispěje k výraznému prohloubení poznatků získaných v analýze prezentované v této dizertační práci. Analýzu stylu odpovídání jsem nicméně již dříve využila ve studii, na které jsem spolupracovala s Martinem Buchtíkem a která prověřovala validitu měření škály pravcového autoritářství v České republice (Chylíková, Buchtík, 2017, v tisku). Uvedená studie ovšem nepřináší meritorní poznatky o českém souhlasném stylu odpovídání.

2. Chyby v datech ze sociálněvědních výběrových šetření

Metoda výběrových šetření svým způsobem umožňuje usuzovat na to, jaké jsou názory, postoje či chování v celé populaci, kterou zkoumáme. Díky ní získáváme jednoznačné hodnoty proměnných, které přinášejí sice redukované, ale do jisté míry přesné a objektivní údaje. Získané ukazatele vlastností, které zkoumáme, nám umožňují matematicky analyzovat společenské fenomény a získávat přehled o stavu v celé společnosti.

V sociologickém kvantitativním paradigmatu se ale v zájmu získání reprezentativních údajů smiřujeme s tím, že získávané informace jsou velice povrchní, a vzdáváme se hlubšího poznávání fenoménů na úkor přesnosti a objektivity. Pokud už tedy činíme takové ústupky ve jménu získání reprezentativních údajů, měli bychom se zabývat i tím, jak spolehlivé ty přesné či objektivní údaje, které získáváme, jsou. Text této kapitoly přináší informace o existenci chyb či odchylek v datech z výběrových šetření a zaměřuje se na výklad o náhodné a systematické chybě měření. Dále prezentuje teoretický rámec pro studium reliability a poskytuje přehled metod pro studium reliability a validity.

2.1 Druhy chyb v sociálněvědních výběrových šetřeních

Ve výzkumu běžně zohledňujeme různé vlivy, které jsou zdroji odchylek či dokonce chyb v datech; je například samozřejmostí, že pracujeme s výběrovou chybou, protože víme, že v datech z výběrového šetření nejsou informace za všechny jednotky v populaci, ale pouze z výběrového souboru, a proto hodnoty získané ze souboru nejsou stejné jako hodnoty v populaci. Rovněž počítáme s tím, že v důsledku výpadku návratnosti naše data nejsou stoprocentně reprezentativní a snažíme se vzniklé odchylky upravovat vážením souboru. Výběrová chyba i odchylka, která vzniká v důsledku výpadku návratnosti, jsou důsledkem toho, že jsme některé jednotky v souboru či v populaci nepozorovali, jsou to tedy chyby chybějícího pozorování. Na druhé straně však existují chyby jiného charakteru, které vyplývají přímo z pozorování. Těmito chybami jsou chyby měření, které se nacházejí přímo v získaných datech.

Všechny odchylky a nepřesnosti v datech jsou zahrnuty do rámce tzv. Celkové chyby šetření (z angl. *total survey error*). Jedná se o koncept amerického výzkumníka Roberta Grovese (Groves, 2004; Groves et al., 2004; Groves et al., 2010), který zastřešuje všechny chyby bez ohledu na to, zda jsou důsledkem pozorování či chybějícího pozorování.

Grovesova *Celková chyba šetření* dělí všechny chyby primárně na systematické (z angl. *bias*) a náhodné, chyby měření pak dělí podle toho, zda jsou jejich zdrojem respondenti,

tazatelé, design dotazníku a mód sběru dat. Předmětem této dizertační práce jsou pouze chyby měření, nikoliv jiné druhy odchylek, a tak zájemce o hlubší porozumění konceptu *Celkové chyby šetření* odkazují na publikace Grovese a kolektivu.

2.2 Chyby měření

Chyby měření jsou části rozptylů měřených proměnných, které neodpovídají měřenému konceptu, resp. znaku. Tato definice vyplývá z tzv. Klasické testové teorie, která je základním předpisem pro analýzu chyby měření. Chyba měření byla v pracích autorů zabývajících se kvalitou dat nejprve definována pouze jako náhodná chyba, tedy fluktuace v hodnotách proměnných, jejíž průměr je nula a která proto neovlivňuje průměry proměnných. Ve skutečnosti však ještě existují systematické chyby, tedy takové rozptyly měřených proměnných, které vychylují jejich průměry jedním či druhým směrem. Typicky je za systematickou chybu považována odchylka, která vzniká vlivem použité metody měření, může jí být ale i vliv tazatelů nebo použitého modu sběru dat. Obecně nejznámějším druhem systematické chyby je odchylka vznikající vlivem sociální desirability, tedy jevu, kdy respondenti zamlčují či přehánějí výpovědi o svém chování, aby byly jejich odpovědi v souladu s panujícími sociálními normami. Když např. provádíme výzkum spotřeby alkoholu, víme, že respondenti systematicky podhodnocují množství spotřebovaných alkoholických nápojů a výsledky z dotazování jsou pak nevalidní. V praktickém výzkumu se však častěji než desirabilitou, jejíž vliv lze empiricky zkoumat velmi omezeně, zabýváme vlivy použitých metod měření, neboť máme přesnější analytické prostředky pro jejich odhad.

V čem spočívá analýza chyb měření? Při přípravě výzkumu a dotazníku postupujeme podle daných pravidel či předepsaných kroků, které musíme udělat, abychom došli od specifikace tématu výzkumu až ke konkrétním dotazníkovým položkám. Nejdůležitější součástí tohoto procesu je volba konstruktů, které chceme zkoumat, nalezení operacionalizací těchto konstruktů, tedy vhodných indikátorů, a formulace položek, které budou součástí dotazníku. O vytvořených položkách pak předpokládáme, že jsou ukazateli konstruktů, který zamýšlíme zkoumat. Většinou si myslíme, že naše položky zkoumaný konstrukt reprezentují dobře, ale nepátráme po tom, do jak velké míry. Zájem o chyby měření je právě takové pátrání a jeho cílem je zjistit, jak velká část rozptylu měřené proměnné konstruktů odpovídá.

O chybách měření je nutné říct, že jsou v datech z výběrových šetření přítomné vždy (Saris, Gallhofer, 2014). Nikdy se nepodaří, aby v praxi uplatněná položka reflektovala zkoumaný koncept (či lépe řečeno znak, jak se lze dočíst v šesté kapitole této práce) stoprocentně, rovněž se nikdy nepodaří, aby bylo měření položky stoprocentně spolehlivé čili reliabilní, tj. aby přinášelo opakovaně stále stejné výsledky. V profesionálním, dobře designovaném a realizovaném výzkumu jsou chyby měření malé, ve výzkumech, které nebyly dostatečně promyšleny a jejich provedení nebylo profesionální, jsou velké. Následující text přináší shrnutí teorie, která stojí za výzkumem chyb měření, a přehled nejznámějších metod, které se k jejich zjištění používají.

2.3 Psychometrie jako základ pro výzkum chyb měření

Vědní disciplínou, ze které vzešly základy pro výzkum chyb měření, je psychometrie, obor psychologie, jenž se zaměřuje na kvantitativní psychologický výzkum. V jejím rámci byla stanovena teoretická východiska a obecně uznávané metody určení chyb měření.

Psychometrie používá dva stěžejní koncepty pro teoretické uchopení chyb měření, a to validitu a reliabilitu. Oba mohla celkem bez problému převzít sociálněvědní metodologie. Validita je v psychometrii i v sociologii nejobecněji definována jako vlastnost měření, které reflektuje to, co chceme měřit, reliabilita je míra, do které je naše měření konzistentní mezi opakovanými měřeními (Řehák, 1998a; Říčan, 1980). Studentům sociálních věd je notoricky známé rčení, že *"validní měření musí být reliabilní, ale reliabilní měření nemusí být validní"*, které v podstatě vystihuje vztah mezi validitou a reliabilitou. Validita je míra, do které se měření blíží skutečné hodnotě pro danou proměnnou, a vztahuje se k tomu, co chceme ve skutečnosti měřit. Na druhé straně reliabilita je přesnost měření, ať už toto měření měří cokoli (Říčan, 1980).

Proč byly základy pro výzkum chyb měření položeny v psychometrii a ne v metodologii výběrových šetření? Je to zřejmě proto, že psychometrie obecně předpokládá existenci přímo nepozorovatelných, tzv. latentních proměnných, které reprezentují nějaký vysoce abstraktní fenomén, např. osobnostní rysy jednotlivce. Tyto abstraktní fenomény, resp. latentní proměnné se v psychometrii měří prostřednictvím indikátorů, tj. položek v psychologickém testu (Groves et al., 2004). Psychometřici tedy přirozeně začali pátrat po tom, do jaké míry se jim tyto latentní proměnné, měřící vlastnosti operacionalizované testem, daří měřit. Naopak ve výběrových šetřeních se výzkumníci a výzkumnice zabývali spíše zjišťováním mnohem méně abstraktních vlastností, tedy např. příjmy, dosaženým

vzděláním, velikostí bytu² atd. (Groves, 2004). Takové měření příliš nesvádělo k pátrání po tom, jak moc položky odráží zkoumaný fenomén, ale s postupným rozvojem sociologie a především zvětšujícími se možnostmi realizace výběrových šetření se rozšířila i oblast zájmu sociologů a do výzkumů se dostaly konstrukty zkoumající vysoce abstraktní fenomény, jako je např. politická orientace či postoje k přistěhovalectví. Tyto konstrukty jsou zjišťovány tzv. subjektivními proměnnými, tj. položkami měřící postoje, hodnoty či názory, u kterých je, stejně jako tomu často bývá v psychometrii, nejasné, do jaké míry skutečně reflektují zkoumaný fenomén. S jejich používáním tedy vznikl opodstatněný důvod používat psychometrickou konceptualizaci a její nástroje odhadu chyb měření (Groves et al., 2004).

2.4 Klasická testová teorie

V předchozím textu bylo naznačeno, že hlavním teoretickým předpisem pro odhad chyby měření, je tzv. Klasická testová teorie (KTT). Tato teorie stanovuje, že rozptyl měřené proměnné je vždy složen ze dvou částí; z části rozptylu odpovídající měřené proměnné (substanciální rozptyl), pro který používá termín pravý skór (z angl. *true score*) a z náhodné chyby (Lord, Novick, 1968; Novick, 1965; Traub, 1997).

Klasickou testovou teorii vyjadřuje rovnice 2.1.

$$(2.1) \quad X = T + E$$

kde E je náhodná chyba měření, T je tzv. pravý skór a X je měření.

Rovnici doplňují následující vlastnosti jejich členů:

1. $E(E) = 0$, tedy že střední hodnota náhodné chyby měření E je nula. Tato vlastnost náhodné chyby měření je právě vyjádřením náhodnosti této veličiny. Důsledkem této vlastnosti je, že chyba E nevychyluje průměr proměnné X.
2. $COV(T, E) = 0$, tedy že pravý skór je nezávislý na chybě měření, což znamená, že neexistuje lineární vztah mezi chybou E a pravým skórem T.

² Uvedené proměnné jsou příkladem tzv. objektivních proměnných. To jsou takové proměnné, které je případně možné validně zjistit i jiným způsobem než výpovědí respondenta (kontrolou úředních záznamů, dotazem na obeznámenou osobu atd.)

3. Při více měřeních jsou náhodné chyby E nekorelované. Z hlediska tohoto požadavku definuje náhodnou chybu Andrews (1984) jako „*odchylky v jedné měřené proměnné, které jsou nezávislé na odchylkách všech ostatních současně měřených proměnných*“.

Pravý skór je teoretický koncept, myšlenkový model, který vyjadřuje předpoklad existence nějaké skutečné hodnoty konkrétní proměnné (znaku) pro konkrétního respondenta.

Rovnici výše lze podle tohoto předpokladu přepsat jako:

$$(2.2) \quad x_{ij} = \tau_{ij} + e_{ij},$$

kde index i odpovídá jednotlivému respondentovi, j vyjadřuje měřený znak (Novick, 1965). Pravý skór má zásadní význam pro úvahy o reliabilitě měření. V empirii je každá proměnná do jisté míry zanesena náhodnou chybou měření vznikající během výzkumného rozhovoru z různých důvodů; takovými důvody může být špatné porozumění otázce, omyl při odpovědi, nedostatečná koncentrace atp. KTT bere tuto náhodnou chybu v úvahu tím, že představuje koncept pravého skóru, který reprezentuje tu část rozptylu proměnné, která zbude po očištění o chybový rozptyl.

2.5 Typy měření

Při používání metod zjišťování proporce náhodné chyby měření vyplývajících z KTT je nutné vědět, jaké vlastnosti mohou mít opakovaná měření té samé proměnné a jaký mají tyto vlastnosti vliv na usuzování o reliabilitě. Existují tři typy měření, které se liší proporcí částí rozptylů náhodné chyby a pravého skóru.

Uvažovanými třemi typy měření jsou (Bollen, 1989):

1. Paralelní měření
2. Tau – ekvivalentní měření
3. Kongenerické měření

Pokud máme dvě měření x_i a x_j , z nichž každé vyjádříme rovnicí:

$$(2.3) \quad x_i = \alpha_i \tau_i + e_i$$

$$(2.4) \quad x_j = \alpha_j \tau_j + e_j$$

kde α_i a α_j jsou regresní koeficienty, jenž značí závislost měření x na pravém skóru τ , a kde platí, že $\text{COV}(e_i, e_j) = 0$ (tzn. náhodné složky měření e nejsou korelovány), a skutečné

skóry τ_i a τ_j jsou stejné, můžeme ukázat rozdíl mezi jednotlivými typy měření uvedenými výše.

- Paralelní jsou taková měření, kde α_i a α_j jsou stejné a $\text{VAR}(e_i) = \text{VAR}(e_j)$, tzn. jsou to měření, ve kterých měřená proměnná reflektuje pravý skór v obou měřeních stejně, a složky rozptylu měřené proměnné odpovídající náhodné chybě měření jsou stejné.
- Pro tau – ekvivalentní měření rovněž platí, že $\alpha_i = \alpha_j$, ovšem rozptyly náhodných složek už se nerovnají, tedy $\text{VAR}(e_i) \neq \text{VAR}(e_j)$.
- Pro kongenerické měření neplatí ani jeden ze dvou uvedených předpokladů, tedy platí, že $\alpha_i \neq \alpha_j$ a $\text{VAR}(e_i) \neq \text{VAR}(e_j)$.

Klasické metody odhadu reliability zpravidla předpokládají právě paralelní měření, které je však ze všech typů měření nejméně časté (Bollen, 1989). Nejčastější je právě kongenerické měření, které ale teoretickým předpokladům metod odhadu reliability zpravidla nevyhovuje. Z toho důvodu je např. notoricky známá metoda odhadu reliability test – retest navzdory všeobecné oblíbenosti často nevyhovující.

2.6 Reliabilita

Vysoká reliabilita měření vyjadřuje předpoklad o tom, že odpovědi respondentů na otázku jsou stabilní, resp. že by byly stabilní, kdyby se proměnná měřila znovu. Čím je fluktuace v hodnotách odpovědí napříč pokusy větší, tím je reliabilita měření nižší. Reliabilitu lze chápat jako stabilitu v odpovědích jednoho respondenta v nekonečném počtu pokusů, ale i jako konzistenci v odpovědích respondentů ve výběrovém souboru, kteří byli dotázáni opakovaně (Bollen, 1989). Z této druhé možnosti vycházejí empirické metody odhadu reliability.

Vzhledem k výše uvedenému výkladu je jasné, že definice reliability vychází z KTT. Pro dvě měření x_i a x_j můžeme rozložit rozptyly měřené proměnné podle následujících rovnic:

$$(2.5) \quad x_i = \alpha_i \tau_i + e_i$$

$$(2.6) \quad x_j = \alpha_j \tau_j + e_j$$

kde podle KTT platí, že: $\text{COV}(\tau_i, e_i) = \text{COV}(\tau_j, e_j) = 0$; $E(e_i) = E(e_j) = 0$; $\text{COV}(e_i, e_j) = 0$.

Samotná reliabilita měření X je dána předpisem:

$$(2.7) \quad \text{Rel}(X) = \text{VAR}(T) / \text{VAR}(X)$$

Rovnice (2.7) vyjádřena slovy znamená, že reliabilita je poměr rozptylu pravého skóru T a rozptylu měřené proměnné X. Definice reliability poměrem rozptylů vychází z faktu, že reliabilita je konzistence opakovaných měření, proto je nutné zkoumat rozptyl, tedy měřítko variability proměnné, a nikoliv jinou míru, např. průměr. Jak dále vyplývá z rovnice, je reliabilita vyjádřitelná číslem mezi 0 a 1, neboť rozptyl proměnné X může být buď stejně velký jako rozptyl pravého skóru T, nebo větší než rozptyl pravého skóru T. Měření s reliabilitou blízké nule je vysoce nereliabilní, měření přibližující se jedné je vysoce reliabilní. O lineárním vztahu mezi měřením X a pravým skórem T platí, že reliabilita je čtvercem korelace mezi měřenou proměnnou a jejím pravým skórem, tedy $\text{Rel}(X) = r^2(T, X)$. To je nutné si uvědomit, když zjišťujeme reliabilitu z modelu strukturálních rovnic jako je např. True Score MTMM model popsáný v Kapitole 6 či QSM model, použitý ve čtvrté kapitole této práce.

2.7 Metody odhadu reliability

Odborná literatura poskytuje přehled několika metod odhadů reliability založených na vztazích vyplývajících z KTT. Cílem tohoto textu není podat vyčerpávající přehled těchto metod a jejich podrobný popis, ale uvést ty nejdůležitější a ty, jejichž aplikace je předmětem analýz v této dizertační práci. Zájemce o přehledy metod odkazují na práce Bollen, Grovese a dalších autorů, citovaných v této kapitole.

Groves (Groves et al., 2004) rozděluje metody odhadu reliability na dvě skupiny:

1. Používající data z opakovaných měření
2. Používající více indikátorů měřeného konceptu

První uvedené jsou metody, které jsou realizovány prostřednictvím opakovaného dotazování položky měřící ten samý pravý skór a těch samých respondentů. Takovými metodami jsou test – retest, kvazisimplex model a metoda alternativních měření. Do druhé skupiny patří metody, které získávají více indikátorů pro odhad vztahu mezi měřeními x a

τ v jednom časovém bodě. Do této skupiny patří metody *split halves*, Cronbachovo alfa³ a True score MTMM model, který kromě reliability nabízí i odhad efektu metody.

2.7.1 Metody odhadu reliability používající opakovaná měření

Obecně pro všechny metody odhadu reliability využívající opakovaná měření existují předpoklady, které musí být splněny. Jsou to (Groves et al., 2004):

1. v období mezi měřeními nenastala změna v pravém skóru τ^4 ;
2. okolnosti dotazování jsou ve všech opakováních rozhovoru stejné;
3. respondent si nepamatuje svou odpověď z předchozího měření, tzn. měření jsou na sobě nezávislá.

Dodržet všechna výše uvedená pravidla v reálném výzkumu bohužel není zcela možné. Detailní vysvětlení, proč jsou tato pravidla v praxi problematická, podávám v této kapitole v popisu metody test – retest a v Kapitole 4, kde ukazují aplikaci kvazisimplexového modelu. Na tomto místě jen ve stručnosti uvádím, že výzkum s opakovanými měřeními je vždy do jisté míry problematický a výsledky z každé takové metody lze podezírat z nepřesnosti. Opakovaným měřením se lze v reálném metodologickém výzkumu však jen stěží vyhnout; jsou totiž jedním z mála způsobů, jak odlišit v měřených proměnných náhodné a systematické rozptyly. Proběhly sice pokusy eliminovat některé z problémů vyplývajících z předpokladů uvedených výše, jak nasvědčuje např. v šesté kapitole popisovaný 2 SB MTMM design sběru dat, ukazuje se ale, že eliminace problému vyplývajícího z opakovaného měření dává vzniknout problému novému a výzkum s využitím metody se stále potýká s komplikacemi.

2.7.1.1 Metoda test – retest

Jednoznačně nejznámější metodou odhadu reliability je metoda test – retest, běžně vyučovaná již v bakalářských kurzech sociálněvědní metodologie. Provedení metody spočívá ve dvakrát opakovaném měření, v každém z nich odpoví ten samý respondent na tu samou dotazníkovou položku. Reliabilita je pak určena jako korelace mezi prvním a druhým měřením. Metoda předpokládá, že obě měření jsou paralelní měření a že měření

³ Existují ještě další alternativy výpočtu vnitřní konzistence škály, ty zmiňuje např. Schubert (2010) nebo Ferjenčík (2000).

⁴ Tato podmínka se netýká kvazisimplexového modelu (QSM).

jsou na sobě nezávislá, tedy že měření v čase t neovlivňuje měření v čase $t + 1$ (Bollen, 1989). Dvakrát opakované měření znázorňují rovnice:

$$(2.8) \quad x_t = \alpha_t \tau_t + e_t$$

$$(2.9) \quad x_{t+1} = \alpha_{t+1} \tau_{t+1} + e_{t+1}$$

kde x_t je první měření, x_{t+1} druhé měření. Dále opět platí předpisy dané KTT, a to že:

$$E(e_t) = E(e_{t+1}) = 0;$$

$$\text{COV}(\tau_t, e_t) = \text{COV}(\tau_{t+1}, e_{t+1}) = 0$$

$$\text{COV}(e_t, e_{t+1}) = 0$$

K metodě test – retest se váží problémy vyplývající z podmínek pro opakované měření:

1. Metoda test – retest předpokládá absolutní stabilitu pravého skóru (τ) v čase, ten se však reálně může v čase mezi měřeními změnit. Buď proto, že doba, která uplynula mezi dvěma měřeními, byla příliš dlouhá, nebo proto, že v době mezi měřeními došlo k nějaké události, která mohla pravý skór τ změnit. Mohlo rovněž dojít k tzv. reaktivitě (Bollen, 1989), což je jev, kdy samotné dotazování při prvním měření indikuje změnu v pravém skóru. Pokud ke změně pravého skóru dojde, korelace mezi prvním a druhým měřením už se nerovná reliabilitě měření a odhad reliability je nepřesný.
2. V důsledku efektu paměti, tedy případu, kdy si respondent pamatuje svou první odpověď, může rovněž dojít k porušení předpokladu nezávislosti měření. Pokud měření proběhla v krátké době po sobě, vede to k tomu, že druhé měření je závislé na prvním, protože si respondent může svou odpověď pamatovat. Tím je porušena i podmínka nekorelovanosti náhodných chyb e a odhad reliability opět není přesný (Sarıs, Gallhofer, 2014).
3. Velmi problematický je předpoklad paralelního měření. Ve výzkumné praxi se paralelní měření vyskytují zřídka, a proto je třeba počítat s tím, že měření je minimálně tau – ekvivalentní. Potom však už není možné počítat s platností rovnic, a tak se korelace x_t a x_{t+1} nerovná reliabilitě (Bollen, 1989).

Vzhledem k výše uvedenému je použití metody test – retest navzdory oblíbenosti poměrně diskutabilní a např. Bollen či Saris a Gallhoferová ji vůbec nedoporučují k použití (Bollen, 1989; Saris, Gallhofer, 2014).

2.7.1.2 Kvazisimplexový model (QSM) a metoda alternativních měření

Dalšími dvěma metodami, které využívají opakovaného pozorování, jsou kvazisimplexový model (z angl. *quasi simplex model*, QSM; Heise, 1969) a metoda alternativních měření (Bollen, 1989). Kvazisimplexový model využívá data z panelových šetření a jeho nesporná výhoda spočívá v tom, že nepotřebuje žádná další měření nad rámec substanciálních měření. Odhady reliability získává v SEM modelu na datech z minimálně tří, resp. čtyř panelových vln měření, přičemž k odhadu změny v hodnotě pravého skóru, která nastala v období mezi dvěma měřeními, používá matematickou funkci. Detailní popis této metody obsahuje Kapitola 4, ve které je kvazisimplexový model aplikován na data z českých šetření EU SILC. Druhá uvedená metoda je založena na opakovaném měření stejného konceptu dvěma různými metodami, odhadem reliability je pak korelace mezi dvěma měřeními (Bollen, 1989). Metoda alternativních měření se od dalších uvedených liší tím, že nezjišťuje reliabilitu konkrétní dotazníkové položky, ale reliabilitu měření jednoho pravého skóru. Trendem v oblasti výzkumu chyb měření je právě zjišťování reliability konkrétních položek, a to z toho důvodu, že je pak možné usuzovat na vliv charakteristik položky na kvalitu měření (Andrews, 1984; Saris, Gallhofer, 2014; Saris, Gallhofer, 2007). I proto je metoda alternativních měření v současné době marginální.

2.7.2 Metody odhadu reliability používající více indikátorů měřeného konceptu

Zdaleka nejznámější multi – indikátorovou metodou odhadu reliability je Cronbachovo alfa. Tato metoda je běžně používaná pro stanovení konzistence položek v baterii, resp. škále, a to jak v sociologii, tak v psychologii. Aby bylo možné hodnotu Cronbachova alfa odhadnout, je nezbytné, aby položky baterie či testu byly formulovány k měřenému konstruktovi ve stejném směru, tj. buď pozitivně či negativně. V případě, kdy používáme data z tzv. vyvážené škály⁵, tedy musíme hodnoty pro část položek překódovat. Metoda má realističtější vlastnosti než např. metoda test – retest; díky měření v jednom časovém bodě eliminuje změnu pravého skóru v čase a umožňuje přesný odhad reliability jak pro

⁵ Vyvážená škála je baterie položek, ve které je jedna polovina položek formulována směrem k měřenému konstruktovi a druhá polovina položek je formulována proti směru měřeného konstruktovi.

paralelní, tak pro tau – ekvivalentní měření (Bollen, 1989). Jejím problémem ovšem je, že neposkytuje odhad reliability pro jednotlivou položku, ale pro celou baterii položek jako celek.

Další multi – indikátorovou metodou odhadu reliability je True Score MTMM model. Učebnice tuto metodu běžně neuvádějí, neboť se jedná o metodu poměrně novou; objevila se v podstatě až po roce 2000 (Saris, Gallhofer, 2014; Saris, 2009). TS MTMM model získává odhady reliability a systematické chyby vzniklé vlivem použité metody měření z dat designu MTMM, resp. 2 SB MTMM, jedná se tedy o způsob zjišťování reliability a validity zároveň. Validita je v tomto pojetí ten rozptyl, který neobsahuje náhodnou chybu a systematickou chybu, vzniklou vlivem použité metody měření. TS MTMM model je předmětem kapitoly 6, kde je plně popsán a aplikován na česká data z šetření European Social Survey (ESS).

2.8 Validita

Validita je překryv mezi konceptem, který zamýšlíme měřit, a naším měřením. Je to tedy míra, do které měřená proměnná přesně odráží zamýšlený konstrukt (Groves, 2004). Některé druhy validity jsou čistě teoretická uchopení problému validity, které nelze přímo aplikovat v analýze kvantitativních dat (obsahová validita), jiné lze využít přímo ke kontrole, zda měření bylo validní (kriteriální validita). Další druhy validit nepodávají přesné informace o validitě, ale spíše naznačují, zda je měření validní (konstruktová validita) či jak je zkoumaný konstrukt náchylný ke změně v důsledku použité metody měření (split – ballot). V následujícím textu krátce popisuji možné druhy validit, o kterých se v metodologii sociálních věd běžně mluví, nepodávám však vyčerpávající seznam metod a typů validity, neboť od toho jsou texty přehledového charakteru. V tomto úvodu mi jde především o to, ukázat, že existují i jiná pojetí validity než to, které používám v této dizertační práci.

2.8.1 Druhy validity, resp. metody určení či odhadu validity

2.8.1.1. Obsahová validita

Čistě teoretickou konceptualizací validity je obsahová validita (z angl. *content validity*) a znamená přesah mezi tím, co chceme měřit, a tím, co skutečně měříme. K chybě, která má za následek porušení obsahové validity, dochází v procesu operacionalizace, a to v případech, kdy jsou zvoleny nevhodné indikátory pro měření zkoumaného konstrukt (Groves et al.,

2004). Obsahovou validitu nelze číselně určit, její stanovení je vlastně jen subjektivní zhodnocení, zda jsou vybrané indikátory vhodné pro měření daného konstruktů.

2.8.1.2 Kriteriaální validita

U objektivních proměnných lze využít nějakou externí kontrolu validity měření a v odborné literatuře bývá tato kontrola označována termínem kriteriaální validita. Tento druh validity předpokládá, že pro některé proměnné lze najít na respondentech nezávislý zdroj informací, který není zanesen chybou. Hodnoty měřené proměnné jsou pak porovnány s údaji z nezávislého zdroje. Zdrojem kritérií při kontrole validity mohou být úřední záznamy, výsledky voleb, informace z cenzů atp., proto je tato validita, resp. její metoda rovněž označována jako tzv. *record check study* (Groves et al., 2004). Validita měření zjištěná prostředky kriteriaální validity je míra, do které měření koreluje s vybraným kritériem (Bollen, 1989; Lord, Novick, 1968). Korelace mezi měřením a kritériem ovšem nemusí zcela odpovídat validitě, neboť i kritéria mohou a často jsou zanesena chybami (např. úřední záznamy jsou neúplné nebo obsahují chyby). V takovém případě samo kritérium není se skutečným skórem v perfektním vztahu a získaný odhad validity je nepřesný (Bollen, 1989).

Příkladem kriteriaální validity může být srovnání výpovědí respondentů o četnosti návštěv zubaře s oficiálními záznamy zubařů. Dalším klasickým příkladem je srovnání výpovědí respondentů o volbě politické strany v posledních volbách s výsledky těchto voleb. Kontrola měření s úředními záznamy je užitečná především v případě proměnných, které mohou být deformovány sociální desirabilitou, tedy jevem, kdy respondent v rozhovoru záměrně podává odpověď, která je blíže sociální normě. Sociální desirabilita pak působí především v šetření citlivějších témat jako je zdraví, příjmy či páchání dopravních přestupků (Chylíková, 2011).

2.8.1.3. Konstruktová validita

Konstruktová validita je soubor teoreticky zdůvodněných předpokladů o vztazích mezi proměnnými v konkrétním modelu (Bollen, 1989). Tyto teoretické předpoklady musí být založeny na dlouhodobě pozorovaných a teoreticky vysoce opodstatněných vztazích mezi proměnnými. Při aplikaci konstruktové validity z teorie vyvodíme různou velikost a směr vztahů mezi měřenými proměnnými a ověřujeme, zda stejné vztahy nalezneme i v datech. Pokud jsou zjištěné vztahy v souladu s předpoklady, je podle tohoto přístupu měření validní; pokud ne, je nevalidní. Měříme-li např. na vzorku populace USA sociálně

konzervativní postoje, očekáváme, že jsou četnější mezi voliči republikánské strany, neboť takový předpoklad stanovuje teorie i předchozí empirické zkoumání. Pokud data předpokládané vztahy potvrdí, usuzujeme, že měření je validní. Konstruktová validita je stejně jako např. výše uvedená obsahová validita metodou, která neposkytuje odhad validity a nepřináší číselně vyjádřitelnou míru validity měření. Tento typ validity je především požadavkem na dobrou praxi, která by měla být samozřejmostí při práci s daty z výběrových šetření.

2.8.1.4 Design split – ballot

Split – ballot je tradiční a poměrně rozšířená metoda, která se uplatňuje především při měření subjektivních proměnných, nejčastěji postojů. Pro měřený konstrukt získává více měření, typicky dvě (Bollen, 1989; Groves et al., 2004). Každé měření je provedeno na podsouboru respondentů, takže se nejedná o metodu tzv. opakovaných měření, ale o multi – indikátorovou metodu. Metoda vychází z empiricky pozorovaného jevu, že různé charakteristiky položky nebo dotazníku mají systematický vliv na pozorované hodnoty proměnných. Cílem použití metody je zjistit, jak konkrétní varianta formulace položky či designu dotazníku⁶ ovlivní pozorování, jde tedy o studium vlivu systematické chyby měření. Praktické použití metody spočívá ve vytvoření výběrových podsouborů, do nichž jsou respondenti vybráni náhodně (Biemer, 2004), každý podsoubor pak dostane jinou variantu metody měření. Častým příkladem použití metody split – ballot je aplikace jedné a té samé dotazníkové položky na různých místech v dotazníku, která zjišťuje efekt kontextu na měření. Často se také zjišťuje vliv formulace položky a použití různých slov. Pozorované rozdíly ve frekvencích odpovědí či v průměrech pak značí systematickou chybu vzniklou v důsledku změny kontextu. Nevýhodou přístupu je, že nedokáže určit, která z použitých metod je „lepší“, tj. neposkytuje číselný odhad systematické chyby pro každou z variant měření. Výzkumníci se musí spokojit jen s rozdílem hodnot použitých proměnných.

2.8.1.5 Konvergentní a diskriminantní validita a design Multitrait multimethod (MTMM)

Konvergentní a diskriminantní validita je vyjádření předpokladu, že měření stejného znaku, tedy indikátoru daného konstruktů, stejnou metodou spolu musí maximálně

⁶ Je možné zkoumat např. i systematické rozdíly v různých modech sběru dat či vliv tazatelů.

korelovat a měření různých znaků různými metodami by spolu mělo korelovat minimálně. Myšlenku představili Campbell a Fiske (1959) a dali jejím prostřednictvím vzniknout experimentálnímu designu Multitrait multimethod (MTMM), který s využitím více položek měří znaky vybraného konstruktů více metodami měření. Nejčastěji používanou formou MTMM je design s devíti položkami, které měří tři znaky třemi metodami. Metodu MTMM inovoval Andrews (1984), který navrhl model strukturálních rovnic, ve kterém lze odhadnout číselné koeficienty validity. V jeho modelu je validita ta část rozptylu odpovídající měřenému znaku a je inverzní k systematické chybě měření vzniklé vlivem použité metody. Nespornou výhodou modelování MTMM dat je, že umí podat číselnou informaci o validitě měření, kterou získá z odhadu parametrů strukturálního modelu. Nevýhodou naopak je, že metoda vyžaduje několik opakovaných měření; vztahují se tak na ní typické problémy metod opakovaných měření a nadměrně zatěžuje respondenty. Detailní popis designu MTMM a strukturálního modelu, ve kterém lze MTMM data modelovat, popisuje text šesté kapitoly této dizertační práce.

3. Teoreticko – analytický rámec dizertační práce

3.1 Rozšíření Klasické testové teorie o systematickou složku

Klasická testová teorie (KTT), popsaná v předcházející kapitole, ve své podstatě považuje za chybu měření tu část rozptylu měřené proměnné, která neodpovídá měřené vlastnosti. Podle KTT je rozptyl měřené proměnné vždy složen ze dvou částí; z části rozptylu odpovídající měřenému znaku (substanciální rozptyl), pro který používá termín *pravý skór* a z náhodné chyby (Lord, Novick, 1968). Přestože KTT bere do úvahy jen rozptyl měřeného atributu a náhodné chyby, není důvod, proč tuto teorii nerozšířit o vliv systematické chyby. Rozšíření spočívá v tom, že je do modelu vztahů mezi měřenou proměnnou a pravým skórem začleněn ještě vliv systematické chyby měření. **Rozptyl měřené proměnné je tak v rozšířené verzi KTT rozdělen do tří částí; na část odpovídající pravému skóru, na část odpovídající náhodné chybě měření a na část odpovídající systematické chybě měření** (Saris, Galhofer, 2014). Rozšířená KTT tak může svým způsobem pracovat s validitou, kterou speciálně definuje jako inverzní k proporcii systematické chyby v rozptylu pravého skóru. Rozšířenou verzi KTT prezentuje Saris a Gallhoferová (2014), kteří rozptyl měřené proměnné dělí podle rovnic:

$$(3.1) \quad y_{ij} = r_{ij} T_{ij} + e_{ij}$$

$$(3.2) \quad T_{ij} = v_{ij} F_i + m_{ij} M_j$$

kde y_{ij} reprezentuje proměnnou, která je indikátorem znaku i a je měřena metodou j , T_{ij} je pravý skór znaku i měřené proměnnou j , a je e_{ij} náhodná chyba proměnné y_{ij} . F_i reprezentuje měřený znak i a M_j použitou metodu měření j . Příslušné koeficienty (r , v , m) vyjadřují sílu vztahu mezi měřenou proměnnou a pravým skórem, resp. konstruktem, resp. efektem metody; r_{ij} odpovídá koeficientu reliability, v_{ij} koeficientu validity a m_{ij} efektu metody. Z rovnic je jasně patrné, že rozdělují rozptyl měřené proměnné y_{ij} na tři části; na část odpovídající pravému skóru a část odpovídající náhodné chybě měření, přičemž pravý skór je tvořen rozptylem odpovídajícím měřenému konstruktovi a systematické chybě vzniklé v důsledku použité metody. Oproti KTT jsou zde tedy dvě inovace; je začleněn vliv systematické chyby a je implementována latentní proměnná konstruktu, jejímž indikátorem je pravý skór měřené proměnné.

Praktické využití Klasické testové teorie šlo plně rozvinout až se vznikem a rozšířením analytické metody modelování strukturálních rovnic (SEM). To ostatně bylo samo o sobě podnětem k rozšíření KTT, když analytici začali v datech kromě náhodné chyby hledat i systematické odchylky (Andrews, 1984). Nejvýznamnějším přínosem analytické metody SEM je totiž její schopnost přímo identifikovat části rozptylů v měřených proměnných. SEM umí rozdělit rozptyl měřených proměnných na různé části podle toho, které latentní proměnné jsou jejich zdrojem. Při správném designu a vhodném modelu tak umí určit složky rozptylů odpovídající měřenému konstrukt, náhodné chybě měření a systematické chybě.

3. 2 Teoreticko – analytický rámec práce

Klasická testová teorie, její rozšířená verze a analytická metoda modelování strukturálních rovnic tvoří teoreticko – analytický rámec pro všechny studie prezentované v rámci této dizertační práce. Uplatněný teoreticko – analytický rámec není ve své podstatě originálním přínosem tohoto textu. KTT a SEM ve svých studiích využívají výzkumnice a výzkumníci již od 80. let (Andrews, 1984; Saris, Andrews, 1991; Saris, Sattora, Coenders, 2004; van Meurs, Saris, 1990; Billiet, McClendon, 2000; Revilla, Saris, 2013a) a rozšířenou verzi KTT ve své knize explicitně prezentují Saris a Gallhoferová (2014), kteří ovšem toto rozšíření nepojmenovávají, a to ani přízviskem „rozšířená verze“. Moje dizertační práce jejich rozšíření přejímá a poukazuje na to, že lze v jeho rámci chápat všechny známé metody odhadu reliability a validity v modelech SEM. Přístup lze aplikovat absolutně, tj. v celé jeho šíři, kdy lze zjišťovat validitu i reliabilitu (TS MTMM), nebo částečně, kdy v jeho rámci můžeme chápat metody zjišťující jen náhodou chybu (QSM) nebo metody zjišťující jen systematickou chybu (CF-RS model). Studie v Kapitole 4 pracuje s metodou určení reliability měření nazvanou kvazisimplexový model (QSM), která zjišťuje proporci náhodné chyby měření, její princip je tedy dán již rámcem KTT. Metoda odhadu chyby, jež vzniká v důsledku souhlasného stylu odpovídání, v tomto textu nazývaná CF-RS model, je prezentována a aplikována v Kapitole 5. Tato metoda bere v úvahu pouze systematickou chybu, nikoliv náhodnou chybu, operuje tak v rámci rozšířené KTT pouze částečně. True Score MTMM model aplikovaný na 2 SB MTMM data z šetření ESS zcela využívá rozšířenou KTT a je předmětem výkladu v Kapitole 6. Teoreticko – analytický rámec KTT a SEM rovněž platí pro sedmou kapitolu této dizertační práce, ve které ukazují vliv chyby měření na výsledky substanciálních analýz a ilustrují metodu očištění dat o chybové rozptyly.

3. 3 Analytická metoda modelování strukturálních rovnic (SEM)

Pro pochopení principu metod popisovaných v této práci a zejména pro porozumění analýz, které používám, je nutné mít alespoň základní představu o tom, co je modelování strukturálních rovnic (z angl. *structural equation modeling*, SEM), na jakém principu je postaveno a jaké problémy doprovázejí jeho aplikaci. Vzhledem k tomu, že SEM není v české kvantitativní sociologii používáno úplně běžně a že je v rámci analytických kurzů na katedrách sociologie vyučováno zpravidla jen okrajově, považuji za nutné v tomto textu uvést alespoň základní a do velké míry zjednodušený výklad o analýze SEM. Na rozdíl od většiny učebních textů nemá tento stručný úvod vyšší matematicko – statistické ambice a jde v něm výhradně o to, přiblížit analýzu méně zkušeným čtenářům a čtenářkám.

V následujícím textu nejdříve obecně popisuji, co je model aplikovaný na sociálněvědní kvantitativní data, a pak představuji metodu SEM, její principy, používaná pravidla a problémy, které se k jejímu používání váží.

3. 3. 1 Model v analýze kvantitativních dat

V této práci se opakovaně uvádí slovo „model“ či „modelování“, proto stojí za to krátce a stručně uvést, v čem modelování kvantitativních dat vlastně spočívá. Na té nejobecnější úrovni jde o to, zjistit, do jaké míry data odpovídají teoretickým předpokladům vyjádřeným právě v modelu. Již na začátku výzkumu zpravidla existují nějaké informace týkající se vztahů mezi měřenými proměnnými. Tyto vztahy jsou uvedeny v nějaké teorii či v hypotézách odvozených z této teorie a lze je vyjádřit ve formě modelu. Pokud chceme zjistit, zda jsou předpokládané vztahy přítomny v datech, použijeme na tato data model a zajímáme se o to, do jaké míry model odpovídá datům. V samotném modelu jde pak o to odhadnout parametry, tedy např. korelace mezi latentními proměnnými, regrese jedné latentní proměnné k jiné nebo faktorové zátěže mezi latentní a měřenou proměnnou. Informaci o tom, nakolik hodnoty odhadnutých parametrů odpovídají skutečným vztahům v datech, podávají ukazatele shody modelu s daty, resp. doporučené intervaly hodnot těchto ukazatelů. Pokud jsou ukazatele shody modelu s daty v přijímaných mezích, lze o modelu říct, že dobře či skvěle odráží vztahy v datech a lze z něj usuzovat na zkoumanou populaci. V následující části textu ilustruji ve velmi zjednodušené formě principy modelování strukturálních rovnic, na jejichž základě jsou odhadnuty parametry SEM modelu a určena shoda modelu s daty. Jde skutečně o velká zjednodušení, která činím proto, aby byly základní charakteristiky SEM dobře srozumitelné začátečníkům.

3. 3. 2 Modelování strukturálních rovnic s latentními proměnnými

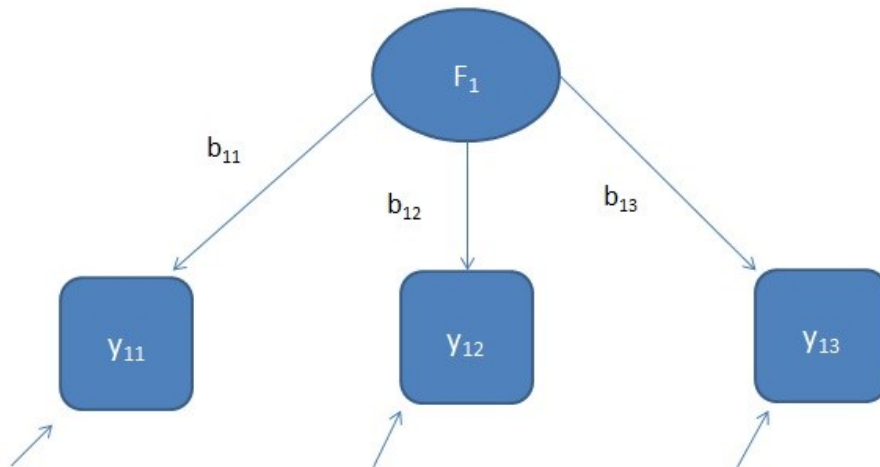
Strukturální modely mohou být různého typu; mohou znázorňovat vztahy jen mezi měřenými proměnnými (pěšinková analýza, z angl. *path analysis*) nebo se zabývají pouze vztahy mezi latentní proměnnou a měřenými proměnnými (konfirmativní faktorová analýza; z angl. *confirmatory factor analysis*, CFA). Nejobecnějším strukturálním modelem je obecný model s latentními i měřenými proměnnými (Bollen, 1989), který je kombinací dvou výše uvedených modelů. Tento model má dvě části, první je model měření (z angl. *measurement model*), který znázorňuje vztahy mezi měřenými a latentními proměnnými, druhou je část s latentními proměnnými (Bollen, 1989) či jinak řečeno strukturální model. V této dizertační práci používám obecný model s latentními i měřenými proměnnými v kapitole 4 a 6, kde je prezentována metoda kvazisimplexového modelu a True score MTMM model. Kapitola 5 a částečně kapitola 6 ilustruje použití metod založených na modelu vztahů mezi latentní a měřenou proměnnou (CF-RS Model, MTMM model).

Modelování strukturálních rovnic je založeno na analýze kovariančních struktur, tj. kovariančních (či korelačních) matic proměnných, které vstupují do analýzy. V tomto se SEM přístup liší např. od známé lineární regrese, kde jsou analyzovány jednotlivé případy (v datech z šetření odpovídající jednotlivým respondentům). Z této vlastnosti SEM analýzy vyplývá i princip, na jehož základě je vypočítána shoda modelu s daty; při výpočtech jde o to, minimalizovat rozdíly mezi maticí kovariancí, jež vstoupila do analýzy, a kovarianční maticí, jež je vypočtena na základě odhadnutých parametrů (Bollen, 1989).

3. 3. 3 Odhad parametrů v modelu

Tzv. parametry modelu jsou faktorové zátěže, korelační a regresní koeficienty a rozptyly měřených a latentních proměnných. Tyto parametry jsou odhadovány z dat, které vstupují do analýzy. Názornou ukázkou principu jejich odhadu v modelu měření ukazují Saris a Gallhoferová (2014). Jejich ukáзка je sice velmi zjednodušující, ale dobře ilustruje princip, na kterém je odhad založen. Autoři v příkladu používají korelace mezi měřenými proměnnými ze souboru dat (nikoliv jejich populační odhady) a ukazují, jak lze s jejich využitím ze vztahů v modelu vypočítat faktorové zátěže.

Obrázek 3.1



Obrázek 3.1 ukazuje model konfirmativní faktorové analýzy (CFA), ve kterém je jedna latentní proměnná F_1 , která má tři indikátory y_{11} , y_{12} a y_{13} . Od latentní proměnné (vždy značené oválem) vedou faktorové zátěže (zde značeny b_{11} , b_{12} a b_{13}) k měřeným proměnným (vždy značeným čtvercem či obdélníkem). Pro úplnost tento grafický model vyjadřují ve zjednodušeném zápisu:

$$(3.3) \quad y_{ij} = b_{ij} F_i + e_{ij}$$

kde i je identifikátor měřené proměnné a j identifikátor latentní proměnné. V rovnici (3.3) reprezentuje y_{ij} měřenou proměnnou, která je indikátorem konstruktů F_i ; b_{ij} je faktorová zátěž mezi měřenou proměnnou y_{ij} a latentní proměnnou F_i , e_{ij} je zbytkový rozptyl proměnné y_{ij} , který neodpovídá konstruktům F_i .

Podle pravidel modelování strukturálních rovnic lze příslušné parametry b_{ij} vyjádřit jako rovné korelacím, jak to ukazují rovnice (3.2) až (3.7):

$$(3.4) \quad \rho(y_{11}, F_1) = b_{11}$$

$$(3.5) \quad \rho(y_{12}, F_1) = b_{12}$$

$$(3.6) \quad \rho(y_{13}, F_1) = b_{13}$$

$$(3.7) \quad \rho(y_{12}, y_{11}) = b_{12} b_{11}$$

$$(3.8) \quad \rho(y_{13}, y_{11}) = b_{13} b_{11}$$

$$(3.9) \quad \rho(y_{13}, y_{12}) = b_{13} b_{12}$$

Všimněme si, že rovnice obsahují nejen korelace mezi měřenými proměnnými, které známe, ale i korelace mezi měřenými proměnnými a latentní proměnou. Vzhledem k tomu, že neznáme hodnoty latentní proměnné, neznáme ani její korelace s měřenými proměnnými, proto si musíme pomoci úpravou rovnic:

$$(3.10)$$

$$\frac{\rho(y_{12}, y_{11})}{\rho(y_{13}, y_{11})} = \frac{b_{12} b_{11}}{b_{13} b_{11}} = \frac{b_{12}}{b_{13}}$$

$$(3.11)$$

$$b_{12} = b_{13} \frac{\rho(y_{12}, y_{11})}{\rho(y_{13}, y_{11})}$$

kde po substituci z (3.7) dostaneme

$$(3.12)$$

$$b_{13}^2 = \frac{\rho(y_{13}, y_{11}) \rho(y_{13}, y_{12})}{\rho(y_{12}, y_{11})}$$

Z rovnice (3.12) již lze při znalosti korelací měřených proměnných odhadnout velikost parametru b_{13} . Obdobně jako v ilustrované úpravě výrazů lze postupovat při hledání hodnot dalších parametrů b_{ij} . Uvedený příklad a rovnice názorně ilustrují, odkud se odhady parametrů v modelu berou, a ukazují, že zdrojem pro odhady parametrů jsou kovariance (či variance, tj. rozptyly; ty však nejsou v příkladu brány v potaz) mezi měřenými proměnnými.

3.3.4 Shoda modelu s daty

Předchozí odstavce ukazují princip, na jehož základě jsou v modelu odhadnuty parametry, následující text naznačuje, jak se vypočítá shoda modelu s daty (z angl. *model fit*), která

přináší údaj o tom, do jaké míry odpovídají odhadnuté parametry skutečným hodnotám v datech, resp. zkoumané populaci. Stejně jako v případě odhadu parametrů, jsou i zde jako vstupní informace využívány rozptyly měřených proměnných a kovariance mezi měřenými proměnnými. Počet kovariancí a rozptylů, které vstupují do analýzy, je zásadní vlastnost analyzovaných dat, od které se odvíjí možnosti modelování. K určení shody modelu s daty je nezbytné, aby v datech bylo více rozptylů a kovariancí než je v modelu parametrů. Výše uvedený příklad z Obrázku 3.1 obsahuje tři měřené proměnné; z tohoto počtu lze zjistit, kolik informací je k dispozici: tři korelace mezi proměnnými a tři rozptyly proměnných, celkem tedy šest údajů, které lze použít pro odhad parametrů a shody modelu s daty. Z nákresu v obrázku lze dále zjistit, jaké parametry je třeba odhadnout: rozptyl latentní proměnné F_i , tři faktorové zátěže mezi latentní proměnnou a měřenými proměnnými (b_{ij}) a tři zbytkové rozptyly měřených proměnných (e_{ij}), tedy celkem sedm parametrů. V modelu je v tuto chvíli o jeden parametr k odhadu víc, než je k dispozici informací, tj. korelací a rozptylů. V modelech SEM je vždy nutné určit měřítko či škálu latentní proměnné. Měřené proměnné škálu již mají, jsou jí např. hodnoty na pětibodové Likertově škále, kterou je proměnná měřena, ovšem latentní proměnné, které odpovídají abstraktním konstruktům, jsou bez měřítka. V modelu SEM získává latentní proměnná měřítko prostřednictvím měřené proměnné, jež je indikátorem konstruktů, kterému odpovídá latentní proměnná. Prakticky se to provádí tak, že se faktorová zátěž mezi měřenou a latentní proměnnou fixuje na hodnotu 1, čímž latentní proměnná získá stejné měřítko, jako měřená proměnná. Zároveň se tím ubere jeden volný parametr k odhadu, takže model v Obrázku 3.1 po fixaci jedné faktorové zátěže už má jen šest volných parametrů, takže je v něm shodné množství informací z dat (kovariancí a rozptylů) a parametrů k odhadu.

Počet rozptylů a kovariancí a parametrů k odhadu určuje, zda je model tzv. identifikovaný (z angl. *just identified*), tj. zda je množství informací z dat a počet parametrů k odhadu vyrovnané, nebo zda je nadidentifikovaný (z angl. *over – identified*), tj. zda je množství informací v datech vyšší než počet odhadovaných parametrů v modelu, případně zda je podidentifikovaný, tj. zda je množství informací v datech nižší než počet odhadovaných parametrů v modelu (českou terminologii přebírám z Matějů, 1989). Důležitou vlastností SEM modelů jsou proto tzv. stupně volnosti (z angl. *degrees of freedom*) a jejich počet. Ten se vypočítá tak, že se od počtu rozptylů a kovariancí v datech odečte počet parametrů, které je nutné odhadnout. Shodu modelu s daty lze určit jen v nadidentifikovaném modelu, tedy v modelu, který má kladný počet stupňů volnosti. V identifikovaném modelu lze

odhadnout parametry modelu, není však možné získat informace o shodě modelu s daty. Parametry modelu, který má záporný počet stupňů volnosti, tedy je podidentifikovaný, nelze vůbec odhadovat, protože neexistuje dostatečné množství informací k odhadu parametrů.

Z výše uvedeného je jasné, že je žádoucí, aby model měl kladný počet stupňů volnosti, tj. aby byl nadidentifikovaný, neboť jen za takové situace lze odhadnout parametry a zároveň zjistit shodu modelu s daty. Při výpočtu shody modelu s daty se stupně volnosti, tedy korelace a rozptyly, které jsou „navíc“, použijí pro „zkoušku“ správnosti odhadů parametrů. Tato zkouška spočívá v tom, že na základě parametrů, které byly vypočteny z rovnic, se vypočítávají hodnoty rozptylů či korelací, které nebyly v předchozím výpočtu použity pro výpočet odhadů (Saris, Gallhofer, 2014). Když se tyto odhadnuté korelace a rozptyly shodují s rozptyly a korelacemi ze souboru dat, má model výbornou shodu modelu s daty. Čím více stupňů volnosti model má, tím více lze provést „zkoušek“ správnosti odhadů parametrů a tím je shoda modelu s dat jistější.

Rozdíly mezi rozptyly a korelacemi z dat a rozptyly a korelacemi, které byly vypočteny z odhadnutých parametrů (z angl. *predicted correlations*), se nazývají rezidua. V analýze jde samozřejmě o to, aby rezidua byla co nejmenší. Výpočet v příkladu výše v tomto textu je velkým zjednodušením; vypočítávat parametry a shody modelu ručně nelze a v praxi se pro výpočty používá statistický software, který používá maticovou algebru a iterační výpočty, jež minimalizují předepsané funkce. Často používanými metodami odhadu jsou Unweighted least squares (ULS) a Weighted least squares (WLS) (Saris, Gallhofer, 2014), které používají rezidua mezi korelacemi ze souboru a odhadnutými korelacemi, které jsou funkcí parametrů. Metody se mezi sebou liší v tom, že ULS dává všem korelacím stejnou váhu, zatímco WLS jim dává různé váhy. V praxi často používaná metoda odhadu Maximum Likelihood (ML) je ve své podstatě WLS odhad se specifickými hodnotami pro tyto váhy (Saris, Gallhofer, 2014). Přesný popis toho, jak uvedené metody odhadují parametry a testové statistiky pro shodu modelu s daty, dalece přesahuje možnosti tohoto textu. Vážné zájemce o SEM a matematické funkce stojící za jeho principy odkazují na klasickou publikaci Bollen (1989), která obsahuje detailní popisy a rovnice.

3. 3. 5 Ukazatele shody modelu s daty a jejich doporučené hodnoty

Shodu modelu s daty vyjadřují statistiky a ukazatele, které jsou založeny na matematických funkcích naznačených v předchozím oddíle textu, a v analytické praxi jejich výpočet poskytuje statistický software, ve kterém modelujeme data. Tradičně se jako ukazatel shody modelu s daty používala signifikance testu χ^2 , jenž testuje nulovou hypotézu, která říká, že rozdíl mezi vloženou kovarianční maticí a vypočtenou kovarianční maticí je nula. Jinými slovy podává informaci o tom, zda se liší rozptyly a korelace (kovariance) v datech od odhadnutých rozptylů a korelací (kovariancí). V současné době se však χ^2 v sociálních vědách prakticky nepoužívá, neboť na velkých datových souborech je vždy signifikantní, tj. odmítá nulovou hypotézu o shodě matic a nepřináší skutečnou informaci o shodě modelu s daty. Statistika χ^2 se rovněž podle většiny autorů nehodí pro nenormálně rozložená data, což je opět komplikace právě pro analýzu sociálněvědních dat (Hooper, Coughlan, Mullen, 2008; Billiet, McClendon, 2000; Byrne, 2009).

Na základě nevhodnosti statistiky χ^2 byly vytvořeny ukazatele shody modelu s daty, které se pro data z výběrových šetření hodí výrazně více. Nejznámějším a nejpoužívanějším je tzv. Root Mean Square Error of Approximation (RMSEA), který patří mezi tzv. absolutní ukazatele shody (Kenny, 2015). Byrneová (2009) uvádí, že ukazatele RMSEA o hodnotách 0,05 resp. 0,08 reprezentují dobrou resp. dostatečnou shodu modelu s daty. V případě tohoto ukazatele platí, že čím nižší hodnota, tím lepší shoda. Podstatná je rovněž šíře intervalů spolehlivosti hodnot RMSEA; horní mez intervalu by ideálně neměla být vyšší než 0,08 (Hooper, Coughlan, Mullen, 2008). Další informaci o shodě modelu s daty přináší tzv. P value RMSEA, což je výsledek testu, který říká, zda je hodnota RMSEA rovná 0,05; hodnota signifikance testu by měla být nad 0,05, někteří autoři však uvádějí přísnější kritérium, a to 0,5 a vyšší (Joreskog, Sorbom, 1996). Ukazatelem, který patří do stejné skupiny jako RMSEA, je Standardized Root Mean Square Residual (SRMR), na jeho velikost se vztahují stejná kritéria jako na velikosti ukazatele RMSEA, ideálně by tedy měl být nižší než 0,08. Dalším často užívaným ukazatelem je Comparative Fit Index (CFI) a patří společně s dalšími, velmi podobnými ukazateli mezi tzv. inkrementální indexy, které mají interval (0;1) a pro které platí, že čím je hodnota bližší 1, tím lepší je shoda modelu s daty (Kenny, 2015). Hodnota CFI by se tedy měla ideálně pohybovat okolo hodnoty 0,95.

3. 3. 6 Specifikace SEM modelu

Důležitou roli při SEM analýze hraje tzv. fixace parametrů modelu. Když jsem v textu výše vysvětlovala, co jsou stupně volnosti modelu, uvedla jsem, že jeden z indikátorů latentní proměnné se standardně fixuje na hodnotu 1, čímž latentní proměnná získává měřítko. Tato fixace je nezbytná úprava, kterou je nutné činit z praktických důvodů, v analýze ale často implementujeme další fixace, abychom specifikovali vlastnosti modelu. Standardně je v modelu většina parametrů ponechána tzv. volná, tj. necháváme software odhadnout velikosti parametru. Pokud ale např. chceme, aby software s daným parametrem vůbec nepočítal, fixujeme jej na hodnotu 0. Pokud např. zafixujeme faktorovou zátěž mezi latentní proměnnou a jejím indikátorem na nulu, vyřadíme tento indikátor z indikátorů latentní proměnné. Tuto fixaci používáme také v případě, kdy nechceme, aby mezi sebou korelovaly zbytkové rozptyly měřených proměnných. Pomocí fixací můžeme také např. určit, aby všechny faktorové zátěže mezi latentní proměnnou a indikátory byly stejné. Možností fixací korelací, faktorových zátěží a rozptylů je více, než uvádím v tomto stručném popisu, obecně však platí, že se parametry ponechávají volné, fixují se na nulu nebo na hodnotu 1 a někdy je opodstatněné fixovat parametr na konkrétní hodnotu.

Fixování parametrů modelu je podstatnou součástí SEM analýzy a často se využívá, hledáme – li model, který nejlépe odpovídá datům. Přestože modelování je do velké míry konfirmativní analýzou, i zde je možné hledat různé varianty a vybírat z konkurenčních modelů ten nejlepší. V případech, kdy jsou modely srovnávány mezi sebou, je ve většině případů nutné porovnávat pouze tzv. vhnížděné modely (z angl. *nested model*) (Bollen, 1989). Model, který má nějaká omezení na volných parametrech, tj. některý z parametrů je fixován na nějakou hodnotu nebo na hodnotu jiného parametru, je vhnížděný identickému modelu, který omezení parametrů nemá (Bollen, 1989: 291). Další možnou definicí vhnížděného modelu pochází od Kennyho (2015), který píše, že model A je vhnížděný model modelu B, když model B je komplikovanější verze modelu A. Jako příklad uvádí model s jedním faktorem, který je vhnížděný modelu se dvěma faktory, neboť model s jedním faktorem může být považován za model se dvěma perfektně korelovanými faktory.

3. 2. 7 Problematické aspekty analýzy SEM

Analýza modelováním strukturálních rovnic má kromě nesporných výhod i svoji stinnou stránku, kterou jsou poměrně časté komplikace při výpočtu řešení. Aplikace konkrétního modelu na data ideálně končí tzv. řádným konvergujícím řešením (z angl. *proper solution*), které dodá odhady všech parametrů v modelu a ukazatele shody modelu s daty. Často se

však stává, že software vydá konvergující řešení s tzv. Heywoodovými případy (z angl. *Heywood cases*, HC) nebo že se mu vůbec nepodaří k řešení dojít a model je tzv. nekonvergující (z angl. *non-convergence*).

Heywoodovy případy jsou stručně řečeno nesmyslné odhady parametrů a může jimi být rozptyl proměnné menší než 0 nebo korelace větší než ± 1 (Kolenikov, Bollen, 2008). Pokud je mezi odhadnutými parametry HC, nedoporučuje se s výsledky dále standardně pracovat. Můžeme se pokusit zbavit se těchto hodnot tím, že re – specifikujeme model nebo data převedeme na jejich přirozený logaritmus, a doufat, že software odhadne model bez HC. Pokud se mu to ale ani tak nepodaří, musíme se smířit s tím, že se řádných výsledků nedobereme. Druhým závažným problémem je nekonvergence modelu, tedy situace, kdy software oznámí, že nemůže vypočítat parametry modelu a dodat ukazatele shody modelu s daty. Možností je opět re – specifikace modelu, případně logaritmování dat, pokud ale ani to nezabere, jsme opět bezmocní.

Důvodem výše popsaných problémů bývají velikosti kovariancí mezi proměnnými v kovarianční matici, jež vstupuje do analýzy, nebo složitost modelu, dále mají svůj vliv nízké hodnoty faktorových zátěží či korelací a regresí mezi latentními proměnnými, případně je na vině velikost analyzovaného souboru. Problematice nekonvergence SEM modelů a výskytu HC se různí autoři a autorky věnují často (Revilla, Saris, 2013a; Revilla, Saris, 2011a, b; Bentler, Chou, 1987; Ding, Velicer, Harlow, 1995), obecně se však tento problém nedaří řešit.

3. 3. 8 Software pro strukturální modelování

Pro SEM byly vytvořeny speciální softwary a v současné době jich je k dispozici několik. V 70. a 80. letech minulého století se používaly programy LISREL a EQS, později je nahradily uživatelsky přátelštější komerční produkty jako AMOS nebo Mplus. Pořízení uvedených softwarů je však velmi finančně náročné a pro běžné studující nereálné. Existují ovšem plnohodnotné alternativy drahých SEM softwarů, které operují v rámci softwarového prostředí R, a to lavaan a sem.

Pro vypracování analýz v této dizertační práci jsem použila dva programy, LISREL (Jöreskog, Sörbom, 2004) a lavaan (Rosseel, 2012). LISREL je nejstarší SEM software a v současné době je mezi běžnými uživateli poměrně neoblíbený, a to kvůli složité práci se syntaxí, která používá maticové značení parametrů. Pro analýzu odchylky souhlasného stylu odpovídání v kapitole 5 jsem použila R package lavaan. Tento software je volně

dostupný zdarma, přesto s ním lze provádět většinu běžných analýz jako s drahými statistickými pakety pro modelování strukturálních rovnic. lavaan dokonce umí simulovat výsledky podle algoritmu různých komerčních softwarů (Rosseel, 2012). Navzdory tomu, že studenti i výzkumníci mají ze softwaru operujícího v rámci R často obavy, je lavaan velmi uživatelsky přátelský a jeho syntaxe je velmi intuitivní. Program LISREL jsem použila pro analýzu 2 SB MTMM dat, neboť v něm lze definovat tzv. TS MTMM model, což ve verzi lavaanu, která byla aktuální v době, kdy jsem analýzy prováděla, nebylo možné. Pracovala jsem s ním rovněž v analýze kvazisimplexovým modelem.

Ze zkušeností, které jsem nabyla při práci s oběma softwary, jsem došla k závěru, že LISREL častěji než lavaan nedokáže najít konvergující řešení. Tento můj poznatek však může být dán tím, že verze LISRELu, kterou používám, je výrazně starší, než verze lavaanu, který se často automaticky updatuje.

4. Kvazisimplexový model pro odhad reliability měření: Odhad reliability vybraných položek z českých šetření EU SILC

Vysoká reliability měření dotazníkových položek ve výběrových šetřeních je podmínkou přijatelné kvality získaných dat. A pouze kvalitní data umožňují sociálněvědní statistickou analýzu, jejíž závěry jsou důvěryhodné a odrážejí sociální realitu. Nabízí se tak doporučení, aby výzkumnice a výzkumníci měli informace o tom, jak reliabilní jsou data, která používají. V běžné výzkumné praxi se ovšem informace o reliability měřených položek zpravidla nedají získat, neboť ve výzkumu provádíme jen substanciální měření daného konstruktů, nikoliv experimentální, opakovaná měření, a toto jedno měření pro zjištění reliability nestačí. Každá metoda zjišťování reliability vyžaduje opakované měření (test – retest) či speciální design s opakovaným měřením (Multitrait – multimethod design (MTMM) analyzovaný jako tzv. True Score MTMM model). V této kapitole prezentovaná metoda kvazisimplexového modelu (z angl. *quasi simplex model*; QSM) se od výše uvedených metod liší. Ne však v tom, že by si vystačila s jedním jediným měřením, ale tím, že využívá opakovaná měření získaná v panelovém šetření. Pracuje tedy s opakovanými měřeními, mezi kterými uběhl delší časový interval a která nebyla získána prostřednictvím speciálního designu, který šel nad rámec potřeb získání substanciálních informací o zkoumaném fenoménu. QSM je proto relativně více přístupná metoda zjišťování reliability měření dotazníkové položky než ostatní metody, neboť panelová data jsou analytikům a analytičkám k dispozici častěji než data z metodologických experimentů, které bývají prováděny relativně vzácně. Pomocí této metody je možné zjistit reliability měření dotazníkových položek, pro které bychom tento ukazatel kvality dat bez provedení dalšího sběru dat nemohli získat. Zjištění reliability prostřednictvím QSM je prostě „něco navíc“, pro co nemusíme vynakládat úsilí při designu výzkumu substanciálních sociálněvědních témat.

QSM vyžaduje měření minimálně ve třech, resp. čtyřech vlnách a data analyzuje v softwaru pro modelování strukturálních rovnic (Alwin, 2007). Samotný model a vztahy mezi měřenou proměnnou a jejím pravým skórem se řídí předpisem Klasické testové teorie. Cílem výkladu v této kapitole je vysvětlit princip kvazisimplexového modelu, ukázat, jak jej lze použít k získání odhadů reliability a prezentovat odhady reliability pro reálná data, která čeští sociologové a socioložky používají. V empirické části této kapitoly metodu QSM aplikuji na data z českých šetření EU SILC – European Union Statistics on

Income and Living Conditions –, jednoho z mála zdrojů panelových dat v České republice. Charakter dat EU SILC umožnil provést analýzu kvazisimplexovým modelem na třech položkách; na položce PH010, která dotazuje subjektivně pocíťovaný zdravotní stav, položce HS120, jež zjišťuje hodnocení finanční situace domácnosti, a položce HS130, která respondenty vyzývá k odhadu nejnižšího možného příjmu, který by stačil na provoz domácnosti.

Vzhledem k tomu, že data EU SILC české analytičky a analytici používají často a že jsou tato data využívána pro analýzu závažných sociálních problémů, jako je např. chudoba, domnívám se, že je znalost o jejich reliabilitě potřebná. Odhady reliability měření položek prezentované v této práci mohou analytičky a analytici využít jako zdroj informací o kvalitě dat položek EU SILC nebo je mohou přímo použít v analýze k očištění dat o rozptyly náhodné chyby. Popis toho, jak lze data očistit o chybový rozptyl, poskytuje sedmá kapitola této dizertační práce.

Cílem tohoto textu je především informovat o reliabilitě vybraných položek dotazníku EU SILC, nezbytnou součástí výkladu je ale i popis metody kvazisimplexového modelu, shrnutí teorie, která za touto metodou stojí, a přehled kritických připomínek vztahujících se k metodě. V české metodologické literatuře nebyl kvazisimplexový model dosud představen; velmi zběžně se o něm zmiňuje Řehák (1998b), popis a ukázka použití však chybí, přestože metoda zažila svůj rozkvět již v sedmdesátých letech dvacátého století. Uvedení této v zahraniční odborné literatuře rozšířené metody do českého prostředí tedy považuji za dílčí přínos tohoto textu.

4. 1 Kvazisimplexový model (QSM) a srovnání s metodou test – retest

Kvazisimplexový model představil v roce 1969 psycholog David Heise (Heise, 1969). Princip QSM měl vyvést z problémů známou metodu test – retest (Lord, Novick 1968), tedy dvakrát opakované měření té samé dotazníkové položky. Opakovaným měřením v metodě test – retest je myšleno, že respondent je dotázán na tu samou otázku pokud možno tím samým tazatelem dvakrát za sebou v relativně krátkém časovém rozmezí. Princip metody test – retest, stejně jako QSM, stojí na vztazích vyplývajících z Klasické testové teorie (Lord, Novick, 1968), popsané ve druhé kapitole této práce. V metodě test – retest je za reliability měření dané dotazníkové položky považována korelace dvou opakovaných měření. Pro odečítání reliability z modelu strukturálních rovnic, což je případ

kvazisimplexového modelu, platí, že reliabilita X je mocnina faktorové zátěže mezi měřenou proměnnou X a jejím pravým skórem T , tedy, že $rel X = r^2(X, T)$.

Metoda test – retest má problémy zásadního charakteru, na které jsem upozornila již v kapitole 2. V tomto bodě výkladu část z nich zopakuji, neboť upozornění na tyto problémy dobře slouží při popisování principu QSM. Zprvée, u metody test – retest nelze vyloučit, že respondenti, kteří na tu samou otázku odpovídají dvakrát, si svoji první odpověď pamatují. V případě, kdy je druhá odpověď závislá na té první, je porušen předpoklad o nezávislosti jednotlivých měření a zjištěná reliabilita je nadhodnocená. Tento tzv. efekt paměti lze eliminovat tím, že se prodlouží časový interval, který uběhne mezi prvním a druhým dotazováním. V takovém případě se však objevuje druhý problém, a to možná změna v hodnotě měřeného konstruktů, a tím pádem v hodnotě pravého skóru. Respondent např. může během času změnit názor, na který se dotazníková položka ptá. Reliabilita měření pak bude odhadnuta nižší, než jaká by byla, kdyby respondenti neměli čas svůj názor změnit. Metoda test – retest se neumí vyhnout oběma těmto problémům najednou; vždy hrozí, že měření reliability bude nepřesné buď z prvního, nebo z druhého uvedeného důvodu.

Kvazisimplexový model řeší nedostatky metody test – retest tím, že teoreticky umožňuje změnu v hodnotě pravého skóru mezi jednotlivými měřeními. Je tak možné prodloužit interval mezi jednotlivými měřeními na tak dlouhou dobu, aby byl vyloučen efekt paměti, a proto lze QSM aplikovat na panelová data, kde je interval mezi měřeními dlouhý. Kvazisimplexový model předpokládá, že proces změny hodnot pravých skóru v čase odpovídá tzv. Markovovým řetězcům, tedy že pravý skór v čase i může být přímo ovlivněn pravým skórem v čase $i-1$, ale nikoli pravými skóry v čase $i-2$ a staršími – jejich vliv může být pouze zprostředkovaný hodnotou v čase $i-1$ (Alwin, 2007). Očekávaná změna mezi pravými skóry tedy vyplývá ze stochastického modelu, nikoliv z nějakého empiricky pozorovaného trendu.

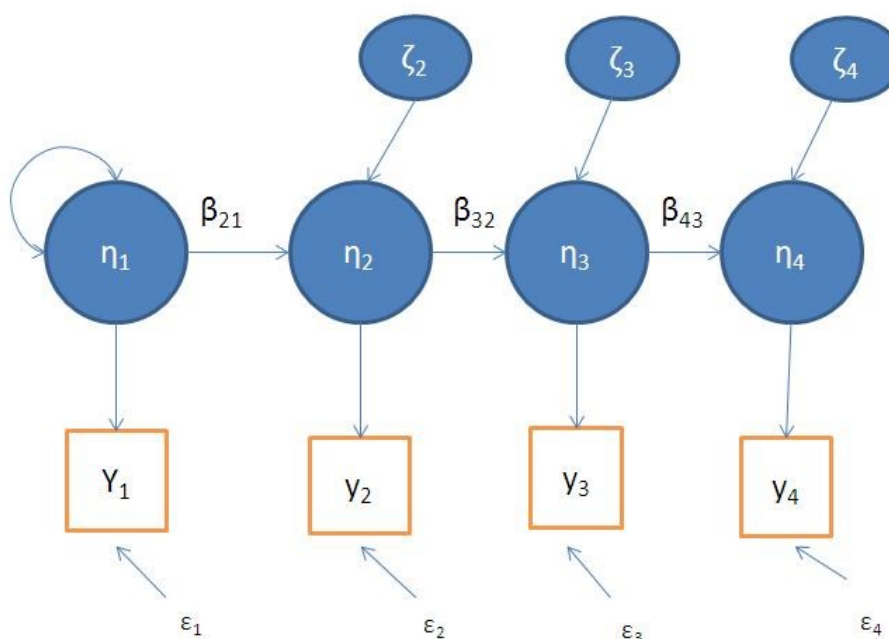
4. 2 Použití kvazisimplexového modelu

Až dosud jsem psala, že metoda QSM vyžaduje měření minimálně ve třech, resp. čtyřech vlnách. Tyto dva rozdílné údaje jsem uváděla proto, že existuje rozdíl v tom, když používáme data ze tří, resp. čtyř a více vln; aplikace kvazisimplexového modelu na data ze tří vln je sice možná, ale významně omezená. Pro data ze tří vln není možné prostřednictvím modelování strukturálních rovnic zjistit ukazatele shody modelu s daty, a

tím ověřit, zda jsou získané odhady reliability spolehlivé. Kvazisimplexový model pro data ze tří vln je identifikovaný, tzn. jeho počet stupňů volnosti je nula. Aby bylo možné testovat shodu modelu s daty, je potřeba zvýšit počet stupňů volnosti minimálně o jeden tím, že se do analýzy zahrnou data z další vlny šetření⁷. Teprve pak je možné získat ukazatele shody modelu s daty a stanovit, nakolik lze odhadnutým parametrům důvěřovat.

Z důvodu konzistence výkladu a prezentované analýzy popisují kvazisimplexový model v tomto textu na příkladu se čtyřmi vlnami měření jedné dotazníkové položky, nikoliv na běžně prezentovaném obecném modelu se třemi vlnami. Vztahy mezi měřeními proměnnými v čase i a jim náležejícími pravými skóry ilustruje Obrázek 4. 1.

Obrázek 4. 1 Kvazisimplexový model se čtyřmi indikátory



Pozn. Obrázek 4.1 používá klasické maticové značení používané v SEM analýze, znaky η_i jsou latentní proměnné pravého skóru v čase i ($\eta_1, \eta_2, \eta_3, \eta_4$), $\beta_{i, i-1}$ jsou tzv. stability ($\beta_{21}, \beta_{32}, \beta_{43}$), ζ_i reprezentuje zbytkový rozptyl latentních proměnných pravého skóru η_i , ($\zeta_2, \zeta_3, \zeta_4$); ε_i je rozptyl měřené proměnné y_i odpovídající náhodné chybě měření ($\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$).

⁷ Přidáním další, v pořadí čtvrté vlny, se počet stupňů volnosti modelu zvýší na dva.

4.3 Popis kvazisimplexového modelu

V souladu s Klasickou testovou teorií (viz rovnice 2.1 a 2.6) rozděluje QSM rozptyl proměnných y_i měřených v čase i do dvou částí: jedna část y_i odpovídá latentní proměnné pravého skóru (η_i), druhá náhodné chybě měření (ε_i). V Obrázku 4.1 reflektují latentní proměnné pravého skóru η_2 , η_3 a η_4 vliv, který reprezentuje možnou změnu v hodnotě pravého skóru, ke které mohlo dojít v čase, který uplynul mezi vlnami dotazování. Tento vliv je reprezentován symbolem ζ_i a jedná se o zbytkový rozptyl neboli náhodnou chybu latentní proměnné. Zbytkový rozptyl prvního latentního skóru η_1 je fixován na hodnotu 1, neboť tato latentní proměnná je ve strukturální části modelu exogenní a tedy nemá v čase předcházející pravý skór, jenž by ji mohl ovlivňovat. Faktorové zátěže mezi pravým skórem η_i a měřenou proměnnou y_i jsou koeficienty reliability a jejich hodnota umocněná na druhou odpovídá reliabilitě měření. V rovnici níže jejich vliv reprezentuje znak Λ_i . Regresní koeficienty $\beta_{i,i-1}$ mezi latentními proměnnými η_i odpovídají koeficientu tzv. stability a ukazují, jak stabilní je měřená charakteristika napříč časem.

Rovnice (4.1) vyjadřující část měření kvazisimplexového modelu (podíly rozptylu měřených proměnných) a (4.2) strukturální část kvazisimplexového modelu (podíly rozptylu latentních proměnných) jsou:

$$(4.1) \quad y_i = \Lambda_i \eta_i + \varepsilon_i,$$

$$(4.2) \quad \eta_i = \beta_{i,i-1} \eta_{i-1} + \zeta_i$$

kde i značí vlnu, resp. čas měření.

Aby bylo možné s kvazisimplexovým modelem analyticky pracovat v softwaru pro SEM, je nutné na některé parametry uplatnit restriktce. Ty zohledňují teoretické předpoklady kvazisimplexového modelu a zároveň jsou nezbytným opatřením, bez kterého nelze model statisticky testovat. Heise (1969) stanovuje, že reliability v modelu, tedy faktorové zátěže mezi y_i a η_i , musí být fixovány tak, aby byly ekvivalentní. Naproti tomu model podle Wileyho a Wileyho (1970) stanovuje podmínku, že ekvivalentní musí být náhodné chyby ε_i . Obě strategie vedou ke stejnému výsledku, pokud se rozptyly měřených proměnných y_i shodují napříč vlnami (Cernat et al, 2014). Existují argumenty pro i proti používání jednoho i druhého modelu, v odborné literatuře ovšem bývá častěji uplatňován model podle Wileyho a Wileyho (Achen, 1975; Green, Palmquist, 1990; Van der Veld, Saris,

2005) nebo jsou použity obě varianty (Erikson, 1979; Jagodzinski, Kuhnel, 1987; Alwin, 2007, Cernat et al., 2014).

Výsledkem SEM analýzy panelových dat s využitím kvazisimplexového modelu jsou odhady parametrů, ze kterých jsou odečteny stability, resp. reliability měření. Metoda QSM byla použita v desítkách odborných prací. Analytičky a analytici s ní pracují buď jako s nástrojem pro zjištění reliability používaných dotazníkových položek (Achen, 1975; Alwin, 2007; Alwin, Krosnick, 1991; Biemer, Christ, Wiesen, 2009; Blok, Saris, 1983; Erikson, 1979; Ehrhardt, Saris, Veenhoven, 2000; Jagodzinski, Kuhnel, 1987; Joreskog, 1970; Lucas, Donnellan, 2011; Van der Veld, Saris, 2005; Wiley, Wiley, 1970; Heise, 1969), nebo jej používají jako prostředek k substantiální analýze, ve které se zaměřují na změnu v hodnotě zkoumaného konstruktů (Alwin, Krosnick, 1989; Judd, Milburn, 1980; Kenny, Campbell, 1989; Green, Pamlquist, 1990; Schilling, 2006).

4. 4 Kritika kvazisimplexového modelu

Jako kterákoliv jiná metoda odhadu reliability vycházející z Klasické testové teorie má i QSM problematické aspekty, které je nutné zohlednit. Hlavní problémy metody odhadu reliability test – retest byly již v předcházejících částech textu uvedeny, v následujícím oddílu shrnuji kritické připomínky, které vznesli někteří autoři a autorky právě k použití kvazisimplexového modelu.

Saris a Gallhoferová (2014) identifikují dva zásadní problémy QSM přístupu. Zaprvé kritizují předpoklad Markovova řetězce, tedy kauzálního řetězce, podle kterého je pravý skóre v čase i ovlivněn výhradně latentním skórem v čase $i-1$ a nikoliv např. pravým skórem v čase $i-2$ nebo $i-3$. Podle nich nejsou v realitě neobvyklé situace, kdy je pravý skóre v čase i ovlivněn např. pravým skórem v čase $i-2$ a nikoliv pravým skórem $i-1$. Takové případy však kvazisimplexový model nedokáže rozeznat a reliabilitu odhadne špatně. Zadruhé autoři tvrdí, že v datech pro každou vlnu existuje část rozptylu měřené proměnné, která je specifická výhradně pro tuto vlnu šetření, a která se již nemusí objevit v následující vlně. Kvazisimplexový model však tento rozptyl nerozpozná jako substantiálně významnou část měřené proměnné, ale jako náhodnou chybu. Autoři udávají příklad s měřením kvality života a jeho ovlivněním aktuální náladou dotazovaného; např. respondentka, která bude mít při prvním dotazování dobrou náladu, odpoví jinak, než ta samá respondentka odpovídající při druhém dotazování ve špatné náladě. Rozptyl měřené proměnné reflektující aktuální náladu ale kvazisimplexový model rozpozná jako náhodnou chybu a

odhad reliability bude v tomto případě nižší, než jaká je skutečná reliabilita. Z výše uvedených důvodů autoři zavrhnou QSM jako nevhodnou metodu měření reliability a razantně upřednostňují vlastní metodu 2 split – ballot MTMM, jejíž popis a aplikaci prezentují v šesté kapitole této dizertační práce.

První připomínka Sarise a Gallhoferové je určitě opodstatněná a poukazuje na nejslabší místo metody QSM. Předpoklad, že pravý skóre v čase i je ovlivněn výhradně pravým skórem v čase $i-1$ je zjednodušením. Bez tohoto zjednodušení však nelze panelová data prostřednictvím QSM metody za standardních okolností modelovat v SEM; kvazisimplexový model, ve kterém by byl pravý skóre i ovlivněn dvěma a více předcházejícími pravými skóry, by měl záporný počet stupňů volnosti. Pro názornost uvedu příklad na modelu se čtyřmi pravými skóry, znázorněném v Obrázku 4.1 a použitým pro analýzu v tomto článku. Tento model s nezbytnými restrikcemi má dva stupně volnosti; aby byla zachována bezpodmínečně nutná možnost testování shody modelu s daty, je dovoleno ukrojit z počtu dvou stupňů volnosti maximálně jeden, tj. přidat do modelu jen jeden další parametr k odhadu. Situace, kdy by byla čtyři měření ve vlnách provázána v čase, by ale vyžadovala specifikaci dalších tří parametrů, čímž by se model stal neidentifikovaným ($df = -1$). Z toho vyplývá, že situaci s kauzálně provázanými pravými skóry nelze statisticky testovat. Výtka Sarise a Gallhoferové je tak spíše připomenutím omezení, které metoda má, nikoliv výzvou k testování jiného kvazisimplexového modelu, ve kterém by byly všechny pravé skóry provázané v čase. Vzhledem k tomu, že je v modelu se čtyřmi vlnami k dispozici jeden stupeň volnosti, nabízí se, aby byl jeden parametr β umístěn někde mezi pravé skóry. Pro takový postup však neexistuje teoretické opodstatnění. Neexistuje důvod, proč přidat další parametr β jen mezi η_1 a η_3 nebo jen mezi η_2 a η_4 nebo jen mezi η_1 a η_4 . Z teoretického hlediska by byl takový krok prostě jen náhodným zkoušením či tipováním, což je strategie, které by se analytici a analytičky měli vyhýbat.

Druhou kritickou připomínku Sarise a Gallhoferové ve své publikaci podrobněji rozebírají van der Veld a Saris (2005), kteří se rovněž domnívají, že reliabilita z QSM je z části odhad reliability dotazníkové položky, ale zároveň reflektuje i další vliv, který je specifický pouze pro konkrétní panelovou vlnu. Van der Veld a Saris tvrdí, že časově specifický rozptyl kvazisimplexový model identifikuje jako náhodnou chybu, a proto je reliabilita získaná jeho prostřednictvím podhodnocená, tedy nižší než jaká ve skutečnosti je. Existenci tohoto časově specifického rozptylu pomocí QSM je možné zkoumat pouze

s využitím dat ze speciálního designu; běžná panelová data totiž neobsahují dostatečné množství informací, aby bylo možné identifikovat časovou složku rozptylu měřené proměnné. Autoři studie měli k dispozici data ze speciálního designu, ve kterém byla v každé vlně panelového šetření položka měřena dvakrát, tedy klasickým test – retestem. Výzkumníci data analyzovali v modelu, kde byly pro každou vlnu k dispozici dva indikátory, a ukázali, že v jejich datech existoval časově specifický rozptyl. Bohužel ale měli k dispozici data pouze ze tří vln šetření, takže důkaz o shodě jejich speciálního kvazisimplexového modelu s daty nemohli dodat.

Kromě této analýzy jejich studie navíc obsahovala porovnání odhadů reliability z test – retest metody a klasického kvazisimplexového modelu. Odhady z obou metod analytici použili k očištění dat od chybových rozptylů a ukázali, že korelace položek, které byly očištěny odhady reliability z QSM, jsou vyšší než 1, což je teoreticky nemožné. Na základě tohoto poznatku usoudili, že odhady reliabilit postojových položek z QSM jsou podhodnocené a doporučují metodu test – retest jako spolehlivější nástroj. K jejich závěru je však nutné dodat, že efekt paměti v test – retest experimentu má za následek odhadnutí vyšší reliability než jaká ve skutečnosti je, což je skutečnost popsána v každém textu o metodě test – retest. Pokud si respondent pamatuje svoji první odpověď a při druhém dotazování ji pouze zopakuje, jsou korelace mezi dvěma měřeními vysoké, a proto je reliabilita odhadnuta vyšší. Naproti tomu v QSM je efekt paměti vyloučen, pokud je doba mezi vlnami v panelu dostatečně dlouhá. I proto vycházejí odhady reliability v QSM obecně nižší než v test – retest datech, jak píší ve stati sami autoři. I v tomto bodě je potřeba připomenout, že autoři nemohli kvazisimplexový model, ze kterého odečetli odhady reliability, statisticky testovat, neboť měli k dispozici jen tři vlny měření.

Biemer, Christová a Wiesen (2009) upozorňují na to, že restriktce týkající se fixace parametrů v kvazisimplexovém modelu jsou sázkou do loterie. Jak už bylo uvedeno, podle Wileyea a Wileyea (1970) jsou fixovány rozptyly náhodných chyb jako ekvivalentní, Heise (1969) naopak doporučuje fixovat jako ekvivalentní reliability. Jedna nebo druhá strategie musí být v SEM analýze použita, neboť bez fixací má kvazisimplexový model záporný počet stupňů volnosti a nelze odhadnout parametry v modelu a testovat shodu modelu s daty. Podle autorů může v datech existovat stav, kdy rozptyly náhodných chyb nebo reliability ekvivalentní nejsou, což model nikdy neodhalí. Problém s utlačivou silou restrikcí autoři řeší zavedením tzv. Generalized Simplex modelu, ten je však použitelný výhradně pro data z více položkových škál. Pokud je testována reliabilita jedné položky,

jako je tomu např. v empirické části této studie, nelze Biemerem a jeho kolegy navržený přístup uplatnit. Autoři práce dále upozorňují na nebezpečí toho, že náhodné chyby mohou v datech korelovat. Tato hrozba se však týká pouze dat z panelových šetření, kde byla časová mezera mezi vlnami příliš krátká na to, aby mohli respondenti svoji předchozí odpověď spolehlivě zapomenout.

Shrnutí kritických připomínek k přístupu QSM lze uzavřít tvrzením, že vědecká komunita není ve věci důvěryhodnosti odhadů z kvazisimplexových modelů celistvá. Existují nadšení podporovatelé přístupu (Alwin, Achen), jiní výzkumníci však považují kritické připomínky za tak závažné, že doporučují QSM vůbec nepoužívat. Mezi nimi je nejhlasitějším odpůrcem přístupu Willem Saris, který vkládá naděje do svého designu 2 split – ballot MTMM (2 SB MTMM), jenž je dlouhodobě používán v metodologických šetřeních European Social Survey a s jehož využitím vzniká obrovská databáze údajů o reliabilitě a validitě dotazníkových položek nazvaná SQP. Detailní popis tohoto designu je obsažen v textu kapitoly 6, v tomto kontextu jen zmíním, že realizace šetření s využitím 2 SB MTMM je organizačně obtížná a stojí více peněz než běžné výběrové šetření. Navíc stejně jako jiné metody opakovaného měření je i 2 SB MTMM problematický; neumí zajistit eliminaci efektu paměti a data, která jsou s jeho využitím získána, jsou obtížně analyzovatelná a často se pro ně nepodaří najít řádně konvergující řešení v SEM analýze (Revilla, Saris 2013a). 2 SB MTMM je tedy navzdory Sarisově doporučení, stejně jako QSM, metodou, která má problémy, jež nelze eliminovat.

Závěrem teoretického popisu metody QSM lze na její obhajobu uvést, že stoprocentní jistotu o reliabilitě sociálněvědních dat neumí podat žádná experimentálně – analytická metoda a všechny známé metody odhadu reliability mají své slabiny. V případě, kdy jsou data analyzována v modelu strukturálních rovnic, jak je tomu v případě QSM, ale třeba i 2 SB MTMM, jsou indikátorem vhodnosti použití metody ukazatele shody modelu s daty. Této zásady jsem se držela v následující analýze, ve které jsem jako důvěryhodné přijala pouze výsledky z modelů, které měly vynikající nebo dobré ukazatele shody modelu s daty.

4. 5 Empirická část: Odhad reliability českých dat EU SILC kvazisimplexovým modelem

4. 5. 1 Zdroje českých dat k analýze kvazisimplexovým modelem

V České republice existuje omezené množství panelových dat vhodných pro analýzu kvazisimplexovým modelem. V současné době jsou k dispozici pouze dva zdroje panelových dat, která obsahují minimálně čtyři vlny; jsou jimi česká šetření EU SILC a data z projektu ELSPAC – Evropské dlouhodobé studie těhotenství a dětství. První uvedený zdroj disponuje daty reprezentativními pro populaci dospělých občanů České republiky a používám jej pro analýzu v tomto článku. Aktuální nedostatek delších panelových řad v České republice je podle všeho dočasnou situací; vzhledem k rozvoji kvantitativního výzkumu v české sociologii se dá v budoucnosti očekávat, že panelová šetření s větším množstvím vln bude možné využívat. Příslibem je např. projekt SHARE, který v současné době nabízí srovnatelná data pro tři vlny⁸. Vzhledem k tomu, že v létě 2015 začal v České republice sběr pro další vlnu, budou data pro čtyřletou řadu k dispozici nejpozději do dvou let⁹. V roce 2015 bylo rovněž zahájeno velké panelové šetření nazvané Proměny české společnosti, v jehož rámci bude proveden sběr dat ve čtyřech vlnách v letech 2015–2018¹⁰.

4. 5. 2 Šetření European Union Statistics on Income and Living Conditions (EU SILC)

Projekt European Union Statistics on Income and Living Conditions (EU SILC) sbírá průřezová a longitudinální data o populaci národních států v Evropské unii od roku 2004; v České republice šetření provádí Český statistický úřad (ČSÚ) pod názvem Životní podmínky od roku 2005. V rámci EU SILC jsou shromažďovány informace o příjmech domácností, chudobě, sociální exkluzi a životních podmínkách obyvatel. Respondenti odpovídají na otázky týkající se na podmínky v bytě, kde bydlí, na náklady na bydlení, na vybavení domácnosti, na všechny příjmy a výdaje domácnosti, na pracovní aktivitu nebo na zdraví a vzdělání. Praktické shrnutí informací o šetření EU SILC v České republice poskytuje článek Martiny Mysíkové (2011).

⁸ Data ze třetí vlny šetření nazvaného SHARE LIFE nejsou srovnatelná s daty z ostatních vln šetření.

⁹ Zdroj: Česká webová stránka projektu, <http://share.cerge-ei.cz>

¹⁰ Zdroj: Webová stránka projektu, <http://www.promenyceskespolecnosti.cz>

Longitudinální část šetření EU SILC je realizovaná ve čtyřletém rotačním panelu. EU SILC má díky tomuto designu k dispozici nejen poměrně velký soubor longitudinálních dat, ale i velmi obsáhlý soubor průřezových dat. Datový soubor pro každý jednotlivý rok totiž obsahuje průřezová data ze čtyř panelových podsouborů (viz tabulka 4.1, údaje ve sloupcích; v tabulce je vidět, že první průřezový soubor obsahující čtyři panelové podsoubory vznikl až v roce 2008). Vzhledem k tomu, že pro analýzu v tomto článku používám panelová data, nikoliv průřezová, podrobněji vysvětluji princip rotačního panelu. Respondenti vybraní pro longitudinální výzkum EU SILC se šetření účastní po dobu čtyř let, dotazováni jsou každý rok jednou. Nový rotační panel je zahajován každý rok; každý rok jsou přibíráni respondenti, kteří v prvním roce vytvoří první vlnu panelového šetření, v následujícím roce druhou atd. Po čtyřech letech šetření zcela opustí. Princip účasti respondentů v rotačním panelu ilustruje Tabulka 4.1, rotační panely jsou vyznačeny v řádcích.

Tab. 4.1 Složení rotačního panelu EU SILC a průřezových souborů, data ČR

| Rok | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|--------------------|------|------|------|------|------|------|------|------|
| Účast v daném roce | X | X | X | X | | | | |
| | | X | X | X | X | | | |
| | | | X | X | X | X | | |
| | | | | X | X | X | X | |
| | | | | | X | X | X | X |

V současné době je pro Českou republiku k dispozici již pět řad čtyřletých panelových dat (v době psaní toho textu jsou k dispozici data z let 2005 – 2012). Čtyřleté panelové řady použité pro studii v tomto textu jsou tyto:

- Panel z let 2005 až 2008
- Panel z let 2006 až 2009
- Panel z let 2007 až 2010
- Panel z let 2008 až 2011
- Panel z let 2009 až 2012

Pro tuto analýzu je existence většího množství čtyřletých panelů v šetřeních EU SILC velkou výhodou, neboť umožňuje testovat QSM pro jednu a tu samou dotazníkovou

položku opakovaně. Vzniká tím možnost porovnání velikosti odhadů reliability z jednotlivých panelů, čímž se v případě jejich větší či menší shody zvyšuje jistota ohledně důvěryhodnosti odhadů.

4. 5. 3 Položky z dotazníků EU SILC vhodné ke QSM analýze

Položky dotazníku EU SILC nejsou typická sociologická data. Drtivá většina těchto položek požaduje výpověď o nějakém faktu, týkajícím se hospodaření domácnosti, jako je např. výpověď o příjmu, pobírání sociálních dávek nebo vybavení bytu, a velká část šetřených proměnných má nominální povahu. Šetření EU SILC má tři různé dotazníky (A, B, C), které respondenti s tazateli během jednoho výzkumného rozhovoru vyplňují. Já jsem vybírala položky z dotazníků pro domácnost (B) a pro jednotlivce (C)¹¹. Výběr položek, které by bylo možné analyzovat prostřednictvím kvazisimplexového modelu, nebyl úplně jednoduchý, neboť data vhodná pro QSM analýzu musí splňovat jisté předpoklady. Prvním z nich je intervalový či ordinální charakter dat¹², dalším je požadavek, aby formulace otázky byla napříč všemi vlnami v panelu naprosto stejná. Dále jsem mohla vybrat jen ty dotazníkové položky, kterým v datových souborech distribuovaných ČSÚ odpovídala jen jedna proměnná. V některých případech jsou totiž v souborech dat proměnné agregované z několika dotazníkových položek a data pro jednotlivou položku nejsou k dispozici (právě z tohoto důvodu byly z analýzy vyloučeny všechny položky týkající se příjmů). Všem výše uvedeným požadavkům odpovídaly jen dvě položky z dotazníku B a tři položky z dotazníku C. Znění těchto položek obsahuje Tabulka 4.2.

Z pěti položek uvedených v Tabulce 4.2 jsem musela vyřadit ještě dvě. Položky PL190 a PL200 podléhají kontrole tazatelů, takže v případě, kdy respondent odpoví jinak než v předchozím dotazování (PL190) nebo jeho odpověď následující rok logicky neodpovídá odpovědi v minulém roce (PL200), tazatel respondenta upozorní na nesrovnalost a vyplní údaj správně. Děje se tak v případě CAPI i PAPI modu sběru dat (Mysíková, 2011)¹³. Data pro položky PL190 a PL200 tedy nereflektovala skutečnou odpověď respondenta, a proto nebyla pro analýzu reliability vhodná. Abych čtenářkám a čtenářům usnadnila orientaci

¹¹ Položky v dotazníku A byly pro analýzu nevhodné.

¹² Princip QSM umožňuje analyzovat i nominální data, v takovém případě je však nutné použít tzv. analýzu latentních tříd (z angl. Latent class analysis), nikoliv v tomto článku používané modelování strukturálních rovnic (SEM).

¹³ Tuto skutečnost mi Martina Mysíková potvrdila i osobně na základě informací, které jí dodaly pracovnice ČSÚ.

v textu, budu v následujícím textu nazývat položku PH010 *PH010 Zdraví*, položku HS120 *HS120 Příjem* a položku HS130 *HS130 Odhad příjmu*.

Tab. 4.2 Znění vybraných položek z dotazníků EU SILC

| Název proměnné v databázi EU SILC: | Znění otázky: | Metoda měření: |
|------------------------------------|--|---|
| HS120 | Jak vychází Vaše domácnost s celkovým měsíčním příjmem? | 1. s velkými obtížemi 2. s obtížemi 3. s menšími obtížemi 4. docela snadno 5. snadno 6. velmi snadno |
| HS130 | Jaký nejnižší možný čistý měsíční příjem by musela mít Vaše domácnost, aby s ním vyšla? Odpovězte prosím s přihlédnutím k současnému složení a podmínkám ve Vaší domácnosti. | |
| PL190 | Kdy jste nastoupil(a) do svého prvního řádného zaměstnání, popř. začal(a) podnikat? | |
| PL200 | Kolik let jste od té doby strávil(a) v práci? (<i>navazuje na položku výše, pozn.aut.</i>) | |
| PH010 | Jak celkově hodnotíte svůj zdravotní stav? | 1. velmi dobrý 2. dobrý 3. přijatelný 4. špatný 5. velmi špatný |

4. 5. 4 Výběr případů ke QSM analýze

Základním požadavkem QSM analýzy reliability je, aby se respondent zúčastnil všech šetřených vln a aby byl ve všech šetřených vlnách totožný. Proto musí ve všech vlnách odpovídat ta samá osoba a není možné, aby za ní odpovídal zástupce. Z toho důvodu byly z panelových dat EU SILC vyřazeny všechny případy, které byly zodpovězeny prostředníkem. Tyto tzv. proxy výpovědi byly četné zejména v posledních dvou rotačních panelech šetření EU SILC, kde jejich proporce dosahovala třiceti, resp. dvaceti procent z celkového počtu šetřených případů. Dalším důvodem zmenšení použitých souborů dat byl jev známý jako tzv. *panel attrition* neboli výpadek panelových respondentů. Z analýzy v tomto článku vypadly všechny případy, kdy se respondent neúčastnil minimálně jedné ze

čtyř vln v panelu, a to z důvodu naznačeném v textu výše: pro analýzu reliability je nezbytné mít odpovědi od všech respondentů, kteří se účastnili všech měření. Existují metody, které jsou výjimkami z této zásady jako je např. speciální design 2 split – ballot MTMM (Sarıs, Gallhofer, 2014), ale obecně v analýze reliability tato zásada platí¹⁴.

Data pro vybrané proměnné jsem analyzovala zvlášť podle módu sběru dat. Chtěla jsem zamezit případnému mísení efektů zodpovědných za vyšší, resp. nižší reliability a případně odhalit rozdíly v reliability dat sbíraných různými módy sběru dat. V šetření EU SILC byly mezi lety 2005–2012 použity módy CAPI (Computer Assisted Personal Interview), PAPI (Paper and Pencil Interview) a velmi okrajově SAQ (Self-administered Questionnaire)¹⁵. V prvních třech rotačních panelech významně převládal mód sběru dat PAPI (cca 90 procent respondentů) a CAPI vůbec použito nebylo. V následujících dvou rotačních panelech (2008–2011 a 2009–2012) se začal používat mód CAPI (řada 2008–2011: 60 procent CAPI, řada 2009–2012: 27 procent CAPI). Aby byly výsledky statistické analýzy spolehlivé a dostatečně robustní, pracovala jsem jenom s těmi módy, pro které existoval velký počet případů, tedy s PAPI a CAPI; data ostatních módu sběru dat jsem z analýzy vypustila. Proporci použitých módů sběru dat analyzovaných položek *PH010 Zdraví*, *HS120 Příjem* a *HS130 Odhad příjmu* ukazuje tabulka 4.3.

Z důvodů uvedených v odstavcích výše jsou soubory, které jsem použila pro analýzu v tomto článku, o něco menší než longitudinální soubory, které pro daný rok a mód sběru dat nabízí ČSÚ. Snížení počtu případů v souborech však není tak velké, aby znemožňovalo úspěšné provedení analýz. V případě položky *PH010 Zdraví* se soubory PAPI dat z prvních třech rotačních panelů zmenšily pouze o 10 až 14 procent¹⁶, CAPI soubor z panelu 2008–2011 se zmenšil o 30 až 42 procent a v případě panelu 2009–2012 se CAPI soubor zmenšil o 37 až 45 procent a PAPI soubor o 19 až 31 procent. Soubory PAPI dat odpovídající položce *HS120 Příjem* pro první tři panely byly v mé analýze o 31 až 33 procent menší než původní soubory. V panelu 2008–2011 se soubor CAPI dat zmenšil o 12

¹⁴ Analýzu obsahující četné chybějící případy je rovněž možné ošetřit některou ze statisticko-analytických metod zaměřených na chybějící data, jako je např. Multigroup Maximum Likelihood (MGML) nebo Full-information Maximum Likelihood (FIML) (více Alwin 2007), a tím do analýzy začlenit i respondenty, kteří neodpověděli ve všech vlnách.

¹⁵ Metodika použitá v datových souborech EU SILC považuje proxy výpověď za jednu z alternativ módu sběru dat. To není standardní postup, proto v tomto textu neuvádím proxy výpověď mezi výčtem použitých módů sběru dat.

¹⁶ Porovnávám změnu velikosti souboru pro každý rok v panelu, proto uvádím interval procentuální změny ve všech čtyřech letech.

až 20 procent, v následujícím panelu byl soubor PAPI dat menší o 38 až 44 procent, soubor CAPI dat se zmenšil o 37 až 48 procent. Pro proměnnou *HS130 Odhad příjmu* platilo to samé nebo velmi podobné, co pro položku *HS120 Příjem*.

Tab. 4.3 Módy sběru dat; počet případů v souborech dat použitelných pro položku *PH010 Zdraví*, pro položku *HS120 Příjem* a pro položku *HS130 Odhad příjmu*.

| | Položka | Časová řada | | | | |
|-----------|---------|-------------|-----------|-----------|---------|---------|
| | PH010 | 2005–08 | 2006–09 | 2007–10 | 2008–11 | 2009–12 |
| Mód sběru | PAPI | 5353 | 4796 | 3572 | 89 | 1678 |
| | CAPI | Nepoužito | Nepoužito | Nepoužito | 1283 | 661 |
| | Ostatní | 59 | 32 | 9 | 373 | 404 |
| | Položka | Časová řada | | | | |
| | HS120 | 2005–08 | 2006–09 | 2007–10 | 2008–11 | 2009–12 |
| Mód sběru | PAPI | 2326 | 2028 | 1560 | 65 | 795 |
| | CAPI | Nepoužito | Nepoužito | Nepoužito | 1222 | 461 |
| | Ostatní | 15 | 10 | 6 | 537 | 274 |
| | Položka | Časová řada | | | | |
| | HS130 | 2005–08 | 2006–09 | 2007–10 | 2008–11 | 2009–12 |
| Mód sběru | PAPI | 2312 | 2028 | 1560 | 90 | 795 |
| | CAPI | Nepoužito | Nepoužito | Nepoužito | 857 | 461 |
| | Ostatní | 14 | 10 | 6 | 182 | 274 |

Pozn. Tab. 4.3.: Kategorie „ostatní“ odpovídá případům, které byly zodpovězeny prostředníkem (proxy výpověď), SAQ nebo mód sběru dat nebyl znám.

4. 5. 5 Analýza kvazisimplexovým modelem a odhady reliability

K analýze dat jsem použila software pro strukturální modelování LISREL (Jöreskog, Sörbom, 2004), vstupními daty byly matice kovariancí. Testovala jsem kvazisimplexové modely s oběma možnými a běžně užívanými restrikcemi, tedy podle Wileyho a Wileyho (1970) i podle Heiseho (1969). Syntaxe k oběma modelům jsou uvedeny v příloze této dizertační práce. Analýza druhého jmenovaného modelu skončila s velmi špatnými ukazateli shody modelu s daty¹⁷, proto odhady reliabilit z této analýzy v tomto textu vůbec neprezentuji. Jako úspěšná se ukázala analýza modelu s restrikcemi podle Wileyho a Wileyho, která měla dobré až vynikající ukazatele shody modelu s daty.

Kvazisimplexový model s restrikcemi podle Wileyho a Wileyho (1970) obsahoval čtyři měřené proměnné a čtyři latentní proměnné (viz Obrázek 4.1 výše v textu). Rozptyly

¹⁷ Ukazatel shody modelu s daty RMSEA byl ve všech analyzovaných panelech v případě obou proměnných nad hodnotou 0,3, která je absolutně nepřijatelná. Většina výsledků obsahovala Heywoodovy případy.

latentních proměnných byly ponechány volné, stejně jako regresní koeficienty mezi latentními proměnnými (parametry Beta). Faktorové zátěže mezi měřenou proměnnou a jí příslušející latentní proměnnou (parametry Lambda) byly fixovány na hodnotu 1, což je standardní a nutný postup, jehož prostřednictvím získává latentní proměnná měřítko. Rozptyl náhodné chyby první měřené proměnné byl ponechán volný (parametr Teta), rozptyly náhodných chyb dalších tří v čase následujících měřených proměnných byly stanoveny jako ekvivalentní prvnímu rozptylu.

Modely s daty pro položky *PH010 Zdraví* a *HS120 Příjem* konvergovaly, tj. software bez problému našel výpočetní řešení. Pouze analýza QSM s daty pro položku *HS130 Odhad příjmu* buď nekonvergovala, nebo hlásila jiné problémy, v jejichž důsledku software nebyl schopen najít řešení. Nakonec jsem tento problém vyřešila logaritmickou transformací dat; data pro položku *HS130 Odhad příjmu* jsem transformovala na jejich přirozený logaritmus a po této úpravě již model konvergoval.

Výsledky analýzy pro položky *PH010 Zdraví*, *HS120 Příjem* a *HS130 Odhad příjmu* pro data z pěti rotačních panelů z let 2005 až 2012 shrnují tabulky 4.4, 4.5 a 4.6. Velikosti tzv. stabilit (β), které nejsou pro odečtení reliability podstatné, jsou uvedeny ve zvláštní tabulce v Příloze dizertační práce.

Tab. 4.4 Výsledky QSM analýzy pro položku PH010 Zdraví

| Max.Likelihood | | | | | | | | | | | |
|----------------|----------------|----------|----------|----------|----------|------------|-------|---------------|---------------|-------|-----|
| položka PH010 | řada 2005–2008 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2005 | rel 2006 | rel 2007 | rel 2008 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 5353 | 0,828 | 0,828 | 0,828 | 0,846 | 0,833 | 0 | 0,0; 0,016 | 1 | 0 | 1 |
| položka PH010 | řada 2006–2009 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2006 | rel 2007 | rel 2008 | rel 2009 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 4796 | 0,846 | 0,828 | 0,846 | 0,846 | 0,842 | 0,027 | 0,011; 0,046 | 0,98 | 0 | 1 |
| položka PH010 | řada 2007–2010 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2007 | rel 2008 | rel 2009 | rel 2010 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 3572 | 0,828 | 0,846 | 0,828 | 0,846 | 0,837 | 0,039 | 0,021; 0,060 | 0,78 | 0,005 | 1 |
| položka PH010 | řada 2008–2011 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2008 | rel 2009 | rel 2010 | rel 2011 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| CAPI | 1283 | 0,774 | 0,757 | 0,774 | 0,774 | 0,770 | 0,04 | 0,0052; 0,078 | 0,6 | 0,007 | 1 |
| položka PH010 | řada 2009–2012 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2009 | rel 2010 | rel 2011 | rel 2012 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 1678 | 0,828 | 0,846 | 0,828 | 0,846 | 0,837 | 0 | 0,0; 0,041 | 0,98 | 0,002 | 1 |
| CAPI | 661 | 0,792 | 0,774 | 0,792 | 0,792 | 0,788 | 0,031 | 0,0; 0,090 | 0,61 | 0,006 | 1 |

Tab. 4.5 Výsledky QSM analýzy pro položku *HS120 Příjem*

| | | | | | | | | | | | |
|----------------|----------------|----------|----------|----------|----------|------------|-------|---------------|---------------|-------|-----|
| Max.Likelihood | | | | | | | | | | | |
| položka HS120 | řada 2005–2008 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2005 | rel 2006 | rel 2007 | rel 2008 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 2326 | 0,828 | 0,81 | 0,81 | 0,792 | 0,81 | 0 | 0,0; 0,031 | 1 | 0,002 | 1 |
| položka HS120 | řada 2006–2009 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2006 | rel 2007 | rel 2008 | rel 2009 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 2028 | 0,81 | 0,81 | 0,792 | 0,81 | 0,806 | 0,018 | 0,0; 0,051 | 0,94 | 0,004 | 1 |
| položka HS120 | řada 2007–2010 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2007 | rel 2008 | rel 2009 | rel 2010 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 1560 | 0,792 | 0,792 | 0,792 | 0,774 | 0,788 | 0 | 0,0; 0,048 | 0,96 | 0,003 | 1 |
| položka HS120 | řada 2008–2011 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2008 | rel 2009 | rel 2010 | rel 2011 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| CAPÍ | 1222 | 0,757 | 0,774 | 0,74 | 0,74 | 0,753 | 0,022 | 0,0; 0,065 | 0,83 | 0,007 | 1 |
| položka HS120 | řada 2009–2012 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2009 | rel 2010 | rel 2011 | rel 2012 | průměr rel | RMSEA | RMSEA 90% int | P value RMSEA | SRMR | CFI |
| PAPI | 795 | 0,81 | 0,81 | 0,792 | 0,81 | 0,806 | 0,039 | 0,0; 0,089 | 0,56 | 0,007 | 1 |
| CAPÍ | 461 | 0,723 | 0,689 | 0,706 | 0,656 | 0,695 | 0 | 0,0; 0,086 | 0,75 | 0,007 | 1 |

Tab. 4.6 Výsledky QSM analýzy pro položku *HS130 Odhad příjmu*

| | | | | | | | | | | | |
|----------------|----------------|----------|----------|----------|----------|------------|-------|--------------|---------|---------|-----|
| Max.Likelihood | | | | | | | | | | | |
| položka HS130 | řada 2005-2008 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2005 | rel 2006 | rel 2007 | rel 2008 | průměr rel | RMSEA | RMSEA 90% | P value | SRMR | CFI |
| PAPI | 2312 | 0,903 | 0,903 | 0,903 | 0,903 | 0,903 | 0 | 0,0; 0,033 | 1 | 0,0013 | 1 |
| položka HS130 | řada 2006-2009 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel2006 | rel 2007 | rel 2008 | rel 2009 | průměr rel | RMSEA | RMSEA 90% | P value | SRMR | CFI |
| PAPI | 2028 | 0,903 | 0,903 | 0,903 | 0,903 | 0,903 | 0,068 | 0,044; 0,096 | 0,11 | 0,0052 | 1 |
| položka HS130 | řada 2007-2010 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2007 | rel 2008 | rel 2009 | rel 2010 | průměr rel | RMSEA | RMSEA 90% | P value | SRMR | CFI |
| PAPI | 1560 | 0,922 | 0,903 | 0,903 | 0,903 | 0,907 | 0,021 | 0,0; 0,059 | 0,88 | 0,0025 | 1 |
| položka HS130 | řada 2008-2011 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2008 | rel 2009 | rel 2010 | rel 2011 | průměr rel | RMSEA | RMSEA 90% | P value | SRMR | CFI |
| CAPÍ | 857 | 0,922 | 0,903 | 0,903 | 0,88 | 0,903 | 0,069 | 0,031; 0,11 | 0,18 | 0,0074 | 1 |
| položka HS130 | řada 2009-2012 | | | | | | | | | | |
| Mód sběru dat | počet případů | rel 2009 | rel 2010 | rel 2011 | rel 2012 | průměr rel | RMSEA | RMSEA 90% | P value | SRMR | CFI |
| PAPI | 795 | 0,922 | 0,922 | 0,922 | 0,922 | 0,922 | 0 | 0,0; 0,0 | 1 | 0,00021 | 1 |
| CAPÍ | 461 | 0,922 | 0,903 | 0,903 | 0,903 | 0,907 | 0 | 0,0; 0,091 | 0,7 | 0,0038 | 1 |

4. 5. 6 Výsledky

Naprostá většina jednotlivých QSM analýz dospěla k výsledku, který ukazoval na vysokou shodu modelu s daty. Hodnoty ukazatelů shody modelu s daty pro položku položky *PH010 Zdraví* a *HS120 Příjem* byly ve všech panelech vynikající až dobré (viz tabulky 4.4 a 4.5). Co se týče položky *HS130 Odhad příjmu*, byly výsledky ve dvou případech ze šesti neuspokojivé. Kvazisimplexový model použitý na data z panelu 2006–2009 a panelu 2008–2011 se s daty neshodoval v ideální míře, neboť devadesátí procentní interval spolehlivosti ukazatele RMSEA překračoval doporučené hodnoty a ukazatel P value RMSEA byl menší než 0,5. Vzhledem k problematickým aspektům metody QSM popsaným v teoretické části tohoto textu jsem v této analýze uplatňovala velice striktní požadavky na shodu modelu s daty. Nespokojila jsem se s hodnotami ukazatelů, které bývají v analýzách přijímány s poznámkou, že se jedná o dostačující shodu modelu s daty, a vyžadovala jsem vynikající shodu. Z toho důvodu výsledky z kvazisimplexových modelů ze zmíněných panelů 2006–2009 a 2008–2011 nepoužívám v závěrečném shrnutí ke zhodnocení reliability položek. Hranice ideálních intervalů spolehlivosti RMSEA byla rovněž překročena v analýzách menších datových souborů, tj. *PH010 Zdraví* 2009–2012 CAPI (661 případů), *HS120 Příjem* 2009–2012 PAPI (795 případů), *HS120 Příjem* 2009–2012 CAPI (461 případů) a *HS130 Odhad příjmu* 2009–2012 CAPI (461 případů). V těchto modelech však bylo překročení limitu intervalů spolehlivosti důsledkem relativně malé velikosti souboru; ostatní ukazatele shody modelu s daty byly v normě, a proto výsledky z těchto analýz považuji za věrohodné.

Reliabilita položky *PH010 Zdraví* se ukázala relativně vysoká ve všech sledovaných panelech; v prvním panelu vyšel odhad reliability shodně 0,828 v prvních třech vlnách, v poslední vlně to bylo 0,846, průměrná hodnota odhadu reliability v tomto panelu byla 0,833. V druhém panelu byl odhad reliability v první vlně 0,846, v druhé 0,828 a ve třetí a ve čtvrté shodně 0,846, průměr činil 0,842. Ve třetím panelu byla reliability 0,828, resp. 0,846, resp. 0,828, resp. 0,846, průměrná hodnota odhadu byla 0,837. V prvních třech panelech byla data primárně sbírána metodou PAPI, v dalších letech výzkumníci začali používat sběr model CAPI. V panelu 200–2011 byla data primárně sbírána metodou CAPI a odhady reliability zde byly nepatrně nižší než v předchozích letech, tento rozdíl byl však malý. V roce 2008 činil odhad reliability 0,774, o rok později 0,757 a dále 0,774, resp. 0,774. Průměrná reliability v tomto panelu činila 0,770. V panelu 2009–2012 byla data sbírána oběma hlavními metodami sběru dat, CAPI i PAPI. V případě PAPI byly odhady

reliability o něco vyšší než odhady z dat CAPI. V první vlně s metodou PAPI byl odhad reliability 0,828, rok poté 0,846, pak 0,828 a v posledním roce 0,846. Průměrně tak byla reliability v souboru PAPI 0,837. Pro data sebraná metodou CAPI byla reliability v roce 2009 0,792, dále 0,774, o rok později 0,792 a pak rovněž 0,792. Průměrně byla reliability v souboru CAPI 0,788.

Výsledky analýzy položky *HS120 Příjem* byly srovnatelné s těmi získanými pro PH010 Zdraví. V rotačním panelu 2005–2008 byl odhad reliability pro data z prvního roku 0,828, z druhého a třetího roku 0,81 a v posledním roce 0,792. Průměrná reliability v tomto panelu byla 0,810. V panelu 2006–2009 byl odhad reliability 0,81 resp. 0,81 resp. 0,792 resp. 0,81 a průměrná reliability tedy byla 0,806. Ve třetím panelu byly odhady v roce 2007 0,792, v roce 2008 0,792, v roce 2009 rovněž 0,792 a v posledním roce 0,774. Průměrně činil odhad reliability 0,788. Stejně jako v předchozím případě byl v prvních třech panelech použit mód PAPI. Následující rotační panel, který užíval CAPI, měl v prvním roce odhad reliability 0,757, v dalším 0,774 a v následujících dvou letech shodně 0,74. Průměrná reliability v tomto panelu byla 0,753. Poslední analyzovaný panel sebraný metodou PAPI měl v roce 2009 odhad reliability 0,81, v roce 2010 rovněž, v roce 2011 0,792 a v roce 2012 0,81, průměrný odhad v celém panelu byl 0,806. Data sebraná metodou CAPI měla odhady reliability o něco nižší, a to 0,723 resp. 0,689, resp. 0,706, resp. 0,656. Průměrná reliability byla potom 0,695.

Nejoptimističtější a nejstabilnější výsledky přinesla analýza dat položky *HS130 Odhad příjmu*. V prvním panelu byl odhad reliability ve všech vlnách shodně 0,903. Z výsledků dalších analýz pro položku *HS130 Odhad příjmu*, které měly vynikající ukazatele shody modelu s daty, vyplynuly ještě vyšší odhady reliability. V panelu 2007–2010 byl v první vlně odhad reliability 0,922, v dalších vlnách 0,903, průměrný odhad z této vlny tedy byl 0,907. V posledním panelu, kde byla data sbírána módem PAPI, byly odhady reliability shodně ve všech vlnách 0,922, v CAPI souboru byly o něco nižší; v první vlně odhad činil 0,922, v druhé a dalších vlnách byly shodně 0,903, průměrná reliability v tomto panelu tedy činila 0,907.

4. 6 Diskuze

Zjištěná reliability položky *PH010 Zdraví* se pohybuje od 0,757 do 0,846, v případě položky *HS120 Příjem* je to interval od 0,656 do 0,828. Položka *HS130 Odhad příjmu* má reliability mezi 0,903 a 0,922. Dáme-li stranou diskuzi k odhadům reliability položky

HS130 Odhad příjmu, které jsou bezesporu vysoké, můžeme se ptát, zda jsou i ostatní hodnoty dostatečné. Existují obecná doporučení, co je a není vysoká reliabilita, např. Peter (1979) uvádí pravidlo podle Nunnallyho, který stanovuje, že dostatečná reliabilita má hodnotu 0,9. Takové pravidlo je však v praxi sociálněvědního výzkumu neudržitelné, neboť reliability o hodnotě 0,9 lze v běžném výzkumu dosáhnout jen výjimečně. Odhady reliability položek bývají zpravidla nižší; často se pohybují kolem hodnoty 0,7, v některých případech se dostávají i pod hodnotu 0,5 (Alwin, 2007; Saris, Gallhofer, 2003; Peter, 1979; Kroh, 2006; Erikson, 1979; Alwin, Krosnick, 1991). Pro ilustraci kvality zjištěných reliabilit položek *PH010 Zdraví* a *HS120 Příjem* jsem porovnála nalezené hodnoty s hodnotami podobných položek, které našli jiní výzkumníci a výzkumnice. Obě položky jsou tzv. subjektivní proměnné, obsahují hodnocení, týkají se zdraví, resp. příjmu a obě používají Likertovu škálu odpovědi – položka *HS120 Příjem* šestibodovou škálu, položka *PH010 Zdraví* pětibodovou. V odborné literatuře lze najít informace o reliabilitě a) položek vyžadujících nějaké hodnocení a b) týkajících se tématu zdraví a finančního příjmu. Z těchto hledisek jsou pro tuto studii relevantní práce Alwina (2007) a Kroha (2006).

Rozsáhlá QSM studie Alwina (2007) analyzovala reliabilitu položek podle mnoha kritérií. Zkoumané položky pocházely ze 112 panelů realizovaných v USA, které šetřily politické postoje, duševní zdraví, vztahy v rodině a některá další témata. Položky v analýze, které měly charakter sebehodnocení (self – assesment) či vnímání sebe sama (self – perception), tedy položky odrážející koncepty odpovídající položkám *PH010 Zdraví* a *HS120 Příjem*, měly v Alwinově studii reliabilitu průměrně 0,5 (ibid 2007:131). Podíváme – li se na reliabilitu položek v Alwinově studii podle tématu, zjistíme, že otázky ohledně finanční situace respondentů či jejich domácností měly průměrnou reliabilitu 0,592 (Alwin, 2007:161). Data týkající se zdraví v této analýze bohužel nebyla. Studie Kroha (2006) byla provedena na experimentálních split – ballot MTMM datech německého Socioekonomického panelu (SOEP) a testovala tři položky dotazující spokojenost se životem, zdravím a finančním příjmem. Reliabilita položky dotazující zdraví se sedmibodovou škálou byla odhadnuta na 0,84, v případě použití jedenáctibodové škály to bylo 0,81. Položka zjišťující spokojenost s příjmem používající sedmibodovou škálu měla reliabilitu 0,69, v případě jedenáctibodové škály to bylo 0,84. Z uvedených údajů lze usoudit, že reliabilita položek *PH010 Zdraví* a *HS120 Příjem* odhadnutá v QSM modelu je poměrně vysoká. Společně s vysokými odhady reliability položky *HS130 Odhad příjmu*

tyto výsledky naznačují, že data EU SILC jsou minimálně v případě těchto položek vysoce kvalitní.

Na solidní spolehlivost zjištěných odhadů reliability ukazuje vysoká konzistence průměrných velikostí reliabilit z různých rotačních panelů EU SILC. V případě položky *PH010 Zdraví* jsou průměrné odhady reliability v PAPI datech téměř totožné a to samé platí pro CAPI data. Průměrné hodnoty odhadů reliability položky *HS120 Příjem* jsou si rovněž velmi blízké, menší odchylka je patrná jen v panelu 2007 – 2010. Průměrné odhady reliability CAPI dat položky *HS120 Příjem* jsou méně konzistentní, zde je ovšem potřeba vzít v úvahu, že v případě panelu 2009 – 2012 byl model aplikován na výrazně menší datový soubor. Konzistence výsledků pro položku *HS130 Odhad příjmů* je ještě přesvědčivější než u ostatních položek.

Pohled do Tabulek 4 a 5¹⁸ napovídá, že mohl existovat rozdíl v reliabilitě CAPI a PAPI dat; hodnoty reliability CAPI položek se zdají o něco málo nižší než hodnoty reliability dat získaných metodou PAPI. Existoval ale rozdíl v reliabilitě CAPI a PAPI dat nebo se jednalo pouze o minimální statisticky nevýznamný rozdíl¹⁹? K zodpovězení této otázky se nabízelo použití statisticko – analytické metody Multigroup Structural Equation Modeling (MG SEM), využívající multigroup Maximum Likelihood (MG ML) odhad shody modelu s daty, která umožňuje srovnávat parametry modelu ve dvou a více skupinách dat. Měla jsem v úmyslu pracovat s předpokladem tzv. metrické invariance (Saris, Gallhofer, 2014; Trusínová, 2014), tedy s hypotézou, že faktorové zátěže v MG SEM modelu odpovídající koeficientům reliability jsou stejné ve skupině CAPI i PAPI dat. V takové analýze bych testovala, zda byly koeficienty reliability stejné ve skupinách podle modu sběru dat, a mohla bych s jistotou říci, zda se reliabilita PAPI a CAPI dat lišila. Tuto analýzu ovšem nebylo možné vzhledem k vlastnostem kvazisimplexového modelu realizovat. V SEM modelu, kde každé latentní proměnné odpovídá pouze jeden indikátor, tedy v modelu, jakým je QSM, nelze metrickou ekvivalenci faktorových zátěží v různých skupinách testovat, protože to neumožňují restriktce parametrů. Obecně v multigroup SEM

¹⁸ Vzhledem k menšímu množství použitelných QSM analýz není ve výsledcích pro položku *HS130 Odhad příjmu* rozdíl v odhadech tak zjevný jako v případě ostatních dvou položek, pro které je k dispozici více výsledků.

¹⁹ V tomto bodě je potřeba připomenout, že reliabilita je druhá mocnina koeficientu reliability. Rozdíly v reliabilitě položek se tak mohou zdát větší, neboť rozdíl v koeficientu reliability mezi položkami z různých let je umocněn na druhou. Zdánlivé rozdíly v reliabilitě mohla způsobit i často velmi rozdílná velikost analyzovaných souborů.

modelech proti sobě stojí dva a více modelů, které se mezi sebou musí v něčem lišit. Základním model, tzv. baseline, je model, kde jsou parametry v jednotlivých skupinách odhadovány nezávisle na parametrech v ostatních skupinách. Další modely s restrikcí napříč skupinami jsou pak porovnávány se základním modelem a na základě tohoto porovnání je stanoveno, zda datům odpovídá spíše základní nebo alternativní model. V testování metrické invariance koeficientů reliability v QSM modelu však nelze v softwaru pro SEM vytvořit základní model a alternativní model, které se od sebe liší: základní model i alternativní model je stejný, neboť všechny faktorové zátěže odpovídající koeficientu reliability jsou fixovány na hodnotu 1. Z tohoto důvodu nelze testovat rozdíl reliability CAPI a PAPI dat metodou MG SEM a jiná statistická procedura testování faktorových zátěží v SEM modelu není možná.

Pro testování ne/shody koeficientů reliability v PAPI a CAPI datech neexistuje ani alternativa v podobě testování shody jiných parametrů kvazisimplexového modelu. Nabízí se sice možnost testovat shodu rozptylů náhodných chyb a rozptylů pravých skóre, jejichž souvislost s reliability vyplývá z rovnice 1 (viz výše), ale tento přístup má závažné omezení. Pokud je náhodná chyba a rozptyl latentní proměnné pravého skóre konstantní napříč měřeními, je konstantní i reliability, faktem je ovšem to, že reliability může zůstat stejná i v případě, kdy se rozptyl pravého skóre či náhodné chyby mění. Naznačená analýza tak umí poznat pouze případy, kdy se reliability napříč skupinami nelišila a zároveň byly stejné rozptyly náhodné chyby a pravého skóre. Případy, kdy se reliability nelišila, ale rozptyly chyby či latentní proměnné se lišily, neumí identifikovat. Jinými slovy, neumí najít všechny případy, kdy mohla být reliability CAPI a PAPI dat stejná. Důvěryhodnou analýzu ne/shody reliability v CAPI a PAPI datech tedy touto cestou získat nelze.²⁰

Přestože nenabízím finální verdikt ohledně shody či neshody reliability CAPI a PAPI dat, shrnuji, jaké poznatky k tomuto tématu prezentuje odborná literatura. Výzkumnice a výzkumníci se rozdílům v reliability mezi CAPI a PAPI modelem příliš nevěnují.

V odborných textech často nacházíme porovnání reliability položek pocházejících z osobního rozhovoru, z telefonického rozhovoru, z dotazníku administrovaného samotným respondentem (SAQ) či z webového dotazníku (Groves, 2004:501; Dillman,

²⁰ Jen pro úplnost dodávám, že v MG SEM analýze invariance chybových rozptylů a rozptylů pravých skóre ve skupinách PAPI a CAPI dat nebyla nalezena jediná shoda příslušných parametrů v příslušných letech.

Christian, 2005; Schwartz et al., 1991; Revilla, Saris, 2013b). Pokud je porovnávána kvalita PAPI a CAPI dat, soustřeďuje se zájem výzkumníků na hodnocení jiných aspektů kvality dat, jako je celková (unit non-response) a položková návratnost (item non-response), vliv sociální desirability nebo eliminace chyb vzniklých při přepisu dat (Schrapler, Schupp, Wagner, 2006; Hox, de Leeuw, Snijkers, 1998; de Leeuw, Hox, Snijkers, 1998). Zmínku o porovnání reliability položek sebraných CAPI a PAPI obsahuje práce Sarise a Gallhoferové (2007), kteří uvádějí, že data získaná za asistence počítače jsou stejně reliabilní jako PAPI data. Jejich závěr potvrzuje práce Kroha (2006), který v datech z německého Socioekonomického panelu rovněž nenašel rozdíl v reliabilitě CAPI a PAPI dat. Pokud by se tedy reliability položek PH010 Zdraví a HS120 Příjem lišila, byl by to odlišný nález od dosavadních poznatků.

Při pohledu na tabulky 4 a 5 se dále nabízí upozornění na rozdíl ve velikosti hodnot odhadů reliability položek PH010 *Zdraví* a HS120 *Příjem*; hodnoty odhadů první uvedené položky jsou o něco vyšší než hodnoty odhadů druhé. Z důvodu popsánoho v předchozím textu nebylo ani v tomto případě možné statisticky testovat, zda byla reliability položky HS120 *Příjem* skutečně nižší. Pokud by rozdíl existoval, mohlo by být jeho příčinou použití pětibodové (PH010 *Zdraví*), resp. šestibodové škály odpovědi (HS120 *Příjem*). O problematice měřících škál píše např. Saris a Gallhoferová (2007: 34), kteří na 2 SB MTMM datech ukazují, že použití střední kategorie odpovědi ve škále významně zvyšuje reliability. Rovněž Alwin (2007:193) ve své rozsáhlé QSM studii uvádí, že pětibodová škála je více reliabilní než šestibodová, o vyšší kvalitě pětibodových škál píše i Revilla, Saris a Krosnick (2014).

Tuto kapitolu lze uzavřít tvrzením, že odhady reliability položek šetření EU SILC analyzovaných metodou QSM jsou vysoké, neboť se téměř všechny pohybují těsně kolem hodnoty 0,8 nebo nad ní. Prezentované odhady jsou rovněž s vysokou pravděpodobností spolehlivé, neboť byly potvrzeny na datech z pěti resp. tří různých panelů. Rozdíly v reliabilitě jednotlivých položek a podle použitého modu sběru dat bohužel nebylo možné prokázat vzhledem k omezením, které kladou principy modelování strukturálních rovnic na analýzu ekvivalence parametrů.

5. Modelování stylu odpovídání v modelu konfirmativní faktorové analýzy: Souhlasný styl odpovídání v měření postojů k přistěhovalcům

Chyby měření mohou mít náhodný a systematický charakter; o náhodné chybě bylo pojednáno v předchozí kapitole, zatímco tato kapitola se zabývá systematickou chybou. Ta může vznikat v důsledku různých činitelů přítomných při dotazování ve výběrových šetřeních; těmito činiteli mohou být vliv charakteristik tazatele, vliv módu sběru dat či vliv použité metody měření. Systematickou chybou, na kterou se soustředím v této kapitole, je odchylka, která vznikla v důsledku tzv. stylu odpovídání. Tato odchylka je do velké míry efektem použité metody měření, takže její analýza spadá do oblasti výzkumu chyb měření vzniklých vlivem použité metody.

Styl odpovídání je jev, kdy respondent odpovídá na položky dotazníku systematickým způsobem bez ohledu na skutečný význam položek. Jeho důsledkem je systematická odchylka, kterou angličtina krátce a výstižně nazývá jako *response style bias*. Vzhledem k tomu, že se v této práci snažím používat české výrazy pro názvy, jež byly vytvořeny v angličtině, volím pro uvedený anglický termín sice poněkud zkratkovitý, ale stručný ekvivalent *odchylka RS*. Tuto odchylku lze modelovat jako latentní proměnou v modelu konfirmativní faktorové analýzy (CFA) a zjistit tak její vliv na odpovědi respondentů. Popisovanou metodu identifikace odchylky RS v CFA je možné chápat v rámci rozšířené klasické testové teorie, která tvoří teoreticko – analytický rámec této dizertační práce. Rozšířená KTT je však v případě odchylky RS využita jen částečně; modelování odchylky RS sice využívá rozložení rozptylu měřené proměnné na substanciální a chybovou část, nedokáže však izolovat pravý skór proměnné. Prostřednictvím modelování odchylky RS tak lze získat jen informaci o vlivu systematické chyby vzniklé vlivem RS, nikoliv o vlivu náhodné chyby.

Jednou z motivací pro vypracování této studie byl fakt, že o českém stylu odpovídání a odchylce, kterou způsobuje, v současné době nic nevíme. Analýzu stylů odpovídání na českých datech dosud nikdo neprezentoval a informace o stylech odpovídání v jiných postkomunistických zemích jsou často omezené a založené na analýze, která využívá méně důvěryhodnou metodu identifikace stylů odpovídání (Meisenberg, Williams – Shillingford, 2008; Harzing, 2006). Přitom je znalost o charakteru českého stylu odpovídání metodologicky hodnotná informace, jejímž prostřednictvím lze získat větší porozumění českým respondentům a která nás může navést k dalšímu, metodologicky i substanciálně

relevantnímu výzkumu. Vzhledem k tomu, že styly odpovídání jsou důsledkem použití Likertových škál, které jsou ve výběrových šetřeních využívány velice často, měl by být náš zájem o studium odchylky RS o to větší; styly odpovídání se teoreticky mají možnost objevit takřka v každém šetření. Přesto v současné době nevíme, jak odchylka RS vychyluje měření v českých datech, a kdo jsou lidé, kteří jsou ke stylům odpovídání náchylnější.

Modelování odchylky RS je užitečné i v běžné substanciální analýze využívající modelování strukturálních rovnic. Tam může ukázat, nakolik styl odpovídání ovlivnil vztahy mezi zkoumaným konstruktem a jeho indikátory, a implementace latentní proměnné odchylky RS pak může zlepšit shodu modelu s daty. Modelování stylů odpovídání nachází využití především v mezinárodním komparativním výzkumu při analýze invariance měření; v datech z různých zemí mohou existovat různě velké odchylky RS a jejich rozptyly mohou bránit prokázání invariance měření.

Výklad i analýza prezentovaná v této kapitole se zaměřuje na souhlasný styl odpovídání (z angl. *acquiescent response style*, ARS), jeden z možných a nejčastějších stylů odpovídání. V analýze ukazují možnosti využití metody modelování odchylky ARS v modelu CFA, který jsem pro účely studie pojmenovala CF-RS model. Použitá metoda plně vychází z popisu aplikace modelu, prezentovaného v článku Billieta a McClendona (2000), kteří stanovili kritéria pro identifikaci ARS odchylky v datech.

Cílem analýzy v této kapitole je v první řadě získat informace o vlastnostech českého ARS a porovnat je s charakteristikami ARS nalezeného v datech ze západoevropských zemí. Studie pracuje s předpokladem o kulturních odlišnostech mezi postkomunistickými a západoevropskými zeměmi, v jejichž důsledku se v těchto zemích může lišit charakter ARS. Analýza se konkrétně zaměřuje na doložení korelace faktoru ARS se vzděláním a věkem, která byla v minulosti nalezena v datech ze západoevropských zemí.

V následujícím textu nejdříve popisují druhy stylů odpovídání a příčiny jejich vzniku. Dále uvádím, které demografické či kulturní proměnné mají na jejich výskyt vliv a jaký vliv mají styly odpovídání na kvalitu dat. Následuje popis metod identifikace stylů odpovídání, kde je důraz kladen především na popis CF-RS modelu, použitého v analýze prezentované v empirické části kapitoly. V analýze pak modeluji odchylku ARS na českých datech ze šetření International Social Survey Programme (ISSP), konkrétně v datech z baterie zjišťující postoje k přistěhovalcům.

5. 1 Styl odpovídání a druhy stylů odpovídání

Styl odpovídání (z angl. *response style*, RS) je habitualizovaný způsob odpovídání na škále odpovědi dotazníkové položky, kdy respondent odpovídá bez ohledu na význam položky (Van Vaerenbergh, Thomas, 2013; Billiet, McClendon, 2000). Odborná literatura rozlišuje několik typů stylu odpovídání; 1) souhlasný styl odpovídání (z angl. *acquiescent response style* (ARS) neboli *yeah – saying*), což je tendence respondentů souhlasit; 2) nesouhlasný styl odpovídání (z angl. *disagreeing response style* (DRS) čili *noy – saying*), tedy tendenci respondentů nesouhlasit; a 3) střední styl odpovídání (z angl. *middle response style* (MRS)), tj. tendenci volit střední bod ve škále odpovědi, v literatuře označovaný také jako *middle saying* (Tourangeau, Rips, Rasinski, 2000). V zahraniční metodologické literatuře lze dále najít studie zabývající se extrémním stylem odpovídání (z angl. *extreme response style*, ERS), což je tendence respondentů volit první či poslední bod na škále odpovědi (Greenleaf, 1992b; Meisenberg, Williams – Shillingford, 2008; Wetzel, Lüdtke, Zettler, Böhnke, 2016; Van Vaerenbergh, Thomas, 2013; Cheung, Rensvold, 2000). Respondenti používající ERS užívají jen krajní body škály odpovědi, přičemž některé studie přinášejí informace o tom, že tato tendence je do jisté míry kulturně podmíněná. Např. Cheung a Rensvold (2000) informují o tom, že Korejci používají ERS výrazně méně než Evropani, a spekulují, že se korejští respondenti krajním bodům škály vyhýbají z důvodu sociální normy, jež trestá jakoukoliv vyhraněnost.

Tab. 5.1 Znázornění typů stylů odpovídání na sedmibodové škále; tučně zvýraznění znázorňuje odpovědi

| Název | Zkratka | Označené odpovědi |
|-----------------------------|---------|----------------------|
| Souhlasný styl odpovídání | ARS | 0 0 0 0 0 0 0 |
| Nesouhlasný styl odpovídání | DRS | 0 0 0 0 0 0 0 |
| Střední styl odpovídání | MRS | 0 0 0 0 0 0 0 |
| Extrémní styl odpovídání | ERS | 0 0 0 0 0 0 0 |

V metodologické literatuře lze najít desítky studií zabývajících se všemi výše uvedenými styly odpovídání, Meisenberg a Williamsová – Shillingfordová (2008) ovšem uvádějí, že v empirii se nejčastěji vyskytují souhlasný styl odpovídání (ARS) a extrémní styl odpovídání (ERS). Studie, jež je prezentovaná v empirické části této kapitoly, se soustředí na ARS, resp. na odchylku ARS, tj. systematickou chybu měření, jež vzniká právě v jeho důsledku.

5. 2 Příčiny vzniku stylů odpovídání

Studium stylů odpovídání doprovází debata o příčinách jejich vzniku. Odborníci v podstatě souhlasí s tím, že RS mají dva zdroje; prvním je použitý měřicí nástroj, druhým je respondent, resp. typ respondenta, který ze své povahy inklinuje k souhlasu (ARS), nesouhlasu (DRS), volbě středního bodu na škále (MRS) atp. (Van Vaerenbergh, Thomas, 2013). Vznik odchylky RS je pak důsledkem reakce respondenta na použitý nástroj měření.

Co se týče první uvedené příčiny, uvádí metodologicky zaměřeni autoři jako jasný zdroj RS používání Likertových škál (McClendon, 1991; Billiet, McClendon, 2000; Groves et al., 2004; Tourangeau, Rips, Rasinski, 2000). Tyto škály odpovědi jsou v empirickém dotazníkovém výzkumu využívány mimořádně hojně; nejčastěji je nacházíme v bateriích položek a zpravidla se s nimi setkáváme ve formě pětibodových škál, často jsou ale užívány i sedmibodové, devítibodové či jedenáctibodové Likertovy škály. Vzhledem k tomu, jak často jsou tyto škály používány, nemůže být sporu o tom, že je nutné zabývat se metodologickým efektem, který způsobují.

V případě respondenta jako zdroje odchylky způsobené styly odpovídání byl prokázán vliv specifických charakteristik, ke kterým lidé inklinují v různé míře. Nejčastější takovou uváděnou specifickou charakteristikou je tzv. *satisficing*, fenomén, pro který není jednoduché najít seriózně znějící český ekvivalent. Z povahy tohoto jevu vyplývá, že se jedná o jistou formu „ulévání se“, kdy respondent, nejčastěji v osobním rozhovoru s tazatelem (face – to – face; 2F2), prostě jenom poskytuje nějaké odpovědi a příliš se nesnaží o to, aby tyto odpovědi skutečně odpovídaly požadovanému vyjádření názoru či údají. Pro účely českého odborného textu je asi nejvhodnější překlad *vyhovění tazateli*, blízký anglickému slovu *satisfy*, který zhruba vystihuje podstatu fenoménu – vyhovět tazateli a odpovědět.

Strategii *vyhovění tazateli* respondent aplikuje, když nemůže najít vhodnou odpověď nebo se nechce v dané chvíli namáhat natolik, aby zodpověděl otázky s plným kognitivním nasazením. V případech, kdy respondent *nemůže* najít odpověď, bývají na vině jeho nižší kognitivní schopnosti a/nebo vysoká kognitivní náročnost položky. Kognitivně náročné položky bývají složitě formulovány, případně jsou dlouhé či používají složitá slova (Krosnick, 1991; Tourangeau, Rips, Rasinski, 2000). Např. práce Knowlese a Condon (1999) prokázala, že vysoká kognitivní náročnost položek zvyšuje výskyt odchylky ARS. Samostatnou kapitolou jsou pak metodologicky špatné položky, které nemají jednoznačný

význam či se ptají na téma, o kterém většina respondentů nic neví. U takových položek je pak výskyt některého ze stylů odpovídání vysoce pravděpodobný. V případech, kdy se respondent *nechce* namáhat, bývá důvodem zpravidla jeho malá motivace odpovědět na položku nebo se účastnit rozhovoru (Krosnick, 1991). Na výskyt ARS v důsledku nízké motivace respondenta upozorňují např. Kaminská, McCutcheon a Billiet (2010), kteří varují před nátlakem na respondenty, aby se účastnili výzkumu. Podle jejich názoru mají snahy o zvýšení návratnosti leckdy skryté úskalí v tom, že data od jednotlivců, kterým se zprvu do výzkumu vůbec nechtělo, mohou mít nižší kvalitu, neboť s vyšší pravděpodobností obsahují chyby měření, mezi nimi i tu způsobenou styly odpovídání. V metodology velice oblíbené knize o kognitivní psychologii respondentů (2000) prezentují Tourangeau, Rips a Rasinski kognitivní fáze, kterými musí respondent projít, aby mohl dodat validní odpověď na dotazníkovou položku. V případě fenoménu *vyhovění tazateli* respondent neprochází všemi těmito fázemi a jeho odpověď vzniká na bázi dojmů a improvizace. Právě v takové situaci, kterou ještě umocňuje použití baterie položek s Likertovou škálou, se uplatňují styly odpovídání.

Na náchylnost k některému z typů RS danou vlastnostmi respondenta ukazuje fakt, že RS, a to především ARS, zpravidla koreluje pozitivně s věkem a negativně se vzděláním. V tomto kontextu upozorňují Billiet a McClendon (2000) na často rozdílné pohledy psychologů a sociologů. Zmiňují práci Rorera (1965), podle kterého není ARS charakteristikou respondenta, ale vzniká výhradně v důsledku konkrétního měřicího nástroje. Jeho pohled na ARS mezi psychology získal řadu významných zastánců, zatímco mezi sociology nikoliv. Billiet a McClendon přičítají tyto rozdílné pohledy charakteru výběrových souborů, které jsou používány v psychologii a sociologii. Vzhledem k tomu, že psychologové používají především soubory z homogenních populací (často se jedná o soubory studentů psychologie) a data sbírají metodou SAQ (z angl. *self-administered questionnaire*; mód sběru dat, kdy respondent vyplňuje dotazník sám), nemají v datech dostatek variance, aby doložili příslušné vztahy mezi ARS a demografickými proměnnými, případně módem sběru dat.

Tourangeau, Rips a Rasinski (2000) na konkrétních případech ilustrují, že souhlasný styl odpovídání (ARS) může být u určitého typu lidí reakcí na přítomnost tazatele. Někteří lidé, obzvláště ti, kteří pocházejí z nižší sociální skupiny než tazatel, podle nich mají tendenci souhlasit s výroky, které jim tazatelé čtou, protože tazatele považují za autoritu, které by

neměli oponovat. Vliv přítomnosti tazatele na výskyt ARS má i další příčinu, a tou je sociální desirabilita. Rossová a Mirowsky ve své práci (1984) studovali vliv sociální desirability na ARS a na datech z USA a Mexika a ukázali, že současný výskyt ARS a sociální desirability je častější u lidí s nižším SES, příslušníků etnické minority a starších lidí. Roli tazatele na výskyt stylu odpovídání zkoumaly rovněž Olsonová a Bilgenová (2011), které přišly s překvapivým závěrem. Navzdory obecnému předpokladu, že zkušenější tazatelé sbírají data s vyšší kvalitou, v analýze našly důkazy o tom, že data od zkušenějších tazatelů obsahují víc chyb v důsledku ARS než data od méně zkušených tazatelů.

Shrnutí příčin výskytu stylů odpovídání vybízí k otázce, zda je používání stylů odpovídání respondenty vyvoláno spíše aktuální situací a okamžitou reakcí na použití konkrétních položek a měřicího nástroje, či zda se jedná o dlouhodobě trvající vlastnost respondenta, která se projevuje při každém výzkumném rozhovoru. Odpověď, která se přiklání k druhé variantě, přinesly např. práce Wetzelové a kolektivu (Wetzel et al., 2016) a Billieta a Davidova (2008). První identifikovali stabilitu stylu odpovídání na německých panelových datech, druzí totéž doložili na datech z Belgie.

5. 3 Vztah ARS k proměnným na individuální úrovni a na úrovni zemí

Odborná literatura opakovaně dokládá, že různé styly odpovídání korelují s konkrétními demografickými charakteristikami (Weijters, Geuens, Schillewaert, 2010b; Meisenberg, Williams – Shillingford, 2008; Moors, 2008; Billiet, McClendon, 2000; Greenleaf, 1992a; Ross, Mirowsky, 1984). ARS byl několikrát doložen jako korelát věku a vzdělání především na datech ze západní Evropy. Ukazuje se, že lidé, kteří mají tendenci souhlasit s protichůdnými výroky, jsou častěji starší a častěji mají nižší vzdělání (Billiet, McClendon, 2000). Různí autoři vidí jako původce těchto vztahů společný faktor, a to nižší kognitivní schopnosti (Weijters, Geuens, Schillewaert, 2010a; Knowles, Condon, 1999; Krosnick, 1991). Vliv vzdělání na odchylku RS pak vysvětlují recipročním vztahem mezi vzděláním a kognitivními schopnostmi s odůvodněním, že lidé s vyššími kognitivními schopnostmi získávají vyšší vzdělání, které zároveň dále zlepšuje jejich kognitivní schopnosti (Weijters, Geuens, Schillewaert, 2010b). Naopak vyšší věk je spojován s útlumem kognitivních funkcí jako je např. paměť (Knauper, 1999), což se odráží ve vyšším výskytu RS mezi staršími lidmi. Pozitivní korelace výskytu ARS a věku byla

doložena v několika studiích (Weijters et al., 2010b; Meisenberg, Williams – Shillingford, 2008; Billiet, McClendon, 2000; Greenleaf, 1992a; Ross, Mirowsky, 1984).

Analýza na datech z Belgie (Weijters, Geuens, Schillewaert, 2010b) prokázala negativní lineární vztah mezi ARS, DRS a ERS a vzděláním, stejně tak studie Billieta našla negativní korelaci ARS se vzděláním (Billiet, McClendon, 2000). Vzhledem k tomu, že se v jednotlivých zemích liší vzdělávací systémy a že v různých zemích panuje různá nerovnost vzhledem k možnosti vzdělání získat, nabízí se otázka, zda je vztah mezi vzděláním a odchylkou RS přítomen ve všech zemích stejně. Meisenberg a Williamsová – Shillingfordová (2008) našli negativní korelaci mezi ARS, resp. ERS a vzděláním v souboru dat z mezinárodního šetření World Values Survey (WVS) z roku 2000 a konstatovali, že vztah vzdělání a RS je celosvětový fenomén. Jejich výsledky jsou však diskutabilní vzhledem k použité metodě identifikace ARS; autoři použili nepříliš důvěryhodnou metodu sumačního indexu kladných odpovědí, která nerozlišuje mezi skutečným souhlasem a souhlasem jako stylem odpovídání. Závěry jejich studie je tak třeba brát s rezervou. Bez ohledu na pochybnosti o kvalitě jejich výsledků je ovšem důležité zmínit, že autoři v článku bohužel neprezentovali korelace ARS a vzdělání pro každou zemi zvlášť, ale jen pro velké geograficko – kulturní celky jako např. Afrika, Blízký východ, postkomunistické země atd. Především s ohledem na analýzu, která je prezentována v empirické části této kapitoly, je škoda, že nejsou k dispozici informace o korelaci vzdělání a ARS v jednotlivých, zejména postkomunistických zemích.

Někteří autoři se ve výzkumu zaměřili na kulturní rozdíly mezi zeměmi a jejich vliv na výskyt odchylky RS (Van Vaerenbergh, Thomas, 2013; Cheung, Rensvold, 2000). Tyto studie analyzují vztah odchylky RS s proměnnými, které indikují kulturní charakter dané země. Cheung a Rensvold (2000) konstatovali, že pozorované kulturní rozdíly mohou být důsledkem kulturně determinovaných fenoménů jako je sociální desirabilita či přesvědčení o tom, že souhlas je žádoucí odpověď. Výsledky studií, které hledaly důkazy o vlivu proměnných na úrovni zemí, se liší podle toho, jakou metodu identifikace odchylky RS autoři zvolili; výzkumníci a výzkumnice, kteří použili metodu sumačních indexů souhlasných odpovědí, která není příliš přesná, jak ukazují dále v textu v popisu metod identifikace RS, důkazy o vlivu různých indikátorů kulturního typu našli. Naopak ti, kteří použili metodu modelování odchylky RS v modelu CFA, jež je nástrojem analýzy v této kapitole, vliv těchto indikátorů neprokázali.

Podle studie Meisenberga a Williamsové – Shillingfordové (2008) je vliv ARS a ERS nejsilnější v méně vyvinutých zemích. Co se týče vztahů mezi výskytem RS a proměnnými na úrovni zemí, je podle nich nejsilnějším determinantem ERS nižší průměrné IQ v zemi a nejsilnějším determinantem ARS je vysoká úroveň korupce. Nejsilněji jsou ARS ovlivněni obyvatelé Afriky, Jižní Asie a Blízkého východu, nejméně naopak obyvatelé anglicky mluvících zemí a evropských zemí s protestantskou historií. Práce Van Herkové, Poortinga a Verhallena (2004) stejně jako předchozí studie používá metodu sumační indexů a dokládá, že ARS je silnější ve středomořských státech jako je Řecko či Itálie a slabší v západoevropských zemích jako je Německo, Anglie nebo Francie. Rozdíly v RS mezi zeměmi zkoumala i Harzingová (2006), která rovněž našla rozdílnou náchyllost respondentů k ARS ze zemí Evropy, Severní a Jižní Ameriky a Asie. Její analýza však proběhla na vzorcích vysokoškolských studentů, takže výsledky studie nejsou reprezentativní za populace národních států. V kontrastu s uvedenými výsledky svědčícími pro rozdílnost RS napříč zeměmi stojí studie Billieta, Cambrého a Welkenhuysen – Gybelse (2002), kteří použili metodu modelování odchylky ARS. Autoři analyzovali data z devíti západoevropských zemí a nenašli rozdíly v síle vlivu ARS. Jedinou výjimkou byla Itálie, kde analytici vůbec neidentifikovali přítomnost ARS.

5. 4 Vliv odchylky RS na kvalitu dat

Jaký vliv má výskyt stylů odpovídání na kvalitu dat z výběrového šetření? Vzhledem k tomu, že se jedná o systematickou chybu, snižuje validitu měření. Všechny typy stylu odpovídání působí v jednom směru, čím systematicky vychylují průměry proměnných a současně ovlivňují velikosti rozptylů měřených proměnných. Odchylka RS se tak projevuje jak v popisných statistikách, jako jsou frekvence a průměry, tak v analýzách založených na korelacích, tj. zejména v často používaných regresních analýzách (Mirowsky, Ross, 1991; Van Vaerenbergh, Thomas, 2013). V analýze SEM odchylka RS deformuje vztah mezi položkou a měřeným konstruktem a ovlivňuje regresní vztahy mezi latentními proměnnými (Billiet, Davidov, 2008). Detailní rozbor vlivu RS na měřenou proměnnou rozepsaný v rovnicích poskytují Welkenhuysen – Gybels, Billiet a Cambré (2003).

Na negativní vliv stylů odpovídání upozorňují metodologové především v případech, jsou – li v analýze porovnávány skupiny. Různé skupiny respondentů se totiž mohou lišit používaným stylem odpovídání a nezahrnutí vlivu odchylky RS do substanciální analýzy může vyústit v chybné závěry o rozdílech mezi skupinami. Takovým nejčastěji

uvažovaným a analyzovaným případem je mezinárodní komparativní analýza (Billiet, Cambré, Welkenhuysen – Gybels, 2002). Autoři často upozorňují, že respondenti z různých zemí mohou používat různé styly odpovídání, případně používají stejný styl odpovídání, ovšem v různé míře, čímž vytvářejí systematickou chybu o různém směru či velikosti. Analýza odchylky RS se však neomezuje jen na porovnávání zemí; např. Di Stefanová a Motl (2009) srovnávali odchylku RS mezi muži a ženami v míře Rosenbergovy škály sebevědomí, rozdíl se jim však nepodařilo prokázat. Rozdíly v RS je možné hledat i v longitudinálních datech; Billiet a Davidov (2008) upozorňují na základě výsledků získaných na panelových datech, že nezahrnutí systematické chyby vzniklé v důsledku RS může vést k deformacím koeficientů stability a k chybným odhadům dalších parametrů modelu aplikovaného na panelová data.

Ve světle popsaného negativního vlivu RS na kvalitu dat se nabízí otázka, zde se lze výskytu stylů odpovídání nějak vyhnout. Někteří výzkumníci navrhují, aby respondenti místo výroku a Likertovy škály dostali přímou otázku a na výběr dvě varianty odpovědi, první reprezentující negativní postoj, druhá pozitivní. Takové položky anglicky psaná odborná literatura označuje jako *forced – choice items* (Billiet, McClendon, 2000). Jiní výzkumníci se proti této praxi staví s nedůvěrou a upozorňují na její nedostatky; data získaná tímto způsobem mají nominální, resp. dichotomickou povahu, což zásadně omezuje možnosti analýzy takových dat. Navíc tato forma odpovědi není údajně oblíbená mezi respondenty, neboť nabízené alternativy některým nepřipadají jako protikladné a cítí se být nuceni do jedné ze dvou variant (Billiet, Davidov, 2008).

Kromě metodologických pojednání, která doporučují brát v úvahu chybový rozptyl vzniklý vlivem stylů odpovídání, existují i kritické pohledy na účel a opodstatnění takové analýzy. Někteří autoři zpochybňují výzkum tohoto fenoménu i snahy o eliminaci chybového rozptylu. He, Fons a Van der Vijver (1995) uvádějí, že podle některých autorů nemá zahrnutí odchylky RS závažnější dopad na výsledky analýz a odvolávají se na práci Onesové a kolegů (1996). Dále však uvádějí odkazy na jiné studie, které vliv odchylky RS prokázaly. Rovněž vyjadřují pochybnosti o přístupu, který odchylku RS chápe jako chybu měření. Upozorňují, že styly odpovídání mohou být reflexí jak osobnostně, tak kulturou podmíněných komunikačních stylů respondentů, a že korekce pro varianci odchylky RS může mít za následek odstranění validní variance. Tato výtky je relevantní kritikou přístupu sociálněvědní kvantitativní metodologie, která nelogičnost v chování respondentů chápe jako cosi, co nereflektuje skutečné vlastnosti respondentů, a lze ji rozšířit o

podezření, že společně s odlišností komunikačních stylů mohou být příčinou identifikovaných stylů odpovídání i rozdíly v chápání společenských fenoménů v různých kulturách. Především s ohledem na kulturní rozdíly mezi zeměmi s dlouhou demokratickou tradicí a zeměmi ve stádiu rozvoje, bychom se mohli začít zabývat otázkou, zda ve všech zemích bez rozdílu skutečně platí západoevropská a severoamerická pravidla ideologické logiky, která stojí za designem většiny měřících nástrojů abstraktních sociálněvědních konceptů a konstruktů, jako je např. politická orientace, postoje k přistěhovalcům či postoje k genderové rovnosti. Je možné, že v zemích mimo bohatý Sever, kde se jmenované ideologické fenomény vytvářely za jiných podmínek a kratší dobu než v západní Evropě a v Severní Americe, jsou pro respondenty zároveň možné a svým způsobem racionálně opodstatněné odpovědi, které jsou z hlediska západní ideologické logiky protikladné. V případě analýzy ARS či DRS, tedy souhlasného či nesouhlasného stylu odpovídání, pak jsou takové odpovědi považovány za systematickou chybu, nikoliv za substanciální informaci. Hypotéza o vlivu rozdílného chápání sociálněvědních konstruktů v zemích s různou úrovní demokratického rozvoje potřebuje detailní studium a může být výzvou pro metodologii výběrových šetření a odrazovým můstkem pro rozsáhlou studii mezinárodních dat.

5. 5 Modelování RS v modelu CFA

Způsobů, jak se vypořádat s chybou měření vzniklou vlivem stylu odpovídání, existuje několik. Odborná literatura nejčastěji uvádí čtyři metody, z nichž všechny lze použít pro identifikaci či eliminaci souhlasného (ARS) a nesouhlasného stylu odpovídání (DRS), a jen některé pro identifikaci a eliminaci středního (MRS) a extrémního stylu odpovídání (ERS). Předmětem výkladu a nástrojem analýzy v této kapitole je modelování odchylky RS v datech z vyvážené škály v modelu CFA, proto se jejímu popisu v tomto textu dostává nejvíc prostoru. Dále však uvádím i další metody identifikace RS, abych čtenářům a čtenářkám ukázala alternativy popisovaného přístupu.

Metodu modelování odchylky RS jako faktoru v modelu CFA do širšího povědomí v sociologické metodologii rozšířil článek Billieta a McClendona (2000), kteří v podstatě použili postup již dříve popsany v článcích Mirowského a Rossové (1991) a Watsonové (1992). Článek se stal v oboru analýzy stylů odpovídání velice vlivným a metodu převzali další autoři a použili ji v navazujících studiích (např. Billiet, Cambré, Welkenhuysen – Gybels, 2002; Welkenhuysen – Gybels, Billiet, Cambré, 2003; Billiet, Davidov, 2008).

Metoda používá data z vyvážené baterie položek, tj. takové baterie, kde je zhruba polovina položek formulována ve prospěch měřeného konstruktů a druhá polovina proti němu. Princip metody pracuje s předpokladem, že situace, kdy respondent souhlasí jak s položkami, které jsou formulovány ve prospěch měřeného konstruktů, tak s položkami, které jsou formulovány proti němu, znamená manifestovaný souhlas bez ohledu na skutečný význam položky. Souhlas s protikladnými výroky v baterii tato metoda nepovažuje za reflexi měřeného konstruktů, ale za systematickou chybu měření vzniklou v důsledku ARS.

Podstata metody spočívá v modelování latentní proměnné odpovídající měřenému konstruktů (*content factor*, CF) a latentní proměnné odpovídající stylu odpovídání (*response style*, RS) (Billiet, McClendon, 2000; Welkenhuysen – Gybels, Billiet, Cambré, 2003; Billiet, Davidov, 2008; Watson, 1992). Vzhledem k tomu, že autoři, kteří model definovali, jej nijak nepojmenovali, nazývám ho v tomto textu pracovní CF-RS model (CF – content factor, RS – response style). Na rozdíl od dalších dvou metod používaných v této dizertační práci, tj. TS MTMM modelu a kvazisimplexového modelu, CF-RS model neobsahuje strukturální část, ale jen část měření, neboť se jedná o model konfirmativní faktorové analýzy. Tuto část měření lze vyjádřit obvyklým zápisem pro SEM jako:

$$(5.1) y = \Lambda_{y1} \xi_1 + \Lambda_{y2} \xi_2 + \varepsilon_y,$$

kde y reprezentuje měřenou proměnnou, Λ_y faktorové zátěže mezi měřenou proměnnou a latentními proměnnými a ξ_1 a ξ_2 latentní proměnné měřeného konstruktů a stylu odpovídání. Názorněji lze CF-RS model popsat zjednodušenou rovnicí:

$$(5.2) y = CF + RS + e$$

V rovnici 5.2 vidíme, že rozptyl měřené proměnné y je rozdělen na rozptyl odpovídající měřenému konstruktů, tj. obsahovému faktoru (CF), rozptyl odpovídající odchylce způsobené stylem odpovídání (RS) a varianci náhodné chyby e , tj. zbytkovému rozptylu, který neodpovídá ani konstruktů ani stylu odpovídání. Restrikce a další náležitosti aplikace CF-RS modelu upřesňují v empirické části této kapitoly, kde je jejich umístění vzhledem k návaznostem praktičtější.

S využitím CF-RS modelu můžeme v datech z vyvážené baterie položek identifikovat rozptyl, který je důsledkem výskytu ARS (souhlasného stylu odpovídání) nebo DRS (nesouhlasného stylu odpovídání). Oproti jiným metodám, které popisují v další části textu, má tento přístup výhodu v tom, že dokáže přímo identifikovat varianci, která je reflexí RS, přičemž nezaměňuje varianci obsahu a stylu, a rovněž nepotřebuje žádná dodatečná měření, ale vystačí si se substantiálními mírami, tj. s položkami ve vyvážené baterii, ze kterých dokáže extrahovat varianci odchylky RS (Billiet, McClendon, 2000). Nevýhodou metody modelování RS je, že jejím prostřednictvím nelze analyzovat ERS, a že nedokáže určit indikátor RS pro jednotlivého respondenta.

Model CF-RS lze aplikovat na data z jedné vyvážené baterie měřící jeden substantiální konstrukt, v takovém případě modelujeme jen jeden ARS faktor, příp. DRS faktor. Je ovšem možné používat data z více baterií měřících více konstruktů, a v tom případě modelujeme pro každý konstrukt jeden RS faktor. Pak je možné pátrat po tom, zda je RS faktor specifický pro každý měřený konstrukt, či zda byla měření všech konstruktů ovlivněna stejným RS faktorem, což by ukazovalo na to, že RS faktor je dán především tendencí respondentů v souboru ne/souhlasit, nikoliv tematickým zaměřením použitého nástroje. Příklad takové analýzy ukazuje článek Billieta a McClendona (2000).

Co získáváme použitím Modelu CF-RS v běžné analýze? Např. v situaci, kdy confirmujeme jedno – faktorové řešení pro data z baterie měřící jeden konstrukt, získáme při implementaci ARS faktoru lepší shodu modelu s daty (Billiet, McClendon, 2000). To se na první pohled nemusí zdát důležité, nicméně existují situace, kdy odchylka ARS může v datech jedno – dimenzionalitu maskovat natolik, že se ukazatele kvality modelu s jedním substantiálním faktorem nacházejí mimo uznávané limity. Právě takovou situaci může implementace ARS faktoru odhalit. Dalším pozitivním efektem modelování ARS je očištění faktorových zátěží mezi měřenými proměnnými a latentní proměnnou konstruktů o vliv ARS. Faktor ARS totiž z měřených položek odčerpává varianci odpovídající souhlasu bez ohledu na formulaci položky, takže faktorové zátěže pak lépe odrážejí vztah proměnných k měřenému konstruktu. Velké využití může najít modelování RS v mezinárodně – komparativní analýze invariance měření; systematický chybový rozptyl stylu odpovídání totiž v datech z některých zemí může působit natolik rušivě, že multigroup model vyhodnotí faktorové zátěže jako variantní i přes to, že vztahy mezi měřeným konstruktem a jeho indikátory v jednotlivých zemích jsou ve skutečnosti stejné.

Příklad takové analýzy ukazují Welkenhuysen – Gybels, Billiet a Cambré (2003), kteří RS faktor modelují jako invariantní mezi jednotlivými evropskými zeměmi.

K čemu je dobré modelování RS faktoru v *metodologické analýze*? Můžeme díky němu zjistit, zda mají respondenti v dané populaci tendenci k ARS nebo DRS, a odhalit sílu vlivu stylu odpovídání. Můžeme zkoumat, zda některé nástroje či výzkumná témata mají spíše větší či menší tendenci k tomu, vzbuzovat v respondentech RS, nebo se můžeme zabývat tím, zda jsou lidé náchylní k RS spíše starší či mladší nebo spíše méně či více vzdělaní či zda hrají roli nějaké další demografické charakteristiky, jako např. příjem nebo pohlaví. Můžeme také zkoumat, zda má RS v dané populaci longitudinální charakter, či nikoliv, nebo odhalovat specifika RS v různých kulturách. Možnosti výzkumu RS s využitím CF-RS modelu jsou zkrátka poměrně široké.

5. 6 Další metody identifikace odchylky RS

Kromě výše popsaného principu lze v odborné literatuře najít ještě tři metody ošetření či identifikace vlivu stylu odpovídání v datech. První z nich předpokládá stejně jako CF-RS model použití vyvážené baterie položek a spočívá v eliminaci chyby vzniklé v důsledku ARS překódováním negativně formulovaných položek baterie (Billiet, McClendon, 2000; Mirowsky, Ross, 1991). Druhý způsob doporučuje výpočet sumačního indexu všech souhlasných odpovědí, které respondent poskytl (Johnson et al., 2005; Welkenhuysen – Gybels, Billiet, Cambré, 2003). Třetí způsob využívá speciální baterii položek, která je ukazatelem RS (Van Vaerenbergh, Thomas 2013; Billiet, Davidov, 2008).

První uvedený způsob se používá především pro vypořádání se s ARS. Užití nachází zejména v psychometrii při práci s dlouhými testy, tj. bateriemi s velkým množstvím položek. Na rozdíl od ostatních uvedených metod nedokáže tato procedura identifikovat či odhadnout proporce či sílu vlivu ARS, ale snaží se jen eliminovat či „neutralizovat“ vliv odchylky ARS. Metodu lze aplikovat jen na data z vyvážené baterie položek a postup eliminace ARS je následující: data pro položky formulované proti konstrukt se překódují, aby nejnižší hodnota byla nejvyšší a nejvyšší hodnota nejnižší (*reverse – coding*). Ze všech položek baterie se pak vypočítá sumační index, který v dalších analýzách zastupuje měřený konstrukt. Výzkumnice a výzkumníci, kteří tento postup volí, o takto vypočítaném indexu soudí, že je zbaven vlivu ARS, neboť předpokládají, že souhlasný styl odpovídání, který ovlivnil pro – konstruktové položky je díky překódování „neutralizován“ souhlasným stylem odpovídání, který účinkoval na proti – konstruktové položky (Billiet, McClendon,

2000). Metoda eliminace ARS překódováním ovšem není vědeckou komunitou přijímána jednoznačně pozitivně. Někteří upozorňují na to, že takto upravené škály mají nižší reliabilitu Cronbachova alfa (Billiet, Davidov, 2008), nebo že metoda mylně předpokládá, že negativně formulované položky jsou měřeným konstruktem a ARS ovlivněny stejně jako kladně formulované položky²¹ (Billiet, McClendon, 2000; Mirowsky, Ross, 1991). Nejzásadnější výtkou zůstává fakt, že metoda nedokáže odhadnout, do jaké míry bylo vlastně měření stylem odpovídání ovlivněno.

Druhá uvedená metoda, tedy odhad RS ze sumačního indexu, je metodology rovněž kritizována. Sumační index jako ukazatel ARS je vytvořen tak, že se sečtou všechny kladné odpovědi respondenta na položky v baterii či v celém dotazníku (případně se sčítají záporné odpovědi; DRS či střední odpovědi; MRS); metoda se využívá i v analýze extrémního stylu odpovídání; ERS). Výsledný sumační index je mírou RS sám o sobě; čím vyšší index, tím je respondent náchylnější ke zkoumanému RS. Častým a oprávněným bodem kritiky tohoto přístupu je upozornění, že v případech, kdy je metoda aplikována na data z nevyvážené baterie či na celý dotazník, neumí rozlišit mezi obsahem a stylem, tj. nedokáže identifikovat, zda je souhlasná odpověď reflexí měřeného konstruktů či projevem stylu odpovídání (Welkenhuysen – Gybels, Billiet, Cambré, 2003). Studie, které používají sumační indexy ARS na datech z nevyvážených baterií proto nejsou důvěryhodné analýzy stylu odpovídání.

Třetí metodou, jejíž použití doporučují např. van Vaerenbergh a Thomas (2013), je speciální míra stylu odpovídání, kterou je samostatná baterie o deseti až čtrnácti položkách. V anglicky psané odborné literatuře je metoda nazývána zkratkou RIRS, odpovídající názvu *representative indicators of response style*. Tuto baterii tvoří maximálně nesourodý set položek, kde každá položka je svým významem co nejvíc nezávislá na všech ostatních. Za tímto designem stojí myšlenka, že takový set položek nemá žádnou společnou varianci reprezentující substanciální konstrukt, a položky tak teoreticky nemají společný obsahový faktor. Případně nalezená společná variance je proto považována za reflexi stylu

²¹ Upozorňuji na to, že výtku o nestejnosti vlivu ARS na pro – konstruktové a proti – konstruktové položky formulují výzkumníci, kteří pro odhad ARS používají CF-RS model, ve kterém, jak ukazují v popisu restrikcí modelu v textu dále, stanovují faktorové zátěže mezi latentní proměnnou RS a jejími indikátory jako stejné pro všechny indikátory. Tyto restriktce znamenají, že se předpokládá, že všechny položky, bez ohledu na to, zda jsou formulovány pro či proti konstrukt, jsou RS faktorem ovlivněny stejně. Tito autoři tak kritizují předpoklad, který má rovněž metoda, kterou používají.

odpovídání (Billiet, Davidov, 2008). ARS pro jednotlivého respondenta je pak zjištěna jako suma souhlasů s položkami baterie RIRS, zjišťovány ale mohou být i ostatní RS včetně ERS. Van Vaerenbergh a Thomas (2013) navzdory tomu, že tuto metodu uvádějí jako nejvhodnější možnou, dodávají, že metoda má omezení v tom, že pro měření RS je potřeba přidávat další položky do dotazníku.

5. 7 Empirická část: Modelování ARS jako latentní proměnné v modelu konfirmativní faktorové analýzy

K analýze českého ARS jsem použila metodu popsanou v teoretické části této kapitoly, kterou pracovně nazývám CF-RS model. Cílem studie je představení metody modelování souhlasného stylu odpovídání v CF-RS modelu a získání informací o charakteru českého ARS. V analýze se soustředím na ověření, zda ARS faktor identifikovaný v českých datech splňuje kritéria pro identifikaci faktoru ARS v CFA modelu, daná v publikaci Billieta a McClendona (2000). Jedním z těchto kritérií je korelace mezi ARS a vzděláním, resp. věkem, která byla nalezena v několika empirických studiích v západní Evropě (Meisenberg, Williams – Shillingford, 2008; Weijters, Geuens, and Schillewaert, 2010b; Greenleaf, 1992a; Ross, Mirowsky, 1984). Studie pracuje s předpokladem, že v České republice jako postkomunistické zemi může být doložení této korelace problematické. V navazující analýze zjišťuji, zda ve vybraných západoevropských zemích přetrvává zkoumaná vlastnost ARS faktoru, tj. jeho korelace s věkem a vzděláním. K tomuto účelu jsou použita data z Belgie a Francie. V první uvedené zemi byla korelace mezi ARS faktorem a vzděláním, resp. věkem již dříve doložena (Billiet, McClendon, 2000), druhá uvedená země byla vybrána jako Belgií kulturně nejbližší země.

5. 7. 1 Popis Modelu CF-RS

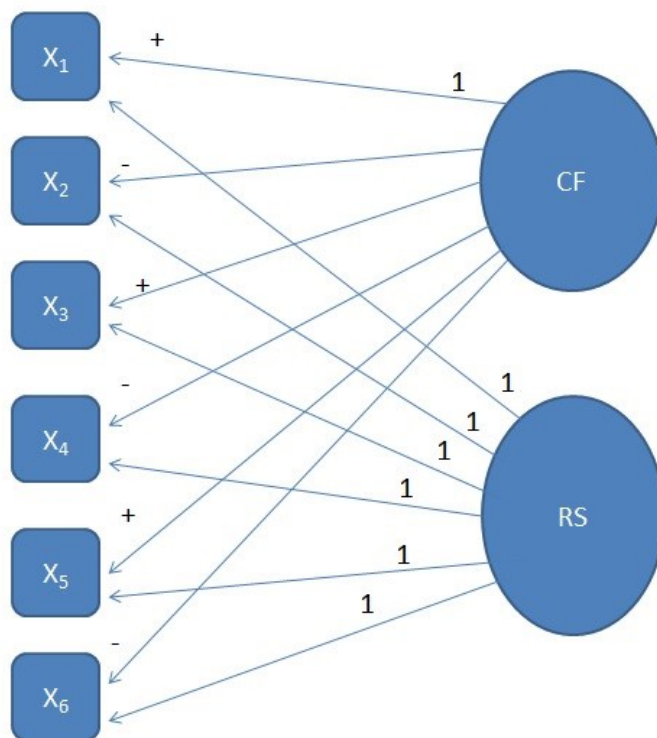
Model CF-RS je specifikován v popisu níže, Obrázek 5.1 ukazuje rozložení proměnných a parametrů v modelu.

- 1) V CF-RS modelu jsou dvě latentní proměnné, první odpovídá měřenému konstrukt (CF), druhá souhlasnému stylu odpovídání (RS). Prvnímu faktoru se obvykle říká obsahový faktor (content), druhému faktor stylu (style).
- 2) Každý faktor ovlivňuje všechny měřené položky v baterii. Všechny faktorové zátěže od obsahového faktoru jsou ponechány volné s výjimkou standardní fixace první položky na hodnotu 1. Všechny faktorové zátěže od faktoru stylu jsou fixovány na hodnotu 1, a to proto, že teoreticky očekáváme, že všechny měřené

položky jsou k ARS náchylné stejně. Welkenhuysen – Gybels, Billiet, Cambré, (2003) k tomuto dodávají, že v případě, kdy jsou některé tyto zátěže ponechány volné, se výrazně snižuje spolehlivost modelu. Svojí roli zde hraje i počet stupňů volnosti; zátěže mezi položkami a ARS faktorem zpravidla nelze ponechat volné, protože by se model stal neidentifikovaným.

- 3) Korelace mezi faktory v CF-RS modelu je zafixována na hodnotu 0, a to ze dvou důvodů, jenž uvádějí autoři, kteří CF-RS model používají; a) korelace mezi faktory není teoreticky očekávatelná, jinými slovy neexistuje důvod, proč by měly faktory mezi sebou korelovat, b) fixace korelace mezi faktory šetří stupně volnosti, které potřebujeme k odhadu shody modelu s daty (Welkenhuysen – Gybels, Billiet, Cambré, 2003).

Obrázek 5.1 CF-RS Model, hypotetický příklad se šesti položkami vyvážené škály



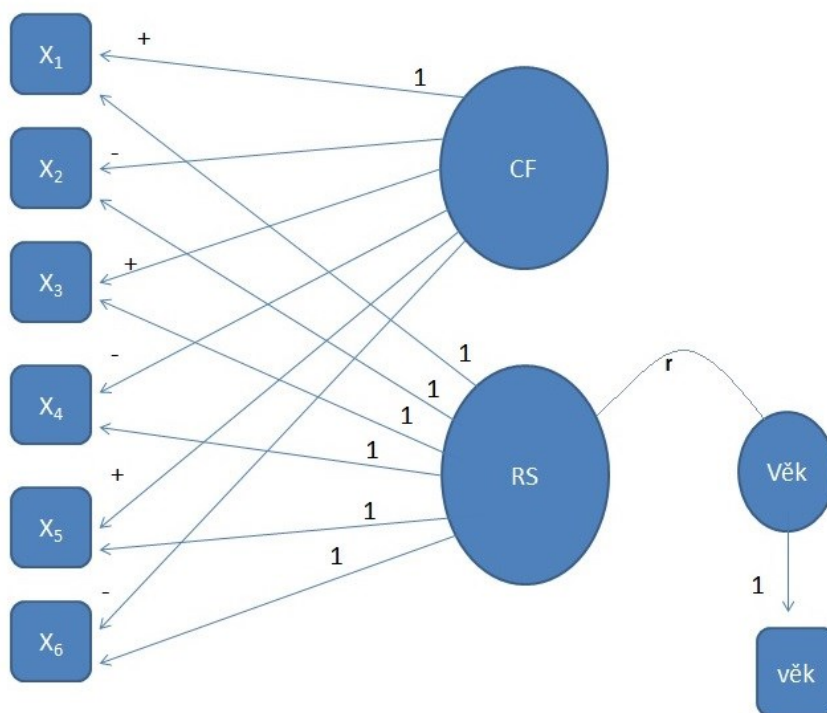
Billiet a McClendon (2000) definují kritéria, na jejichž základě lze rozhodnout, zda lze modelovaný RS faktor označit za varianci odpovídající odchylce ARS. Těmito kritérii jsou:

- 1) Přidání faktoru stylu (RS) do modelu se substanciálním, obsahovým faktorem (CF) musí zlepšit ukazatele shody modelu s daty.
- 2) V CF-RS modelu má RS faktor nenulový rozptyl, který je menší než rozptyl obsahového faktoru²².
- 3) Když do CF-RS modelu přidáme latentní proměnnou věk, je odhadnuta pozitivní korelace mezi věkem a ARS faktorem (viz Obrázek 5.2).
- 4) Když do CF-RS modelu přidáme latentní proměnnou vzdělání, je odhadnuta negativní korelace mezi vzděláním a ARS faktorem.
- 5) Když do CF-RS modelu přidáme latentní proměnnou sumační index souhlasu²³, je odhadnuta vysoká pozitivní korelace mezi indexem souhlasu a ARS faktorem.
- 6) Billiet, Cambré, Welkenhuysen – Gybels (2002) v návaznosti na článek Billieta a McClendona dodávají další kritérium, a to že faktorové zátěže mezi položkami a ARS faktorem musí být menší, než faktorové zátěže mezi položkami a obsahovým faktorem. Toto kritérium vychází z opodstatněného předpokladu, že ARS faktor ovlivňuje měřené proměnné výrazně méně než substanciální konstrukt, který baterie měří.

²² Srovnávají se nestandardizované rozptyly latentních proměnných. Tato poznámka se může zdát zbytečná, neboť standardizovaný rozptyl latentní proměnné v SEM modelu má vždy hodnotu 1, nicméně ji raději uvádím pro ty, kteří by chtěli Model CF-RS aplikovat a kteří se v SEM analýze ještě příliš nevyznají.

²³ Součet kladných odpovědí respondenta na všechny položky baterie.

Obrázek 5.2 CF-RS model s latentní proměnnou věk²⁴



5. 7. 2 Použitá data

Zdrojem dat pro analýzu v této kapitole jsou výběrová šetření projektu International Social Survey Programme (ISSP), a to ISSP Národní identita z roku 2003 a ISSP Národní identita z roku 2013²⁵. Souhlasný styl odpovídání byl modelován na datech z baterie zjišťující postoje k přistěhovalcům, která byla s menšími úpravami použita v obou uvedených vlnách ISSP. Výsledky z dat 2003 a 2013 tak bylo možné porovnat, čímž tato studie do jisté míry získala longitudinální charakter. Možnost srovnání výsledků ze dvou rozdílných období je v analýze výhodou, která zvyšuje její důvěryhodnost.

²⁴ Obdobně vypadá CF-RS model s latentní proměnnou vzdělání a latentní proměnnou indexu souhlasu.

²⁵ Data ISSP jsem vybrala proto, že dotazníky ISSP často obsahují poměrně dlouhé vyvážené baterie položek měřící jeden konstrukt. Nejdříve jsem jako možný zdroj zkoušela i data z mezinárodního projektu European Social Survey (ESS), ale jeho dotazníky obsahují vyvážené baterie v menší míře. Baterie ESS jsou navíc velice krátké, často obsahují jen tři nebo čtyři položky, což není dost k získání řádného konvergujícího řešení CF-RS modelu.

Při přípravě studie jsem analyzovala data pro všechny vyvážené baterie ze všech dostupných českých dat ISSP, tj. z šetření, která proběhla mezi lety 1995 až 2014. Tato rozsáhlá analýza byla nutná z důvodů náročných požadavků na úspěšnou SEM analýzu. V modelování strukturálních rovnic bývá docela často náročné najít řádné konvergující řešení, jak uvádím v Kapitole 3 této dizertační práce. I v případě analýz s CF-RS modelem se ukázalo, že naprostá většina modelů aplikovaných na data nekonverguje nebo hlásí výsledky s Heywoodovými případy. Důvodem je pravděpodobně fakt, že CF-RS model je poměrně složitý v tom, že každá položka je ovlivňována oběma latentními proměnnými v modelu, a situace se stává ještě složitější v případě, kdy je do CF-RS modelu přidána latentní proměnná věku, resp. vzdělání, resp. indexu souhlasu, která má jen jeden indikátor (viz Obrázek 5.2). Software si zřejmě s takovým modelem nedokáže dobře poradit v případě, kdy se vyskytnou problémy typické pro nekonvergence a Heywoodovy případy, tedy nízké faktorové zátěže mezi měřenými a latentními proměnnými či nízké korelace mezi latentními proměnnými v modelu nebo nízké hodnoty kovariancí v matici vstupující do analýzy. Nakonec se ukázalo, že jediná česká data, pro která bylo možné všechny popsané modely odhadnout, byly dvě téměř totožné baterie z šetření z roku 2003 a 2013. Shodou okolností se jedná o baterii, jejíž podobné verze použili autoři, jejichž práce jsou v tomto textu citovány (Billiet, McClendon, 2000; Billiet, Cambré, Welkenhuysen – Gybels, 2002; Welkenhuysen – Gybels, Billiet, Cambré 2003; Billiet, Davidov, 2008).

Výčet položek, které byly součástí použitých baterií, obsahují Tabulky 5.2 a 5.3. Za každým výrokem baterie je uvedeno, zda je položka formulována směrem k měřenému konstrukt (P) nebo proti němu (N). K baterii z ISSP 2003 dodávám, že prezentované položky nebyly součástí jedné baterie, ale jedná se o spojení pěti – položkové baterie (A10) a části baterie (A15). Položky v61 a v62 z ISSP 2003 jsou ovšem ekvivalentní položkám v53 a v54 z baterie ISSP 2013; v šetření ISSP 2003 byly tyto položky z nějakého důvodu součástí jiné baterie a k jejich sloučení s položkami A10 došlo až v dotazníku ISSP 2013. Vzhledem k povaze výroků v obou bateriích lze očekávat, že baterie měří jeden konstrukt, a to postoje k přistěhovalcům. Všechny výroky v obou bateriích byly měřeny na pětibodové Likertově škále odpovědi s hodnotami od 1 „naprosto souhlasím“ po 5 „naprosto nesouhlasím“.

Tab. 5.2 Baterie Postoje k přistěhovalcům, ISSP 2003

| | |
|-----|---|
| v50 | A101 Přistěhovalci zvyšují míru kriminality. (N) |
| v51 | A102 Přistěhovalci jsou obecně pro českou ekonomiku přínosem. (P) |
| v52 | A103 Přistěhovalci berou práci lidem, kteří se narodili v České republice. (N) |
| v53 | A104 Přistěhovalci zušlechťují českou společnost tím, že přinášejí nové myšlenky a kulturu. (P) |
| v54 | A105 Vláda vynakládá na pomoc přistěhovalcům příliš mnoho peněz. (N)* |
| v61 | A153 Legální přistěhovalci do České republiky, kteří nejsou občany, by měli mít stejná práva jako občané České republiky. (P) |
| v62 | A154 Česká republika by měla zavést přísnější opatření, aby zabránila ilegálnímu přistěhovalectví. (N) |

Tab. 5.3 Baterie Postoje k přistěhovalcům, ISSP 2013

| | |
|-----|---|
| V48 | Přistěhovalci zvyšují míru kriminality. (N) |
| V49 | Přistěhovalci jsou obecně pro českou ekonomiku přínosem. (P) |
| V50 | Přistěhovalci berou práci lidem, kteří se narodili v České republice. (N) |
| V51 | Přistěhovalci zlepšují českou společnost tím, že přinášejí nové myšlenky a kulturu. (P) |
| V52 | Obecně řečeno, přistěhovalci podkopávají českou kulturu. (N)* |
| V53 | Legální přistěhovalci do České republiky, kteří nejsou jejími občany, by měli mít stejná práva, jako občané ČR. (P) |
| V54 | Česká republika by měla přijmout tvrdší opatření na vyhoštění ilegálních přistěhovalců. (N) |
| V55 | Legální přistěhovalci by měli mít stejný přístup k veřejnému vzdělání jako občané České republiky. (P)* |

Pozn. Tab. 5.2 a 5.3; Položky jsou označeny názvy, které mají v originálních datových souborech ISSP 2003 a 2013.

Pozn. Tab. 5.2 a 5.3; Hvězdičkou (*) jsou označeny položky, které byly v baterii jen v jedné vlně.

5. 7. 3 Výsledky analýzy

K analýze jsem použila software RStudio (RstudioTeam, 2015) a R package lavaan (Rosseel, 2012). Pro odhad všech modelů jsem použila metodu odhadu parametrů Satorra –

Bentler Mean Adjusted Maximum Likelihood (MLM). Tato metoda je pro analýzu vybraných dat vhodnější než běžně používaný Maximum Likelihood, neboť je učena pro abnormálně rozložená data. Běžně je tato metoda označována jako metoda robustního odhadu (Kline, 2005). Následující text přináší výsledky analýzy českých dat v šetřeních ISSP v roce 2003 a 2013 a francouzských a belgických dat ze stejných šetření.

5. 7. 3. 1 Výsledky, česká data ISSP 2003

Model 1 obsahuje jednu latentní proměnnou odpovídající měřenému konstrukt. Tato latentní proměnná ovlivňuje všech sedm položek baterie, její rozptyl je ponechán volný. Faktorová zátěž první položky (v50) je standardně fixována na hodnotu 1, ostatní zátěže jsou ponechány volné. Tabulka 5.4 ukazuje, že Model 1 s minimálním množstvím restrikcí má sice nedostačující, ale celkem slušnou shodu modelu s daty (CFI 0,888; RMSEA 0,078). Model 2 je vhnížděný model Modelu 1, který navíc obsahuje latentní proměnnou ARS (viz Obrázek 5.1). Všechny faktorové zátěže od tohoto faktoru jsou fixovány na hodnotu 1. Ukazuje se, že tento model má výrazně lepší shodu s daty než Model 1 (CFI 0,961, RMSEA 0,047); Model 2 lze považovat za dobře padnoucí model, neboť i hodnota ukazatele P value (0,579) je nadstandardně vysoká. Nic nebrání konstatovat, že implementace faktoru ARS zlepšila shodu modelu s daty, čímž byla splněna první podmínka pro identifikaci ARS faktoru.

Z Modelu 2 lze odečíst odhad variance latentních proměnných; odhad rozptylu obsahového faktoru (CF) činí 0,352**, rozptyl ARS je 0,05**. Faktor ARS tedy splňuje další kritérium identifikace, tj. má nenulový rozptyl, který je menší než rozptyl substantiálního faktoru.

V následujících modelech jsou prověřovány korelace faktoru ARS s teoreticky očekávanými koreláty.

Tab. 5.4 Data ČR, N=945 případů; ISSP 2003; MLM metoda odhadu; přehled hodnot ukazatelů shody modelu s daty

| | Chi ² | df | CFI | RMSEA | 90%int RMSEA | P value RMSEA |
|---------|------------------|----|-------|-------|--------------|---------------|
| Model 1 | 93,646 | 14 | 0,888 | 0,078 | 0,064; 0,092 | 0,000 |
| Model 2 | 40,628 | 13 | 0,961 | 0,047 | 0,032; 0,063 | 0,579 |
| Model 3 | 86,374 | 18 | 0,951 | 0,063 | 0,051; 0,076 | 0,038 |
| Model 4 | 52,027 | 18 | 0,949 | 0,048 | 0,034; 0,063 | 0,557 |
| Model 5 | 49,087 | 18 | 0,958 | 0,043 | 0,030; 0,057 | 0,773 |

Do Modelu 3 byla přidána latentní proměnná sumačního indexu souhlasných odpovědí²⁶ (viz Obrázek 5.2). Odhadnutá korelace mezi ARS faktorem a indexem souhlasu je zde 0,820**²⁷. Tato korelace je bez pochyby silná, což značí, že modelovaný ARS je skutečně reflexí souhlasných odpovědí.

Model 4 a Model 5 prověřovaly velikost korelace mezi ARS a vzděláním, resp. věkem. Příslušné latentní proměnné byly do modelu vloženy stejným způsobem jako index souhlasu. Obě korelace se ukázaly jako velice nízké a nesignifikantní; odhadnutá korelace mezi ARS a vzděláním v Modelu 4 byla -0,01, korelace mezi ARS a věkem v Modelu 5 byla 0,068. Je tedy zřejmé, že ARS faktor v českých datech z roku 2003 nekoreluje s věkem ani se vzděláním.

Tabulka 5.5 níže ukazuje, jak se změnily faktorové zátěže směrem k obsahovému faktoru po přidání faktoru stylu. Změny nejsou zásadního charakteru, nicméně přítomny jsou. Vidíme, že přidáním ARS faktoru se zvýšily hodnoty negativních faktorových zátěží. Tyto zátěže byly v Modelu 1 nižší z toho důvodu, že na položky rušivě působil rozptyl reflektující souhlasné odpovědi na pro i proti – konstruktové položky. Pozitivní faktorové zátěže se naopak o něco snížily. Středem pozornosti je zde i velikost faktorových zátěží k ARS faktoru. Vidíme, že jsou nižší než zátěže k substantiálnímu faktoru, čímž je splněno další kritérium identifikace ARS.

Tab. 5.5 Data ČR, ISSP 2003, standardizované faktorové zátěže latentních proměnných

| | Model 1 | Model 2 | |
|---------|----------------------|----------------------|--------------------|
| Položka | Obsahový faktor (CF) | Obsahový faktor (CF) | Faktor stylu (ARS) |
| v50 | 0,632 | 0,613 | 0,232 |
| v51 | -0,525 | -0,584 | 0,228 |
| v52 | 0,480 | 0,452 | 0,203 |
| v53 | -0,533 | -0,595 | 0,226 |
| v54 | 0,570 | 0,547 | 0,216 |

²⁶ Index souhlasu vznikl jako suma odpovědí, kdy respondent odpověděl na Likertově škále variantou o hodnotě 1, tj. naprosto souhlasím. Vyšší hodnota indexu značí vyšší počet souhlasných odpovědí na položky v baterii.

²⁷ V samotné analýze pracuji s nepřekódovanou Likertovou škálou, tj. odpověď o hodnotě 1 = naprosto souhlasím, odpověď o hodnotě 5 = naprosto nesouhlasím. Latentní proměnná RS je tak tvořena odpověďmi o nižších hodnotách. Když latentní proměnnou RS koreluji s indexem souhlasu, kde vyšší hodnota znamená větší souhlas, věkem, kde vyšší hodnota znamená vyšší věk, a vzděláním, kde vyšší hodnota znamená vyšší vzdělání, jsou v modelu korelace odhadnuty v opačném směru, než jaký uvádí literatura, tj. jako negativní, resp. negativní, resp. pozitivní. Přímou v textu však korelace uvádím v očekávaném směru, tedy jako pozitivní, resp. pozitivní, resp. negativní, neboť nechci čtenáře a čtenářky mást.

| | | | |
|-----|--------|--------|-------|
| V61 | -0,306 | -0,315 | 0,171 |
| V62 | 0,334 | 0,315 | 0,269 |

5. 7. 3. 2 Výsledky česká data ISSP 2013

Modely v této analýze byly specifikovány obdobně jako modely v analýze dat z roku 2003. Model 1 obsahuje pouze obsahový faktor a z Tabulky 5.6 můžeme vyčíst, že se tento model s daty příliš neshoduje (CFI 0,752; RMSEA 0,124). Model 2 je vhnízděný model předchozího modelu, který obsahuje latentní proměnnou ARS a je definován stejně jako model v předchozí analýze. Vidíme, že implementace ARS faktoru zlepšila shodu modelu s daty (CFI 0,828; RMSEA 0,106), nicméně ukazatele shody stále nebyly dostačující. Proto jsem do modelu na základě tzv. modifikačních indexů²⁸ přidala korelaci mezi položkami V53 a V55²⁹, která je navíc teoreticky opodstatněná, neboť tyto položky mají podobný význam. Zahrnutí korelace v Modelu 3 sice zlepšilo shodu s daty (CFI 0,936, RMSEA 0,066) na takovou úroveň, se kterou si již dá pracovat, nicméně je třeba říct, že shoda modelu s daty zdaleka není ideální vzhledem k hodnotě P value (0,001). Závěry činěné na základě tohoto modelu je proto třeba brát s malou rezervou.

Tab. 5.6 Data ČR, N=1724 případů; ISSP 2013; MLM metoda odhadu; přehled hodnot ukazatelů shody modelu s daty

| | Chi2 | df | CFI | RMSEA | 90%int RMSEA | P value RMSEA |
|---------|---------|----|-------|-------|--------------|---------------|
| Model 1 | 547,132 | 20 | 0,752 | 0,124 | 0,116; 0,132 | 0,000 |
| Model 2 | 384,207 | 19 | 0,828 | 0,106 | 0,098; 0,114 | 0,000 |
| Model 3 | 154,77 | 18 | 0,936 | 0,066 | 0,058; 0,075 | 0,001 |
| Model 4 | 317,459 | 24 | 0,906 | 0,084 | 0,077; 0,092 | 0,000 |
| Model 5 | 188,309 | 24 | 0,927 | 0,064 | 0,056; 0,072 | 0,002 |
| Model 6 | 178,785 | 24 | 0,93 | 0,061 | 0,054; 0,069 | 0,007 |

Rozptyl měřeného konstruktů odhadl Model 3 na 0,319**, variance ARS byla odhadnuta na 0,062**. I v tomto případě tedy platí, že rozptyl faktoru stylu odpovídání má nenulový rozptyl, který je menší než rozptyl měřeného konstruktů. Model 4 s latentní proměnnou indexu souhlasu odhadl korelaci mezi ARS faktorem a indexem souhlasu na 0,801**, i v tomto případě se tedy jedná o velmi silnou korelaci, která potvrzuje, že ARS faktor je tvořen kladnými odpověďmi na všechny položky baterie. Model 5 odhadl velice slabou,

²⁸ Modifikační indexy jsou ukazatele generované softwarem pro SEM, které naznačují změnu shody modelu s daty v případě, že se nastavení příslušného parametru změní (Byrne, 2009).

²⁹ Korelace mezi V53 a V55 byla v Modelu 3 odhadnuta na 0,42, tj. byla v očekávaném směru.

avšak signifikantní korelaci mezi ARS faktorem a vzděláním ($r = -0,081^{**}$), naopak v Modelu 6 se ukázalo, že ARS faktor nekoreluje s věkem ($r = 0,034$).

Stejně jako v analýze českých dat z roku 2003 nezpůsobila implementace faktoru RS velkou změnu ve velikosti faktorových zátěží mezi obsahovým faktorem a jeho indikátory; u většiny zátěží se jedná o změny o velikosti několika setin. Velikosti faktorových zátěží mezi indikátory a faktorem souhlasného stylu odpovídání byly srovnatelné se zátěžemi z analýzy dat z roku 2003, většinou se pohybovaly nad hodnotou 0,200.

Tab. 5.7 Data ČR, ISSP 2013, standardizované faktorové zátěže latentních proměnných

| | Model 1 | Model 3 | |
|---------|----------------------|----------------------|--------------------|
| Položka | Obsahový faktor (CF) | Obsahový faktor (CF) | Faktor stylu (ARS) |
| V48 | 0,612 | 0,611 | 0,250 |
| V49 | -0,566 | -0,615 | 0,227 |
| V50 | 0,639 | 0,630 | 0,221 |
| V51 | -0,543 | -0,587 | 0,223 |
| V52 | 0,577 | 0,558 | 0,199 |
| V53 | -0,411 | -0,374 | 0,199 |
| V54 | 0,297 | 0,311 | 0,247 |
| V55 | -0,349 | -0,296 | 0,222 |

5. 7. 3. 3 Porovnání faktorových zátěží od obsahového konstruktů v letech 2003 a 2013

V CF-RS modelu na datech z roku 2003 lze vidět, že faktorové zátěže mezi obsahovým faktorem a položkami v61 (*Legální přistěhovalci do České republiky, kteří nejsou občany, by měli mít stejná práva jako občané České republiky*) a v62 (*Česká republika by měla zavést přísnější opatření, aby zabránila ilegálnímu přistěhovalectví*) jsou poměrně nízké, a proto by měly být v případné substanciální analýze vyloučeny jako nekvalitní ukazatele měřeného konstruktů. Obdobně v datech z roku 2013 nedostatečně sytí položky v54 (*Česká republika by měla přijmout tvrdší opatření na vyhoštění ilegálních přistěhovalců*) a v55 (*Legální přistěhovalci by měli mít stejný přístup k veřejnému vzdělání jako občané České republiky*). Všimněme si, že formulace položek v62 (2003) a v54 (2013) mají podobný význam. To, že je model v obou případech vyhodnotil jako špatné indikátory, značí, že tyto položky nejsou vhodné k měření takto operacionalizovaného konstruktů. Když se podíváme na ostatní zátěže od obsahového faktoru v obou souborech, vidíme, že položky, jejichž formulace jsou stejné či jsou si velice podobné, v obou vlnách sytily faktor podobně. Výjimkou jsou položky v52 (2003) a v50 (2013); tyto položky mají shodné znění „*Přistěhovalci berou práci lidem, kteří se narodili v České republice*“, přesto se ve vlnách

liší (0,452** (2003), 0,630** (2013)). Během deseti let se tedy tento výrok zřejmě stal silnějším indikátorem měřeného konstruktů.

5. 7. 3. 4 Využití modelování faktoru stylu odpovídání v substanciální analýze

Výše uvedené modely s faktorem stylu lze využít v substanciální analýze, která využívá metodu modelování strukturálních rovnic. Kdybychom měli SEM model vztahů mezi latentními proměnnými, z nichž by všechny nebo některé měly více indikátorů, tj. měřených proměnných, mohlo by očištění latentní proměnné konstruktů Postoje k přistěhovalcům o vliv ARS ovlivnit velikost regresních vztahů mezi touto latentní proměnnou a dalšími latentními proměnnými. Vzhledem k relativně malému rozptylu faktoru ARS v modelovaných případech by však tento vliv byl minimální, a pravděpodobnou pozitivní stránkou implementace faktoru ARS by bylo spíše zlepšení shody modelu s daty, než významné změny ve vztazích mezi latentní proměnnými.

5. 7. 4 Analýza dat z Belgie a Francie a srovnání s výsledky z České republiky

Korelace mezi faktorem ARS a indexem souhlasu, resp. vzděláním, resp. věkem v CF-RS modelu byly doloženy v článku Billieta a McClendona (2000) na souboru belgických dat z roku 1995; $r(\text{ARS}; \text{index souhlasu})=0,900^{**}$; $r(\text{ARS}; \text{vzdělání}) = -0,230^{**}$; $r(\text{ARS}; \text{věk}) = 0,150^{**}$). Je však otázkou, zda jsou tyto korelace dlouhodobou charakteristikou belgického ARS, kterou můžeme stále pozorovat, či zda se jednalo o ojedinělý nále. Odpověď jsem hledala v analýze, jejímž cílem bylo ověřit, zda zmíněné korelace v Belgii stále existují. Výhodou je, že původní analýza z části proběhla na podobné baterii, proto jsou moje a původní výsledky do velké míry srovnatelné. Z důvodů omezení daných zdrojem dat, jsem mohla analyzovat belgická data jen v šetření ISSP 2013, v předchozí vlně Národní identity totiž Belgie chyběla. V ISSP 2003 jsem proto použila data z Francie, tj. Belgii kulturně nejbližší země.

5. 7. 4. 1 Výsledky Francie ISSP 2003

Na data z Francie i Belgie jsem použila stejný postup jako v analýze českých dat. Tabulka 5.8 poskytuje přehled ukazatelů shody modelu s daty ve francouzských datech, které jsou dobré. Model 1 obsahuje jen faktor konstruktů, Model 2 má navíc faktor ARS. Do Modelu 3 byla navíc dodána teoreticky opodstatněná korelace mezi položkami v51 a v53, která zlepšila shodu modelu s daty³⁰. V tomto modelu má rozptyl obsahového faktoru hodnotu 1,176**, rozptyl faktoru ARS 0,026**, obsahový faktor je tedy výrazně větší, což se

³⁰ Hodnota této korelace v odhadnutém Modelu 3 činila 0,287, tj. byla v očekávaném směru.

shoduje s kritérii pro identifikaci faktoru ARS. Všimněme si toho, že velikost rozptylu faktoru ARS ve francouzských datech je menší než velikost rozptylu v českých datech z roku 2003. V Modelu 4 koreluje ARS faktor s indexem souhlasu poměrně silně ($r = 0,632^{**}$), stejně tak korelace ARS faktoru se vzděláním a věkem jsou silné (Model 5: $r(\text{ARS}; \text{vzdělání}) = -0,170^{**}$; Model 6: $r(\text{ARS}; \text{věk}) = 0,395^{**}$). Odhadnuté parametry CF-RS modelu tedy prokázaly, že jsou splněna všechna kritéria pro identifikaci ARS faktoru.

Tab. 5.8 Data Francie, N= 1398 případů; ISSP 2003; MLM metoda odhadu; přehled hodnot ukazatelů shody modelu s daty

| | Chi2 | df | CFI | RMSEA | 90%int RMSEA | P value RMSEA |
|---------|---------|----|-------|-------|--------------|---------------|
| Model 1 | 213,663 | 14 | 0,957 | 0,101 | 0,090; 0,112 | 0,000 |
| Model 2 | 117,316 | 13 | 0,978 | 0,076 | 0,065; 0,088 | 0,000 |
| Model 3 | 84,452 | 12 | 0,984 | 0,066 | 0,054; 0,078 | 0,015 |
| Model 4 | 108,779 | 17 | 0,982 | 0,062 | 0,052; 0,073 | 0,028 |
| Model 5 | 113,688 | 17 | 0,978 | 0,066 | 0,056; 0,077 | 0,006 |
| Model 6 | 148,277 | 17 | 0,973 | 0,074 | 0,064; 0,085 | 0,000 |

5. 7. 4. 1 Výsledky Belgie ISSP 2013

Shody modelů aplikovaných na belgická data jsou opět dobré, neboť se nacházejí v mezích daných odbornou literaturou (viz Tabulka 5.9). Stejně jako v předchozím případě zlepšila implementace faktoru ARS shodu modelu s daty. Faktor ARS odhadnutý v Modelu 2 má rozptyl o velikosti 0,055, obsahový faktor v tom samém modelu má rozptyl o hodnotě 0,754, i v tomto případě je tedy potvrzeno kritérium pro identifikaci ARS. V Modelu 3 koreluje faktor ARS s indexem souhlasu rovněž celkem silně ($0,521^{**}$) a Model 4 a Model 5 doložily korelaci ARS se vzděláním a věkem; $r(\text{ARS}; \text{vzdělání}) = -0,121^{**}$; $r(\text{ARS}; \text{věk}) = 0,161^{**}$. Analýza tedy ukázala, že faktor ARS v belgických datech z roku 2013 má vlastnosti, které byly identifikovány již v datech z roku 1995.

Tab. 5.9 Data Belgie, N= 1696 případů; ISSP 2013; MLM metoda odhadu; přehled hodnot ukazatelů shody modelu s daty

| | Chi2 | df | CFI | RMSEA | 90%int RMSEA | P value RMSEA |
|---------|---------|----|-------|-------|--------------|---------------|
| Model 1 | 284,049 | 20 | 0,938 | 0,088 | 0,081; 0,096 | 0,000 |
| Model 2 | 120,164 | 19 | 0,976 | 0,056 | 0,048; 0,065 | 0,111 |
| Model 3 | 206,617 | 25 | 0,959 | 0,065 | 0,058; 0,073 | 0,000 |
| Model 4 | 161,267 | 25 | 0,97 | 0,057 | 0,050; 0,065 | 0,060 |
| Model 5 | 167,357 | 25 | 0,969 | 0,058 | 0,051; 0,066 | 0,036 |

Tabulka 5.10 ukazuje standardizované zátěže mezi latentními a měřenými proměnnými v modelech aplikovaných na francouzská a belgická data. Zde si můžeme všimnout, že ve francouzských datech je vliv ARS na položky poměrně malý, a že zátěže k obsahovému faktoru jsou faktorem stylu ovlivněny velice málo. V porovnání s výsledky z českých dat je vliv na velikost těchto zátěží menší. Vliv ARS v datech z Belgie je o něco větší než v datech z Francie, nicméně ani zde nevidíme zásadní rozdíly mezi zátěžemi ke konstrukturu v modelu bez ARS faktoru a v modelu s ARS faktorem. V obou frankofonních zemích jsou tedy zátěže k substantiálnímu konstrukturu stylu odpovídání ovlivněny méně než v České republice.

Tab. 5.10, ISSP 2003, 2013, Francie, Belgie, standardizované faktorové zátěže latentních proměnných

| Šetření | Francie, ISSP 2003 | | | Belgie, ISSP 2013 | | | |
|---------|----------------------|----------------------|--------------------|-------------------|----------------------|----------------------|--------------------|
| | Model 1 | Model 3 | | Model 1 | Model 2 | | |
| Položka | Obsahový faktor (CF) | Obsahový faktor (CF) | Faktor stylu (ARS) | Položka | Obsahový faktor (CF) | Obsahový faktor (CF) | Faktor stylu (ARS) |
| v50 | 0,782 | 0,786 | 0,116 | V48 | 0,746 | 0,746 | 0,202 |
| v51 | -0,701 | -0,679 | 0,137 | V49 | -0,734 | -0,748 | 0,216 |
| v52 | 0,740 | 0,734 | 0,118 | V50 | 0,677 | 0,671 | 0,194 |
| v53 | -0,745 | -0,731 | 0,129 | V51 | -0,728 | -0,744 | 0,207 |
| v54 | 0,812 | 0,815 | 0,118 | V52 | 0,765 | 0,765 | 0,189 |
| V61 | -0,576 | -0,586 | 0,106 | V53 | -0,597 | -0,607 | 0,19 |
| V62 | 0,703 | 0,707 | 0,126 | V54 | 0,629 | 0,627 | 0,205 |
| | | | | V55 | -0,473 | -0,471 | 0,232 |

5. 8 Diskuze

Modelování faktoru ARS v českých datech baterie zjišťující postoje k přistěhovalcům v šetřeních ISSP 2003 a 2013 přineslo nové poznatky o charakteru českého ARS.

Vzhledem ke shodě modelů s daty lze tvrdit, že výsledky jsou důvěryhodné, ovšem s upozorněním, že v případě dat z roku 2013 byla shoda o něco horší a ne zcela ideální.

Parametry CF-RS modelů v datech z obou vln z velké části splňovaly požadavky dané kritérii pro identifikaci ARS faktoru. Přidání ARS faktoru do modelu s jedním obsahovým faktorem zlepšilo shodu modelu s daty, faktor ARS měl nenulový rozptyl menší než obsahový faktor, faktorové zátěže od ARS modelu byly menší než faktorové zátěže od obsahového faktoru a korelace faktoru ARS s latentní proměnnou indikující počet souhlasných odpovědí respondenta byla vysoká. Velikost faktorových zátěží mezi

proměnnými a obsahovým faktorem se po implementaci ARS faktoru změnila, ale ne výrazně. Praktickým přínosem CF-RS modelu pro substantiální analýzu je v obou analyzovaných případech zlepšení shody modelu s daty a získání možnosti modelovat vztahy mezi obsahovým konstruktem a případnými dalšími konstrukty či měřenými proměnnými, které budou očištěny o slabý efekt ARS.

Analýza českých dat se zaměřila především na ověření dvou specifických kritérií identifikace ARS faktoru, kterými je korelace ARS faktoru se vzděláním, resp. věkem. Uvedené korelace se v podstatě nepodařilo nalézt. V datech z roku 2013 sice byla nalezena signifikantní korelace ARS faktoru se vzděláním ($r = -0,081^{**}$), byla ovšem velice slabá a je otázkou, zda tak velká korelace, jejíž signifikance byla doložena na nadstandardně velkém souboru (1724 případů) je i věcně významná. Tato korelace rovněž pochází z modelu, který neměl zcela ideální shodu s daty.

Navazující analýza prokázala, že v datech z Belgie a Francie jsou splněna všechna kritéria identifikace ARS faktoru, včetně předpokládaných korelací ARS faktoru se vzděláním a věkem. Ve francouzských datech ISSP z roku 2003 koreloval faktor ARS s indexem souhlasu silně ($r = 0,635^{**}$), stejně tak byly přítomné korelace se vzděláním ($-0,170^{**}$) i s věkem ($0,395^{**}$). V datech z Belgie ze šetření ISSP 2013 existovala poměrně silná korelace ARS faktoru s indexem souhlasu ($r = 0,521^{**}$), korelace se vzděláním a s věkem byla podstatná a signifikantní ($r = -0,121^{**}$, resp. $r = 0,161^{**}$).

Při interpretaci výsledků z Česka, Francie a Belgie je nutné zohlednit vliv umístění baterie položek v dotazníku. Respondenti se v průběhu rozhovoru unavují a ztrácejí pozornost; čím je dotazník delší, tím jsou unavenější, méně se soustředí a do odpovídání vkládají méně energie. Vzhledem k tomu, že únava a snížená pozornost může mít vliv na inklinaci respondentů ke stylu odpovídání, je potřeba vědět, zda se umístění v dotazníku v jedné zemi výrazně nelišilo od umístění v dotazníku v jiné zemi. V českém a francouzském dotazníku Národní identity 2003 bylo umístění škály postojů k přistěhovalcům stejné a rovněž realizace šetření byla unikátní; v jeho rámci nebyla sbírána data pro nějaký jiný projekt, nejednalo se tedy o omnibusové šetření (Scholz, Harkness, Faab, 2005). Respondenti v obou zemích tak před odpovídáním na analyzovanou škálu shodně odpovídali na cca 50 otázek či výroků, průměrně tedy byli vystaveni zhruba stejné kognitivní zátěži. V případě analýzy dat Národní identity 2013 z Česka a Belgie byla situace jiná; výzkumníci v Belgii sbírali data pro šetření Národní identita 2013 a Rodina

2012 dohromady, přičemž položky modulu Rodina v dotazníku předcházely položkám Národní identity (Joye, Sapin, 2016). Belgičtí respondenti tak byli v porovnání s českými respondenty unavenější, neboť před zkoumanou škálou odpověděli na cca 110 otázek a výroků, zatímco Češi za sebou měli jen cca 40 otázek. Vliv únavy mohl hrát roli v porovnání vztahu mezi ARS a vzděláním, resp. věkem. Za předpokladu, že jsou starší a méně vzdělaní lidé náchylnější k únavě a ztrátě pozornosti, nemůžeme si být jisti tím, zda se ve vztahu ARS a vzdělání, resp. věku skutečně Česká republika a Belgie liší. Spolehlivé informace o rozdílech mezi českým a belgickým ARS a jeho vlastnostech bychom mohli získat, jen pokud bychom měli další data, ve kterých by nebyl přítomen vliv umístění položky v dotazníku, tj. potřebovali bychom měření z vyvážené škály, jež byla v obou zemích umístěna na stejném místě v dotazníku.

Na okraj lze podotknout, že výsledky analýzy ze všech souborů naznačují, že francouzský a belgický ARS je o něco slabší, než ten český. V datech z Francie a Belgie mělo zahrnutí vlivu ARS do modelu minimální vliv na velikost faktorových zátěží mezi proměnnými a obsahovým faktorem. V českých datech byl tento vliv o něco větší, i když ne zásadní. Porovnání velikostí rozptylů faktorů ARS ve třech analyzovaných zemích ukázalo, že i zde jsou patrné rozdíly; ve Francii je tento rozptyl menší než v Belgii a v Česku. Jistotu ohledně rozdílných variancí RS faktoru a velikosti faktorových zátěží mezi měřeními a latentními proměnnými by přinesla multigroup CFA analýza CF-RS modelu zaměřená na prokázání invariance měření v jednotlivých zemích, která by prokázala, zda se statisticky významně liší vlastnosti RS faktoru v Česku, ve Francii a v Belgii. Tento druh analýzy by byl užitečný také v případě, kdybychom chtěli srovnávat, zda je měření škály postojů k přistěhovalectví ekvivalentní ve všech analyzovaných zemích, jinými slovy, kdybychom chtěli vědět, zda škála měří ve všech zemích stejný konstrukt. Tato analýza by s vysokou pravděpodobností přinesla informaci, že měření konstruktů ve třech analyzovaných zemích se liší, neboť to je patrné z velkých rozdílů ve velikosti faktorových zátěží v odhadnutých modelech (viz Tabulky 5.5, 5.7 a 5.10). Je zjevné, že položky baterie ve francouzských a belgických datech reflektují analyzovaný konstrukt ve větší míře, než v českých datech. To znamená, že měření pomocí zkoumaných baterií je v těchto frankofonních zemích validnější než u nás. Tento jev má dva pravděpodobné důvody, které se sebou vzájemně souvisí; Francie i Belgie jsou země s dlouhou demokratickou tradicí, kde se politický a společenský ideologický diskurz vytvářel poměrně dlouhou dobu a jinak než v postkomunistických zemích; lidé v těchto zemích tak uvažují o společenských

fenoménech jinak, než lidé v Česku; rovněž zkušenost s přistěhovalectvím je v Česku a ve Francii, resp. Belgii naprosto odlišná. Dále je potřeba vzít v úvahu, že většina konstruktů a baterií měřících nejrůznější abstraktní společenské fenomény vznikly v průběhu 20. století právě v západní Evropě, případně v Severní Americe, a není proto divu, že tamější obyvatelé jsou na jejich aplikaci připravenější, než lidé v postkomunistických zemích.

Nejpodstatnější zjištěním analýzy v této kapitole je poznatek, že český ARS zřejmě nekoreluje se vzděláním ani s věkem. Ačkoliv jsou tyto korelace v odborné literatuře uváděny jako dvě z kritérií pro identifikaci ARS faktoru v datech, domnívám se, že v tomto případě není jejich absence znakem toho, že v modelu nebyl identifikován ARS. Při aplikaci modelu na data, která nepocházejí ze západní Evropy, nemusí být porušení těchto kritérií při současném splnění dalších kritérií indikátorem toho, že nebyl objeven ARS, ale může se jednat spíše o nalezené kulturní specifikum. Výsledek méj analyzy tedy interpretuji tak, že český ARS je v tomto ohledu specifický a odlišný od západoevropského.

Jaké mohou být důvody toho, že v Česku tyto vztahy mezi stylem odpovídání a věkem, resp. vzděláním nepozorujeme? Nabízí se možnost, že v postkomunistickém Česku vzdělání nediferencuje mezi méně a více kognitivně schopnými lidmi tak efektivně jako v západní Evropě. Vzdělanost české populace je nepochybně ovlivněna vzdělávacím systémem socialistického Československa, který usiloval o co největší počty lidí s vyučením a co nejmenší počty lidí s vysokoškolským vzděláním. Naopak porevoluční vzdělávací systém se zejména v posledních letech profiluje jako podporující získávání vysokoškolského, především magisterského vzdělání, a vycházející vstříc lidem, kteří chtějí mít místo vyučení maturitu. V západních zemích nedošlo v poslední době k tak velké, nárazové změně v systému vzdělávání, takže se tam neliší šance na získání vzdělání mezi staršími a mladšími lidmi v tak velké míře. Tyto rozdíly mezi západní Evropou a Českem mohly mít vliv na to, že v Česku vzdělání tolik nediferencuje mezi kognitivně méně a více schopnými lidmi. V případě, že by platila výše naznačená hypotéza, by se však měla v českých datech objevit alespoň korelace ARS faktoru s věkem. Lidé ve všech západních zemích se zvyšujícím se věkem do jisté míry přicházejí o kognitivní funkce a je dost nepravděpodobné, že např. Belgičané či Francouzi o ně přicházejí více než Češi.

Absence korelací faktoru ARS s věkem a vzděláním v Česku a jejich přítomnost v západních zemích by mohla ukazovat na jinou příčinu, a to na vliv postmoderního

západoevropského ideologického diskurzu na formování uvažování o společenských fenoménech. To, že položky zkoumané baterie ve Francii a v Belgii měří silněji, tzn. validněji než v Česku, jak vyplynulo z analýzy, je svým způsobem důkaz, že minimálně v případě postojů k přistěhovalectví se ideologické diskurzy v těchto zemích liší, přičemž v západních zemích je zkoumaný konstrukt přítomen silněji. Baterie tedy reflektuje ideologický diskurz spíše západních zemí, než postkomunistického Česka. Budeme-li uvažovat nad vztahem mezi ARS a vzděláním, resp. věkem, nalezeným v západních zemích, mohla by být v západní Evropě neschopnost odpovídat konzistentně s ideologickým principem použitým v baterii generační záležitost týkající se především starších lidí; naopak v Česku, kde dosud lidé nejsou zvyklí uvažovat o společenských problémech způsobem běžným v západní Evropě, by se týkala všech věkových skupin. Tuto hypotézu lze aplikovat i na vztah ARS faktoru a vzdělání; v západní Evropě je vliv ideologického diskurzu do jisté míry podmíněný vzděláním, neboť vzdělanější lidé se tomuto diskurzu spíše naučí, zatímco v Česku vzděláním podmíněn není, neboť příslušný diskurz není obecně rozšířen a tudíž se ho spíše nenaučí ani více vzdělaní lidé. S touto hypotézou se nabízí i úvaha o vlivu sociální desirability na odpovědi, a to zejména ve výzkumu sociálně citlivých témat, jako je např. v této studii analyzované přistěhovalectví. K sociální desirabilitě jsou v těchto případech náchylnější vzdělanější lidé, případně mladší lidé, kteří prezentují odpovědi v souladu s očekáváním, se kterým byly baterie navrženy. Vzhledem k tomu, že v Česku se vyskytuje tzv. politická korektnost minimálně, je vliv sociální desirability v případě citlivých sociálních témat zanedbatelný. V odpovědích se tak mohou častěji objevovat rozpory, vzniklé v důsledku toho, že respondent neaplikuje ucelený ideologický rámec, ale své aktuální rozpoložení nad danými položkami, aniž by uvažoval nad tím, co je vhodné říct.

Dalším vlivem by mohla být zkušenost lidí s rolí respondenta, která se může v západní Evropě a v Česku lišit. Lidé, kteří jsou s rolí respondenta obeznámeni ve větší míře, odpovídají „lépe“, tj. jejich odpovědi mají vnitřní logiku a jsou konzistentní. V západní Evropě by pak podle tohoto vysvětlení byla zkušenost s rolí respondenta záležitostí především mladších lidí a do jisté míry i vzdělanějších lidí, kteří se roli dokáží spíše naučit. V Česku, které je fenoménem výzkumu veřejného mínění zasaženo o něco méně, pak tato „respondentská“ zkušenost není tolik rozšířena a neliší se mezi věkovými a vzdělanostními skupinami. Podle tohoto vysvětlení by pak byli čeští respondenti obecně méně racionální

při volbě odpovědí, než respondenti v západní Evropě, a v datech ze souboru by se tak nacházelo větší množství chyb měření.

Výše uvedené hypotézy do problematiky ARS v Česku mnoho světla nevnášejí a pohybují se spíše v rovině spekulací než seriózního výzkumu. V současné době ovšem nemáme dost informací k tomu, abychom mohly specifitu českého ARS objasnit, a tak je třeba věnovat se dalšímu výzkumu. Ke studiu fenoménu je potřeba analyzovat další, již existující data, a doložit na nich absenci vztahu ARS a vzdělání, resp. věku. Je nezbytné najít další vyvážené baterie měřící jiné konstrukty, než jsou v této kapitole použité Postoje k přistěhovalcům. V české společnosti, která nemá zkušenosti s imigrací, nemusí mít respondenti strukturované postoje k této problematice, a tudíž nemusí řadu výroků analyzované škály chápat jako protikladné. V návaznosti na analýzy jiných škál je také vhodné realizovat unikátní měření, prověřující vztah mezi kognitivními schopnostmi, vzděláním a ARS. Přínosem pro další výzkum by rovněž bylo prověření jiných používaných měř souhlasného stylu odpovídání, např. v tomto textu zmíněné indikátory stylu odpovídání, tzv. RIRS.

6. True Score MTMM metoda odhadu validity a reliability a analýza českých 2 SB MTMM dat z šetření European Social Survey

Tato kapitola se zaměřuje na popis a vysvětlení metod odhadu chyb měření odvozených z principu Multitrait multimethod (MTMM), které plně využívají principů rozšířené Klasické testové teorie. Cílem textu je poskytnout základní uvedení do problematiky sběru a analýzy tzv. 2 split – ballot MTMM dat (2 SB MTMM) a ukázat analýzu validity a reliability měření prostřednictvím tzv. True Score MTMM modelu (TS MTMM model). Tento text by měl zájemcům o studium chyb měření prostřednictvím MTMM přinést všechny potřebné informace nezbytné pro porozumění komplexnímu problému analýzy 2 SB MTMM dat a použití TS MTMM modelu. Specifikem textu je zaměření na problematické aspekty aplikace TS MTMM modelu na 2 SB MTMM data, v teoretické části kapitoly podávám přehled těchto problémů a s pomocí odborné literatury poskytuji jejich zdůvodnění.

Při práci na této kapitole jsem vycházela především z publikací členek a členů výzkumného týmu pracoviště Research and Expertise Centre for Survey Methodology (RECSM) ze španělské Univerzity Pompeu Fabra, který se dlouhodobě podílí na přípravách metodologických šetření European Social Survey (ESS). Výzkumníci RECSM připravují 2 SB MTMM experimenty pro šetření ESS a jsou zodpovědní za zpracování těchto dat. Vzhledem k tomu, jak náročné je získat odhady chyb měření z 2 SB MTMM dat, má RECSM na tuto specifickou analýzu v podstatě monopol. Na základě informací z publikací týmu RECSM jsem mohla sama analyzovat česká 2 SB MTMM data z šetření ESS a získat vlastní zkušenost se všemi analytickými úskalími, které popisuje tým RECSM ve svých publikacích a dalších instruktážních materiálech. Výsledky, které jsem získala z českých 2 SB MTMM dat jsou do velké míry omezené právě z důvodů problémů, které analýza těchto dat obnáší. Tyto omezené výsledky ale nakonec posloužily k názornému srovnání s výsledky týmu RECSM, jenž prezentuji v empirické části kapitoly.

Na následujících stránkách nejdříve popisuji experimentální design MTMM a na něj navazující, upravený design 2 split – ballot MTMM, který byl použit pro sběr experimentálních dat v šetřeních mezinárodního projektu ESS. Text dále obsahuje popis a vysvětlení TS MTMM modelu, který je používán k odhadu koeficientů validity a reliability v MTMM, resp. 2 SB MTMM datech. Empirická část této kapitoly prezentuje výsledky

vlastní analýzy dat z 2 SB MTMM experimentu nazvaného Role mužů a žen ve společnosti, použitého v šetření ESS.

6.1 Multitrait multimethod (MTMM) experiment

Princip metody Multitrait multimethod (MTMM) představili psychologové Campbell a Fiske už na konci padesátých let (1959), jejich tzv. MTMM korelační matice sloužila jako prostředek k určení tzv. konvergentní a diskriminantní validity vybraných dotazníkových otázek. Z jejich myšlenky vzešel MTMM experimentální design, který ve svých studiích validity měření využil Andrews (1984) a stanovil pravidla pro analýzu takových dat v SEM (Andrews, 1984, Saris, Andrews, 1991). Všiml si totiž, že MTMM data mají takový charakter, který umožňuje identifikovat jedinečné zdroje rozptylu měřeného konstruktů a použité metody měření, tedy latentní proměnné konstruktů a metody. Tato vlastnost MTMM dat je dána tím, že set dotazníkových položek, který tvoří MTMM experiment, měří několik konstruktů několika metodami zároveň. Nejčastěji se o MTMM uvažuje jako o devíti proměnných, měřících celkem tři latentní znaky (z angl. *traits*) třemi různými metodami (*methods*). Znaků, tedy reprezentantů měřeného konstruktů, a metod ale může být i více, v tom případě však stoupá počet položek v MTMM experimentu, čímž se design stává velmi obtížně použitelným v reálných situacích.

V klasickém MTMM experimentu musí být všech devět položek zodpovězeno každým jedním respondentem ve výběrovém souboru (to však zcela neplatí pro 2 split – ballot MTMM design, který popisuji v další části kapitoly). Získaná matice korelací resp. kovariancí mezi devíti proměnnými pak slouží jako vstupní data pro získání odhadů kvality. Data jsou analyzována v softwaru pro SEM, jehož prostřednictvím lze získat odhady validity měření (Andrews, 1984), případně odhady validity, reliability a tzv. společného rozptylu metody (Saris, Galhoffer, 2014).

Při měření tří latentních znaků třemi metodami získáváme devět proměnných y_{ij} , kde (Saris, Satorra, Coenders, 2004):

i... latentní znak (trait), který proměnná měří, kde $i = 1;3$,

j... metoda, kterou je proměnná měřena, kde $j = 1;3$.

Metodu MTMM ilustruje následující příklad z druhé vlny šetření ESS. Tento experiment měřil latentní konstrukt Role mužů a žen ve společnosti následujícími třemi latentními znaky:

- Latentní znak F1: Omezení placeného zaměstnání,
- Latentní znak F2 : Odpovědnost za děti a domácnost,
- Latentní znak F3 : Právo na zaměstnání.

Každý z těchto latentních znaků (F_i) byl měřen třemi metodami (M_j), které se v tomto případě od sebe lišily formulací výroků a škálou, na které byl výrok měřen. Dotazníkové položky měřící znaky metodami jsou vypsány v Rámečku 6.1.

Tab. 6.1 Znění dotazníkových položek v MTMM experimentu Role mužů a žen ve společnosti

| |
|---|
| F1M1 Žena by měla být připravena omezit svou placenou práci kvůli rodině. |
| F2M1 Muži by měli přijmout stejnou odpovědnost za domov a děti jako ženy. |
| F3M1 Když je málo pracovních míst, muži by měli mít větší právo na zaměstnání než ženy. |
| Rozhodně souhlasím |
| Spíše souhlasím |
| Ani souhlas ani nesouhlas |
| Spíše nesouhlasím |
| Rozhodně nesouhlasím |
| F1M2 Žena by neměla být nucena omezovat placenou práci kvůli rodině. |
| F2M2 Ženy by měly mít větší odpovědnost za domov a děti než muži. |
| F3M2 Když je málo pracovních míst, ženy by měly mít stejné právo na zaměstnání jako muži. |
| Rozhodně souhlasím |
| Spíše souhlasím |
| Ani souhlas ani nesouhlas |
| Spíše nesouhlasím |
| Rozhodně nesouhlasím |
| F1M3 Žena by měla být připravena omezit placenou práci kvůli rodině 1 2 3 4 5 Žena by neměla být nucena omezovat placenou práci kvůli rodině. |
| F2M3 Muži by měli přijmout stejnou odpovědnost za domov a děti jako ženy 1 2 3 4 5 Ženy by měly mít větší odpovědnost za domov a děti než muži. |
| F3M3 Když je málo pracovních míst, muži by měli mít větší právo na zaměstnání než ženy 1 2 3 4 5 Když je málo pracovních míst, ženy by měly mít stejné právo na zaměstnání jako muži. |

6. 2 Design sběru MTMM dat

Sběr dat pro klasický MTMM experiment je nepraktický s ohledem na reálné schopnosti a možnosti respondentů. Experiment předpokládá, že dotázaní budou opakovaně odpovídat na velice podobné otázky, jak ukazuje příklad v Rámečku 6.1. V tomto MTMM experimentu musí jeden respondent na otázku, zda by žena měla omezit placenou práci kvůli rodině, odpovědět třikrát, a to jednou na kladně formulovaný výrok se škálou „1 – rozhodně souhlasím až 5 – rozhodně nesouhlasím“, podruhé na opačně formulovaný výrok s tou samou škálou odpovědi a potřetí na otázku, kde je pětibodová škála uvedena přímo v těle dotazníkové položky. Vzhledem k tomu, že v každém MTMM experimentu jsou tři znaky, odpovídají respondenti v jednom experimentu celkem na devět otázek. Respondenti jsou při takovém postupu nadměrně zatěžováni a tato zátěž má negativní vliv na kvalitu jejich odpovědí. S touto zátěží se pojí několik problémů. Prvním je ztráta motivace dostatečně se soustředit na rozhovor; tento jev je v anglicky psané odborné literatuře nazýván *satisficing* (Tourangeau, Rips, Rasinski, 2000) a v této dizertační práci jej překládám jako *vyhovění tazateli*. V takovém případě dotazování unavený respondent odpovídá nedbale a jeho odpovědi nemusí reflektovat jeho skutečné postoje či zkušenosti. Druhým problémem týkajícím se klasického MTMM designu je fakt, že respondenti odpovídající na tři vzájemně velmi podobné otázky si svoje předchozí odpovědi mohou pamatovat (Revilla, Saris, 2011a) a své následující odpovědi na podobnou otázku pak už jen zopakují. Opakování odpovědí se v datech projeví jako tzv. korelované chyby měření, které není možné v modelu vztahů MTMM odhalit a které zastírají skutečné chybové rozptyly, vzniklé vlivem použité metody v dotazníkové položce (Saris, Sattora, Coenders, 2004).

6. 3 Design sběru 2 split – ballot MTMM

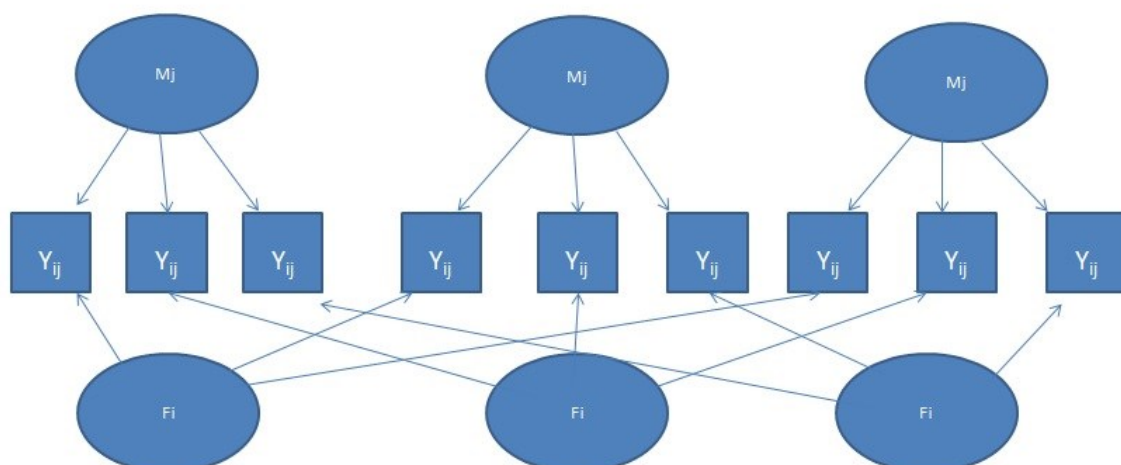
Výše popsané problémy designu sběru MTMM dat řeší metodologická inovace označovaná jako 2 split – ballot MTMM (2 SB MTMM) (Saris, Sattora, Coenders, 2004; Revilla, Saris, 2013a; Revilla, Saris, 2011a). Tento design sběru dat je inspirován metodou split – ballot, která se v metodologii sociálněvědních šetření používá ke zjišťování variability proměnných a jejíž popis je uveden v kapitole 2. Design 2 SB MTMM snižuje respondentovu zátěž z celkových devíti odpovědí v jednom experimentu na šest, a to tím, že rozdělí soubor na dvě stejně velké části, ve kterých každý respondent odpovídá na položky, jež jsou měřeny jen dvěma ze tří použitých metod měření. Redukce počtu položek v experimentu má dva pozitivní efekty; zaprvé se výrazně snižuje pravděpodobnost, že si

respondent pamatuje své předchozí odpovědi na otázku zjišťující jeden ze znaků v MTMM, zadruhé se rovněž snižuje pravděpodobnost výskytu jevu *vyhovění tazateli* (van Meurs, Saris, 1990). Design 2 SB MTMM použili výzkumníci v metodologických šetřeních projektu ESS právě proto, aby se zbavili nepraktických vlastností designu MTMM. Respondenti v prvním ze dvou podsouborů odpovídali na otázky měřené metodou 1 a 3, respondenti v druhém podsouboru odpovídali na otázky měřené metodou 2 a metodou 3.

6.4 Modelování MTMM dat

Pro analýzu MTMM dat či 2 SB MTMM dat, ve kterých chceme identifikovat jen systematický efekt metody měření, se používá model konfirmativní faktorové analýzy. S jeho využitím můžeme identifikovat složku rozptylu odpovídající znaku a složku odpovídající efektu metody (Bollen, 1989; Werts, Linn, 1970; Kenny, Kashy, 1992). V CFA modelu získáme odhady faktorových zátěží mezi měřenými indikátory y_{ij} a jejich latentními znaky F_i (viz obrázek 6.1) (Andrews, 1984), z nichž odečteme koeficienty validity (faktorové zátěže mezi F_i a y_{ij}) a koeficienty systematického vlivu metody (faktorové zátěže mezi M_j a y_{ij}). Je patrné, že tento model nedokáže odhadnout složku rozptylu náhodné chyby měření, nezjistíme v něm tedy reliabilitu.

Obrázek 6.1 Multitrait multimethod (MTMM): znázornění vztahů mezi měřenými a latentními proměnnými (Andrews, 1984)



6. 4. 1 True Score MTMM Model

Tzv. True Score MTMM je analytickou inovací modelu popsaného výše. „True score“ v názvu značí *pravý skór*, čímž odkazuje k ústřednímu konceptu Klasické testové teorie. Díky zavedení pravého skóru do navržených vztahů mezi měřenými a latentními proměnnými je možné v tomto modelu odhadnout proporcí náhodné chyby měření a koeficient reliability. Na rozdíl od předchozího modelu se již nejedná o model CFA, ale o strukturální model, neboť zde nacházíme kauzální vztahy mezi latentními proměnnými. V TS MTMM modelu existuje pro každou položku y_{ij} latentní proměnná T_{ij} , čímž vzniká samostatná faktorová zátěž odpovídající koeficientu reliability. Druhá mocnina této faktorové zátěže potom odpovídá reliabilitě měření (Sarıs, Andrews, 1991; Sarıs, 2009).

TS MTMM model lze vyjádřit jako (Sarıs, Sattora, Coenders, 2004):

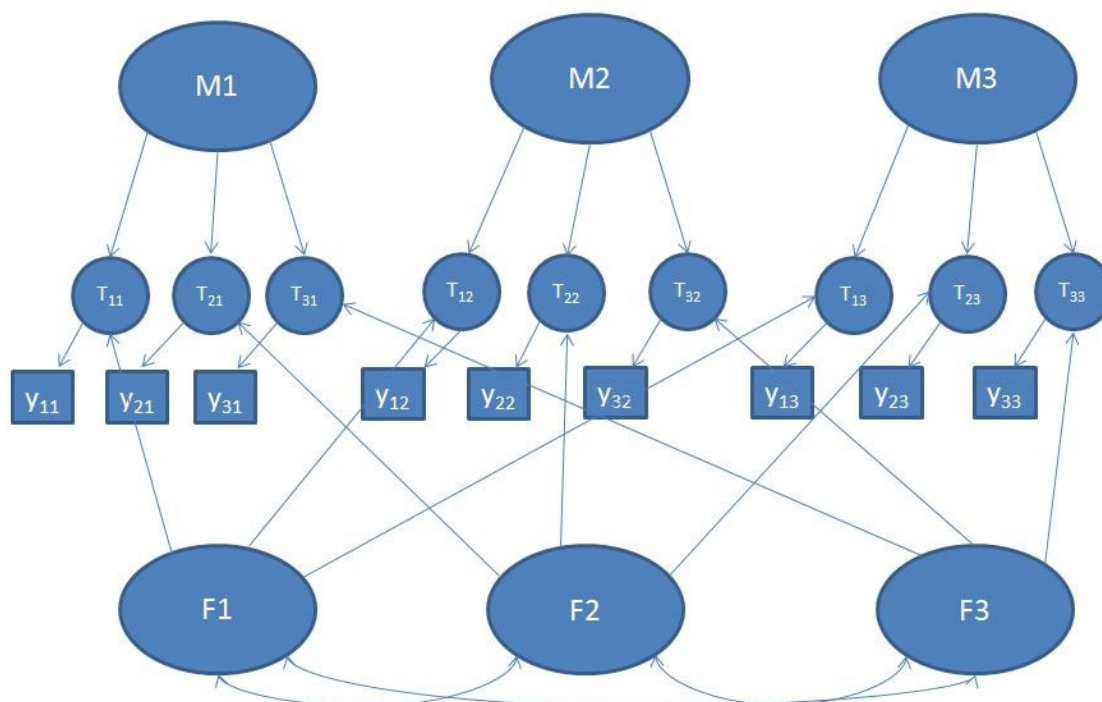
$$(6.1) y_{ij} = r_{ij} T_{ij} + e_{ij} \quad \text{kde } i = 1;3 \text{ a } j = 1;3$$

$$(6.2) T_{ij} = v_{ij}F_i + m_{ij} M_j \quad \text{kde } i = 1;3 \text{ a } j = 1;3$$

- y_{ij} je měřená proměnná. Reprezentuje latentní znak i měřený metodou j ,
- T_{ij} je True score čili pravý skór. Reprezentuje složku rozptylu měřené proměnné y_{ij} očištěnou o náhodnou chybu,
- F_i je latentní znak, který chceme změřit (trait),
- M_j je latentní proměnná, jejíž rozptyl reprezentuje systematický efekt metody j ,
- e_{ij} je náhodná chyba pro každé y_{ij} , která má nulový průměr a není korelována s dalšími náhodnými chybami, s M_j ani s F_i ,
- r_{ij} je ve standardizovaném modelu měření interpretován jako koeficient reliability. Umocněný na druhou r_{ij}^2 odpovídá reliabilitě měření,
- m_{ij} ve standardizovaném modelu měření reprezentuje efekt metody. Umocněný na druhou m_{ij}^2 se rovná podílu rozptylu měřené proměnné, který odpovídá systematické chybě,
- v_{ij} je ve standardizovaném modelu měření interpretován jako koeficient validity. v_{ij}^2 odpovídá konstruktové validitě měření, jejíž výpočet je $v_{ij}^2 = 1 - m_{ij}^2$.

Vztahy mezi měřenými položkami y_{ij} , latentními proměnnými T_{ij} , latentními znaky F_i a latentními proměnnými metod (M_j) v TS MTMM modelu ilustruje obrázek 6.2. V tomto příkladu existuje devět měřených proměnných y_{ij} , tři latentní znaky F_i a tři latentní proměnné metod M_j .

Obrázek 6.2 Znárodnění True Score MTMM modelu; zdroj: (Saris, Galhoffer, 2014)



6. 4. 2 Specifikace True Score MTMM modelu

V analýze MTMM dat či 2 SB MTMM dat v softwaru pro SEM je na TS MTMM model nutná aplikace některých restrikcí (Saris, Sattora, Coenders, 2004). První z nich je fixace faktorové zátěže pro latentní proměnné metod tak, aby byly stejné pro všechny proměnné y_{ij} , tj. že:

$m_{ij} = m_m$ pro všechna i , kde index m značí, že pro danou metodu j jsou faktorové zátěže stejné.

Tato restrikce vyjadřuje, že o každé jedné metodě použité v MTMM či 2 SB MTMM designu se předpokládá, že působí stejně na všechny tři indikátory, které jsou touto metodou měřeny. Aplikace této restrikce navíc zvyšuje stupně volnosti modelu, čímž se zvyšuje pravděpodobnost, že software dojde k řádnému konvergujícímu řešení.

Další restrikce TS MTMM modelu jsou (Saris, Sattora, Coenders, 2004):

- latentní znaky F_i nejsou korelovány s latentními proměnnými metod M_j ,
- latentní znaky F_i jsou mezi sebou korelovány,

- zbytkové rozptyly měřených indikátorů e_{ij} nejsou korelovány mezi sebou ani s dalšími prvky modelu,
- latentní proměnné metod (M_j) mezi sebou nekorelují.

I tyto restriktce zvyšují stupně volnosti modelu a zlepšují šance na získání řádného konvergujícího řešení, všechny jsou zároveň teoreticky opodstatněné. V případě první zmíněné restriktce se předpokládá, že není důvod, aby spolu korelovaly znaky a metody. Druhá restriktce vyjadřuje, že znaky jsou podkonstrukty či latentní indikátory jednoho zastřešujícího konstrukt, v ilustraci v Rámečku 6.1 je to konstrukt Role mužů a žen ve společnosti. Třetí restriktce zajišťuje, že se rozptyl odpovídající metodě nerozplyne v přímé korelaci s jinými položkami, které byly měřeny stejnou metodou, a poslední restriktce předpokládá, že neexistuje důvod pro korelaci latentních proměnných metod.

6. 4. 3 Odhady koeficientů validity a reliability z TS MTMM modelu

Odhady parametrů z TS MTMM přinášejí informace o proporcích rozptylu proměnných odpovídajících substanciálnímu rozptylu a chybovému rozptylu. Z odhadnutých faktorových zátěží odečítáme koeficienty reliability, reliability a vyjádření vlivu metody. Koeficient validity (v_{ij}) je faktorová zátěž mezi latentní proměnnou T_{ij} a odpovídajícím latentním znakem F_i ; umocníme – li ho na druhou, získáváme odhad validity měření v_{ij}^2 . Nezapomeňme, že validita je v rozšířené Klasické testové teorii specifickým druhem validity a vyjadřuje inverzi k efektu metody. Ten je v TS MTMM modelu vyjádřen odhadem faktorové zátěže mezi latentní proměnnou T_{ij} a latentní proměnnou metody M_j . Inverzní vztah mezi validitou a efektem metody vyjadřuje následující rovnice (Saris, Gallhofer 2014:201):

$$(6.3) \quad v_{ij}^2 = 1 - m_{ij}^2$$

Číselná hodnota m_{ij}^2 je ta část rozptylu daného indikátoru, která je vysvětlena vlivem latentní proměnné metody M_j . Její odmocnina je použita k výpočtu tzv. společného rozptylu metody (CMV), který sdílí dvě či více proměnných, které byly měřeny stejnou metodou měření.

$$(6.4) \quad CMV = r_i m_i m_j r_k,$$

kde r_i a r_k jsou koeficienty reliability položky i a položky k , které byly měřeny metodou M_j (Saris, Gallhofer, 2014:290)³¹. Tento ukazatel nachází využití v případě, kdy chceme vědět, zda a do jaké míry se mohla zvýšit korelace mezi dvěma proměnnými, které byly měřeny stejnou metodou. V substanciální analýze dokáže tento ukazatel odstranit zvýšení korelace mezi dvěma proměnnými vzniklé vlivem použité metody.

Koeficient reliability (r_{ij}) je faktorová zátěž mezi latentní proměnnou T_{ij} a jejím indikátorem y_{ij} . Druhá mocnina tohoto čísla odpovídá reliabilitě (r_{ij}^2) a je to rozptyl měřeného indikátoru očištěný o náhodnou chybu.

Koeficient validity a reliability lze zkombinovat v jeden celkový ukazatel kvality položky, tzv. *total quality of a measure* (q_{ij}^2) (Saris, Gallhofer 2014:294), pro který volím český jazykový ekvivalent *celková kvalita položky*.

Celková kvalita položky q_{ij}^2 se vypočítá podle:

$$(6.5) \quad q_{ij}^2 = (r_{ij} \cdot v_{ij})^2.$$

Ukazatel číselně vyjadřuje, jak velká část rozptylu měřené proměnné odpovídá výhradně měřenému znaku, je to tedy o všechny chyby očištěný rozptyl měřené proměnné. Pro ilustraci, v položce v Rámečku 1 „*TIMI Žena by měla být připravena omezit svou placenou práci kvůli rodině*“ q^2 odpovídá výhradně latentní proměnné vyjadřující názor respondentů na omezení práce žen kvůli rodině.

6. 4. 4 Využití odhadů kvality z TS MTMM modelu

Jaké využití mají odhady kvality z TS MTMM modelu? Samy o sobě slouží jako informace o kvalitě konkrétních měření; na základě odhadů reliability či efektu metody můžeme zhodnotit kvalitu měření konkrétních položek a stanovit, zda jsou takové položky pro substanciální analýzu vhodné. Druhou možností je použití odhadů kvality přímo v substanciální analýze, kde místo celých rozptylů proměnných použijeme očištěný rozptyl. Příklady substanciální analýzy s očištěnými daty a srovnání výsledků s očištěním o

³¹ Saris a Gallhoferová 2014:290 uvádějí vzorec $CMV = r_i m_i m_j r_k$, s rozdílnými indexy u metod (m_i, m_j), přestože CVM nachází využití pouze v případě, kdy jsou metody, kterými byly korelující položky měřeny, stejné. Logicky by tedy i indexy u metod v rovnici měly být stejné. Autoři však uvádějí rovnici ve znění výše, proto je cituji tak, jak uvádějí, a na tuto nesrovnalost upozorňuji.

chyby a bez něj ukazují v Kapitole 7, která obsahuje i návod, jak data o chybové rozptyly očistit.

6.5 Problematické aspekty analýzy 2 SB MTMM dat

Díky metodologickým šetřením ESS máme v současné době k dispozici velké množství 2 SB MTMM dat. Tento speciální design využívající dva podsoubory respondentů byl v šetřeních použit z dobrých důvodů; řeší zásadní problém klasického MTMM designu, tedy následek situace, kdy si respondent odpovídající na tři velmi podobné otázky svoje předchozí odpovědi pamatuje a pouze je zopakuje, nebo kdy se v důsledku častého opakování podobné otázky unaví a odpovídá nedbale. Protože 2 SB MTMM design po respondentovi vyžaduje pouze dvě odpovědi na podobné otázky, výrazně se snižuje zátěž na respondenta. Podle (Van Meurs, Saris, 1990; Revilla, Saris 2011b) respondenti zapomínají své odpovědi na otázky v průměru po dvaceti pěti minutách, resp. po zodpovězení 75 otázek, takže při použití 2 SB MTMM by během jednoho výzkumného rozhovoru v šetření ESS měli v době odpovídání na druhou otázku svoji první odpověď již zapomenout. Není sice možné vyloučit, že někteří respondenti si pamatují svoji první odpověď i po delší době, nicméně je potřeba uznat, že pravděpodobnost efektu paměti přístup split – ballot zásadně snižuje, a proto je přínosem a metodologickou inovací.

Data z designu 2 SB MTMM lze analyzovat metodou strukturálního modelování stejně jako MTMM data pocházející z úplného designu, jejich analýza je však komplikovanější. Vstupní matice korelací totiž není úplná, neboť v ní chybí údaje o korelacích mezi proměnnými měřenými metodou 1 a 2. Výzkumný tým RECSM řeší tento problém aplikací multigroup SEM analýzy (Multigroup Maximum Likelihood metoda odhadu), ve které je příslušnost ke skupině (group) dána příslušností respondenta k jednomu ze dvou podsouborů. Ve statistickém softwaru LISREL, ve kterém RECSM dělá své výpočty, jsou jako vstupní data použity dvě matice korelací, první odpovídající podsouboru, jenž odpovídal na otázky měřené metodami 1 a 3, druhá, obsahující data ze souboru, který odpovídal na otázky měřené metodami 2 a 3. Ukázku toho, jak tyto matice vypadají, a jak je na ně aplikován TS MTMM model obsahuje Příloha 6.1.

Multigroup analýza matic s chybějící korelacemi je sice originální způsob, jak analyzovat neúplná data 2 SB MTMM, není však stoprocentně spolehlivá; v porovnání s běžnou analýzou úplných MTMM dat vede častěji k nekonvergujícímu řešení nebo výskytu Heywoodových případů (Revilla, Saris, 2011b). Inovativní 2 SB MTMM design

se tak stává příčinou dalších problémů v analýze MTMM dat, s tím rozdílem, že problémem již není paměť a únava, ale obtíže získat výpočty. Nejčastějšími komplikacemi při využívání 2 SB MTMM dat pro odhady validity a reliability jsou následující tři; 1) problémy při získávání řádného konvergujícího řešení TS MTMM modelů na souborech dat o velikosti 1000 až 3000 případů, 2) maximální náročnost provádění analýz a 3) nemožnost získání řádných konvergujících řešení TS MTMM modelu v případě některých konkrétních 2 SB MTMM experimentů.

6. 5. 1 Nekonvergence TS MTMM modelů na „malých“ souborech

V analýze 2 SB MTMM dat jsou nekonvergence nebo řešení s HC velmi běžné zejména v případech, kdy jsou použita data pouze z jednoho národního státu, tedy soubor o velikosti 1000 až 3000 případů (Oberski, Gruner, Saris, 2011; Revilla, Saris, 2013a; Saris, Gallhofer, 2014: 220). Ve většině případů nelze pro tyto relativně malé datové soubory vůbec získat řádné konvergující řešení a tedy ani odhady validity a reliability. Podle týmu RECSM jsou obtíže důsledkem sníženého množství stupňů volnosti modelu, pro jehož odhad jsou použita data s chybějícími korelacemi mezi proměnnými měřenými metodou 1 a 2. Tento problém nelze prakticky nijak řešit, což je poměrně závažný nedostatek designu 2 SB MTMM.

Možností, jak sbírat split – ballot MTMM data, a přitom mít korelace mezi všemi devíti proměnnými v experimentu, je používání 3 split – ballot MTMM designu, který využívá tři podsoubory respondentů. V tomto designu respondenti v prvním podsouboru odpovídají na otázky měřené metodou 1 a 2, v druhém podsouboru na otázky měřené metodou 2 a 3 a v třetím podsouboru na otázky měřené metodou 1 a 3. Tento 3 SB MTMM design uchovává výhody split – ballot MTMM přístupu a zároveň poskytuje korelace mezi všemi položkami; z těchto dat je tedy možné běžně získávat řádná konvergující řešení (Revilla, Saris, 2011). Design 3 SB MTMM se však téměř nepoužívá a data z něj nejsou až na výjimku několika studií k dispozici. Hlavním důvodem je jeho nepraktičnost: neposkytuje úplná data za všechny jednotky v celém výběrovém souboru. Na každou položku z MTMM experimentu odpoví jen dvě třetiny respondentů v souboru, takže pokud potřebujeme, aby některé proměnné z experimentu byly změřeny na celém souboru (což je např. případ ESS), nemůžeme 3 SB MTMM design použít.

6. 5. 2 Řešení pro úspěšnou analýzu 2 SB MTMM dat

Řádná konvergence TS MTMM modelu na 2 SB MTMM datech je tím pravděpodobnější, čím je analyzovaný soubor dat větší (Revilla, Saris, 2013a), takže jedinou možností, jak získat odhady kvality měření z modelů, je použít maximálně velké soubory dat.

Výzkumnice a výzkumníci pracoviště RECSM proto používají strategii hromadné analýzy 2 SB MTMM dat, do které vstupují data ze všech zemí, které se v dané vlně účastnily ESS. Tato strategie je v podstatě jediným způsobem, jak získat důvěryhodné odhady parametrů z TS MTMM modelů aplikovaných na experimentální data z ESS. Je však časově a technicky velice náročná, takže její provedení závisí na jistém know how a spolupráci většího množství analytiků.

Tým RECSM analyzuje soubor obsahující data ze všech zemí ESS s použitím metody odhadu Multigroup Maximum Likelihood, kde je jako třídící proměnná (group) použita proměnná „země“ (Saris, Gallhofer, 2014:220; Oberski, Gruner, Saris, 2011). Kombinuje tak dvojí multigroup analýzu, kdy je příslušností ke group jak členství ve split – ballot souboru, tak v zemi. Analytici pak upravují restriktce modelu, až najdou unikátní řešení pro každou zemi zvlášť. Analýzu provádějí dva na sobě nezávislí analytici, kteří porovnávají své výsledky po každém analytickém kroku. Při hledání nejlepšího modelu postupují do jisté míry exploračně, protože upravují restriktce modelu tak, aby získali co nejlepší shodu modelu s daty. Analýza končí, když je nalezena nejlepší shoda modelu s daty. K určení nejlepšího modelu z řady vhnížděných modelů používá RECSM program JRule (Van der Veld, Saris, Satorra, 2008), který kontroluje chybu 1. a 2. druhu.

Popsaná analytická procedura je nepochybně náročná a pro výzkumnici či výzkumníka mimo tým RECSM je obtížné zorientovat se v používaných postupech. RECSM na svých webových stránkách sice publikuje většinu informací potřebných k pochopení problému a dodatečné dokumenty jako syntaxe k výpočtu a vzorce výpočtu, ale některé více či méně detailní informace je těžké dohledat. Například není jasné, jak v analýze nakládat s chybějícími hodnotami; v programu LISREL jsou vstupními daty matice kovariancí a údaj o počtu případů v souboru. Vyvstává otázka, zda používat metodu *listwise deletion* nebo *pairwise deletion*, a pokud budeme používat *pairwise deletion*, musíme se nějak vypořádat s faktem, že máme různé počty případů pro kovariance dvojic proměnných. Odpověď na tuto otázku jsem získala od analytičky RECSM Melanie Revillové, která mi sdělila, že používají metodu *pairwise* a pro počet chybějících hodnot volí průměr ze všech případů

pro všechny dvojice korelací³². Uvedený příklad je však jen jednou z mnoha nejasností. Další z nich je např. nejistota o tom, jak přesně analytici RECSM v multigroup SEM analýze dat ze všech zemí ESS kombinují příslušnosti případů k zemi a k souboru split – ballot. Odpověď na tuto otázku se mi v materiálech prezentovaných RECSM nepodařilo najít.

6. 5. 3 Nemožnost získat odhady pro některé 2 SB MTMM experimenty

Největším problémem metody 2 SB MTMM je, že některá 2 SB MTMM data nelze úspěšně analyzovat vůbec, tedy ani při analýze dat ze všech zemí ESS najednou. Jedním z těchto případů je experiment z ESS nazvaný Media, který obsahuje otázky na zjištění četnosti sledování médií, použitý v šetřeních ESS 1, ESS 4 a ESS5. Saris a Gallhoferová (2014:220) uvádějí, že pro tento experiment se vůbec nepodařilo najít konvergující model. Nekonvergence modelu pro experiment Media je způsobena tím, že skutečná korelace mezi latentními znaky (F_i) v TS MTMM modelu je blízká nule (Saris, Gallhofer, 2014:218). Proč je tato vlastnost pro TS MTMM model problematická, vysvětlují na následujících řádcích; v SEM analýze obecně je možné najít unikátní konvergující řešení v případě, kdy je v modelu dostatečné množství stupňů volnosti. To má model v případě, když má latentní proměnná, která nekoreluje s jinou latentní proměnnou, alespoň tři indikátory, nebo když má latentní proměnná nejméně dva indikátory a zároveň koreluje s jinou latentní proměnnou. Protože v datech z 2 SB MTMM má při použití metody odhadu Multigroup Maximum Likelihood každý latentní znak (F_i) pouze dva indikátory, je nulová korelace mezi F_i příčinou toho, že software pro SEM nemůže najít konvergující řešení. Revilla a Saris (2011a) uvádějí ještě další okolnosti analýzy 2 SB MTMM dat, kdy mohou nastat problémy. Získat konvergující řešení je nemožné, když je rozptyl některého z faktorů metody M_j příliš malý, nebo když jsou korelace mezi latentními znaky F_i stejně velké. Pokud nastanou tyto případy, není 2 SB MTMM design tzv. empiricky identifikovaný a v analýze není možné dojít k odhadům validity a reliability (Saris, Satorra, Coenders, 2004).

6. 6 Empirická část: Aplikace TS MTMM modelu na česká data šetření ESS

Na základě výše popsaných problematických aspektů analýzy 2 SB MTMM dat, jsem se rozhodla analyzovat tato data z českých šetření ESS. Cílem analýzy bylo zjistit, v kolika případech se podaří získat řádné konvergující řešení a odhady kvality měření položek.

³² Zdroj: emailová komunikace se členkou týmu RECSM Melanií Revillovou v červnu 2014.

Navazujícím cílem bylo porovnat odhady kvality měření z mé analýzy s odhady, které prezentoval tým RECSM. Vzhledem k tomu, že v době vypracování této analýzy měl RECSM k dispozici jen odhady z první a druhé vlny šetření ESS, analyzovala jsem jen data z prvních dvou vln.

6. 6. 1 Data European Social Survey

V této práci jsem použila data z mezinárodního projektu výběrových šetření European Social Survey (ESS). Tento projekt mapuje postoje obyvatel evropských zemí k různým společenským fenoménům, oproti jiným mezinárodním projektům však poskytuje velký prostor metodologickým aspektům sociálněvědního výzkumu. Ve všech vlnách a všech zúčastněných zemích jsou měřeny 2 SB MTMM experimenty a v každém šetření se nachází několik takových experimentů. V této analýze používám česká data z prvních dvou vln ESS, první vlna proběhla v roce 2002, druhá následovala o dva roky později.

6. 6. 2 Výsledky analýzy

Na soubory českých 2 SB MTMM dat o velikosti mezi 1000 a 2000 případů jsem aplikovala TS MTMM model s restrikcemi tak, jak byl popsán výše v této kapitole. K analýze jsem použila software pro strukturální modelování LISREL (Jöreskog, Sörbom, 2004), použité syntaxe jsou v příloze této dizertační práce³³. K odhadu parametrů jsem použila metodu Multigroup Maximum Likelihood.

Analýza nepřinesla překvapivé výsledky. Pro naprostou většinu experimentů nebylo podle očekávání možné získat řádné konvergující řešení. První vlna ESS obsahovala šest 2 SB MTMM experimentů, z nichž jsem analyzovala všechny, druhá vlna ESS obsahovala rovněž šest experimentů, z nichž jsem analyzovala pět. Pro 10 z těchto celkem 11 experimentů jsem získala nekonvergující řešení nebo řešení s Heywoodovými případy (HC). V datech z první vlny jsem z celkem šesti experimentů obdržela čtyři řešení s HC, jedno nekonvergující a jedno s blíže neurčenou chybou, kterou se mi nepodařilo najít a odstranit. V datech z druhé vlny jsem získala tři řešení s HC, jedno nekonvergující a jedno řádné konvergující řešení. Tím řádně konvergujícím řešením byl experiment Role mužů a žen ve společnosti popsán v této kapitole výše v Tabulce 6.1. Ukazatele shody tohoto

³³ Syntaxe ke všem modelům, které jsem použila na data z 2 SB MTMM experimentů z ESS1 a EES2, jsou stejné. Liší se pouze zadání dat, tj. korelační matice, průměry a směrodatné odchylky.

modelu s daty byly vynikající (χ^2 121.3; df 57; RMSEA 0,035; 90% RMSEA 0,026; 0,044; P value 1,00; CFI = 0,97), odhady z tohoto modelu tak mohou být poměrně přesné.

Sloupec v Tabulce 6.2 označený Model ukazuje, že validita pro 6 z 9 položek experimentu je poměrně vysoká, v případě reliability jsou výsledky o něco horší – reliability je nižší u těch samých třech položek, v ostatních případech je ale také poměrně nízká, pouze pro dvě položky je vyšší než 0,8. Dílčím cílem analýzy bylo porovnání odhadů z mé analýzy s odhady, které prezentuje tým RECSM ve svém online programu SQP³⁴. Toto srovnání je rovněž v Tabulce 6.2, kde sloupec SQP prezentuje odhady, které vzešly z analýz RECSM. Ze srovnání vyplývá, že odhady z mé analýzy se od těch v programu SQP do jisté míry liší. Příčinou jsou zřejmě rozdíly v použité analytické strategii, a to především fakt, že RECSM používá multigroup analýzu pro jednotlivé země a data ze souboru všech zemí ESS, zatímco já jsem použila jen data z Česka. Za povšimnutí stojí, že SQP vůbec neobsahuje odhady pro položky F1M2 a F2M2, což naznačuje, že model, který použil tým RECSM, nedokázal tyto parametry odhadnout.

Tab. 6.2 Srovnání odhadů validity a reliability položek z experimentu *Role mužů a žen ve společnosti* získaných z vlastní analýzy českých dat ESS2 a z databáze programu SQP 2.0

| Ukazatel | v ² validita | | r ² reliability | |
|----------|-------------------------|-------|----------------------------|-------|
| | Model | SQP | Model | SQP |
| F1M1 | 0,846 | 0,962 | 0,640 | 0,619 |
| F2M1 | 0,624 | 0,976 | 0,372 | 0,249 |
| F3M1 | 0,902 | 0,972 | 0,723 | 0,612 |
| F1M2 | 0,240 | NA | 0,185 | NA |
| F2M2 | 0,864 | NA | 0,830 | NA |
| F3M2 | 0,593 | 0,907 | 0,292 | 0,421 |
| F1M3 | 0,922 | 0,908 | 0,846 | 0,648 |
| F2M3 | 0,903 | 0,889 | 0,656 | 0,633 |
| F3M3 | 0,941 | 0,930 | 1,00 | 0,734 |

NA – Not available. SQP 2.0 z blíže nespecifikovaného důvodu neposkytuje pro tyto položky odhady kvality.

Zdroj dat: European Social Survey, 2. vlna

³⁴ Hodnoty z SQP v Tabulce 6.2 jsou platné pouze pro data získaná v ČR.

6. 7 Diskuze

Text této kapitoly přinesl informace o problematice analýzy 2SB MTMM a popsal použití TS MTMM modelu jako metody, která umí odhadnout validitu i reliabilitu měřených položek. Zvláštní pozornosti se dostalo problémům, ke kterým dochází v SEM analýze 2 SB MTMM dat, mezi kterými jako nejpodstatnější vyčnívá fakt, že v některých případech z tohoto experimentálního designu není vůbec možné získat odhady kvality měření. Tuto skutečnost potvrdila i vlastní analýza, ve které jsem získala řádně konvergující model pouze pro jeden z celkem jedenácti 2 SB MTMM experimentů, které byly měřeny v prvních dvou vlnách ESS.

Odhady parametrů ukázaly, že poměrně slušnou validitu a zároveň reliabilitu mají čtyři z devíti položek:

- F3M1, „Když je málo pracovních míst, muži by měli mít větší právo na zaměstnání než ženy“ ($v^2=0,902$; $r^2=0,723$);
- F2M2, „Ženy by měly mít větší odpovědnost za domov a děti než muži“ ($v^2=0,864$; $r^2=0,830$);
- F1M3, „Žena by měla být připravena omezit placenou práci kvůli rodině 1 2 3 4 5 Žena by neměla být nucena omezovat placenou práci kvůli rodině“ ($v^2=0,922$; $r^2=0,846$);
- F3M3, „Když je málo pracovních míst, muži by měli mít větší právo na zaměstnání než ženy 1 2 3 4 5 Když je málo pracovních míst, ženy by měly mít stejné právo na zaměstnání jako muži“ ($v^2=0,941$; $r^2=1,00$)

Srovnání odhadů z mé analýzy s odhady, které našel tým RECSM, odhalilo, že analýzy došly do jisté míry k odlišným výsledkům. Vzhledem k rozdílům v proceduře, kterou jsem použila já a kterou používají analytičky a analytici RECSM, odkazují případné zájemce o přesné odhady reliability, validity a dalších ukazatelů chyby měření na zdroj RECSM, kterým je online software SQP. Tento program a možnosti, jak využít odhady validity a reliability, popisují v následující kapitole, která je věnována praktickému zacházení s chybami měření v substanciální analýze.

7. Využití odhadů chyb měření pro substanciální analýzu

V teoreticko – analytickém rámci, použitém v této práci, definuji chyby měření jako ty části rozptylu měřené proměnné, které neodpovídají měřenému konstrukt. Za chybu měření tak považuji náhodný rozptyl, jehož vlastnosti definuje Klasická testová teorie, a rozptyl, který vznikl v důsledku použité metody měření, např. konkrétní Likertovy škály. Pokud měřenou proměnnou očistíme o všechny tyto složky, dostáváme rozptyl, který je reflexí pouze té vlastnosti (znaku či konstrukt), kterou chceme zkoumat. Nabízí se otázka, zda bychom v substanciální analýze neměli zkoumat právě jenom tu část měřených proměnných odpovídající konstrukt a zabývat se vztahy mezi proměnnými, které byly očištěny od všech rušivých vlivů. Na tuto otázku však není jednoznačná odpověď; rozhodnutí, zda brát v substanciální analýze v úvahu chybu měření či nikoliv, je věcí osobního názoru a možná i aktuálního rozpoložení. Postoj odborné veřejnosti k chybám měření lze s trochou fantazie ilustrovat na pomyslné škále možných přístupů k této problematice. Na jednom konci škály se nacházejí ti, kteří jsou hluboce přesvědčeni, že v rozptylu proměnné lze určit validní a nevalidní část, a že část proměnné, která nekoreluje s daným konstruktem a jejíž variabilita má náhodný charakter, je jakýsi šum, který v analýze nemá co dělat. Slova „konstrukt“ a „operacionalizace“ jsou pro ně svatá a v dotazníkové položce vždy vidí spíše reflexi konstrukt, než konkrétní otázku. Mezi tyto sociální vědce rozhodně patří v této práci mnohokrát citovaní metodologové Saris, Revillová nebo např. Billiet, kteří se ve své vědecké práci věnují téměř výhradně studiu a analýze chyb měření. Naopak na druhém konci pomyslné škály jsou ti odborníci, podle kterých je svým způsobem „správná“ každá respondentova odpověď. Pro tyto výzkumníky nese každá hodnota v datech relevantní informace a nesrovnalosti je neznepokojují natolik, aby pochybovali o důvěryhodnosti a výpovědní hodnotě svých dat. Za jejich zástupce lze považovat politického vědce Philipa Converse, podle kterého nejsou data z výběrových šetření zanesena chybami, ale přinášejí hodnoty, které „správně“ reflektují, že podstatná část lidí nemá na velkou část společenských záležitostí žádný názor (Feldman, 1989; Converse, 1964). Converse získal podklady pro své přesvědčení v analýze panelových dat týkajících se postojů k veřejné politice, ze kterých zjistil, že postojové velké části lidí jsou značně nestabilní. Nestabilitu však nevysvětluje chybou měření, ale tím, že lidé nemají skutečné, dlouhodobé postoje a v danou chvíli odpovídají tak, jak se zrovna hodí. Je zjevné, že odborníkům a odborníkům, kterým je bližší přístup Converse, není tato práce určena. Nicméně ti, kteří se blíží prvnímu konci pomyslné škály, by mohli ocenit ukázkou

vlivu chyb měření na výsledky substanciálních analýz a popis, jak lze data vstupující do analýzy očistit o chybové rozptyly. Obě ukázky prezentuje tato kapitola, v jejím závěru je popsán online program SQP, který obsahuje odhady reliability, validity a dalších ukazatelů vlivu systematické chyby získané z dat ESS. Tento program navíc umožňuje predikci kvality jakékoliv dotazníkové položky.

7.1 Vliv snížené validity a reliability na substanciální analýzu

Vliv chyby měření v této kapitole ilustruji na příkladech bivariační analýzy, tj. korelaci dvou proměnných, a multivariační analýzy, tedy vztazích vyplývajících z kovariance více proměnných. V dalším textu uvádím několik konkrétních příkladů, které ukazují, že zahrnutí chyby měření do analýzy proměnných může v některých případech zásadně ovlivnit velikost a směr nalezených lineárních vztahů, a ilustrují, že analýza s daty, které nebyly očištěny o chybu měření, dochází k jiným výsledkům, než analýza dat, ve které je uvažován vliv chyb měření.

Vliv snížené reliability na korelaci dvou proměnných je dlouhodobě v centru pozornosti v psychometrii. V tomto vědním oboru výzkumníci používají k eliminaci vlivu náhodné chyby tzv. *correction for attenuation*, která ukazuje změnu ve velikosti korelace mezi dvěma proměnnými, které nejsou stoprocentně reliabilní. Překlad tohoto anglického odborného výrazu do češtiny je problematický, a tak jej pro účely tohoto textu volně překládám jako *korekci náhodné chyby*. Korekci náhodné chyby lze při existenci dvou měření (x, y) tohoto samého znaku vyjádřit následovně (Bollen, 1989):

$$(7.1) \quad r_{x'y'} = r_{xy} / \sqrt{r_{xx} r_{yy}}$$

kde

$r_{x'y'}$ = korelace x a y po korekci

r_{xy} = pozorovaná korelace x a y

r_{xx} = reliabilita x

r_{yy} = reliabilita y

Z formule (7.1) vyplývá, že očištění korelací o vliv náhodné chyby vždy vede ke zvýšení hodnoty korelace. Můžeme to ilustrovat na příkladu: známe – li reliabilitu měření x

($r_{xx}=0,34$) a měření y ($r_{yy}=0,66$) a korelace těchto dvou proměnných se rovná 0,2, zjistíme dosazením do formule (7.2), že jejich korelace po korekci je 0,43.

V české akademické sociologii o vlivu náhodné chyby měření na korelace proměnných stručně informoval např. Řehák (1971, 1998a), zahraniční odborná literatura je v tomto ohledu bohatší a nabízí např. ilustrace vlivu náhodné chyby na multivariační analýzu. Bollen (1989) v klasické publikaci o modelování strukturálních rovnic ukazuje hypotetické příklady vlivu snížené reliability na velikost regresních koeficientů v modelu s jednou závislou a sedmi nezávislými proměnnými. Používá konkrétní příklad modelu vztahů mezi znečištěním ovzduší a úmrtností; úmrtnost η je v modelu definována jako lineární funkce sedmi proměnných ξ .

$$(7.1) \quad \eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \gamma_4 \xi_4 + \gamma_5 \xi_5 + \gamma_6 \xi_6 + \gamma_7 \xi_7 + \zeta$$

η = úmrtnost

ξ_1 = znečištění siřičitany

ξ_2 = částice v ovzduší

ξ_3 = hustota populace

ξ_4 = procento populace nad 65 let

ξ_5 = procento nebělochů v populaci

ξ_6 = procento rodin s příjmem pod hranicí chudoby

ξ_7 = přirozený logaritmus populace

ζ = náhodný rozptyl

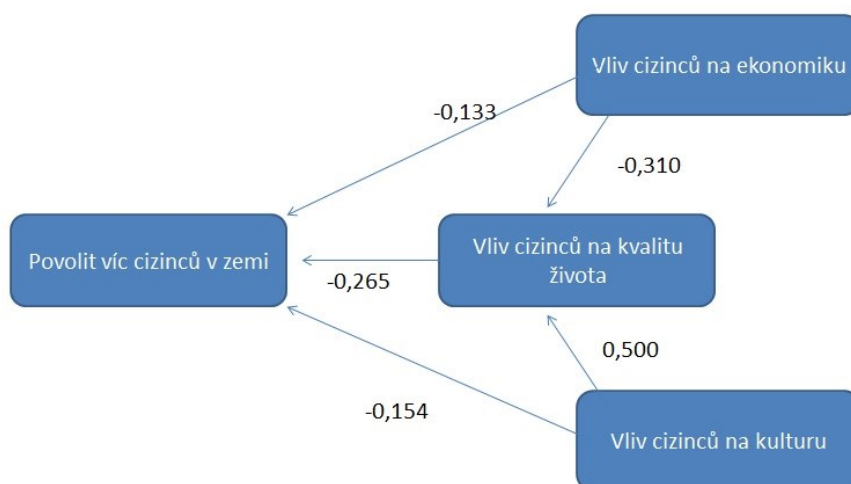
Proměnné úmrtnost η odpovídá měřená proměnná x , o které se v příkladu uvažuje jako o proměnné se sníženou reliabilitou, o nezávisle proměnných ξ se předpokládá, že jsou bez chyby měření, tedy že se jejich reliabilita rovná jedné. V tabulce 7.1 vidíme, jak různé úrovně snížené reliability závislé proměnné x ovlivňují pozorované vztahy mezi závislou a nezávislými proměnnými. Zatímco při reliabilitě x rovné jedné, tedy v případě, kdy v rozptylu měřené proměnné neexistuje náhodná chyba, je regresní koeficient mezi x a ξ_1 roven 0,107, při reliabilitě 0,5 je to 0,291. V tomto konkrétním modelu má náhodná chyba ten vliv, že regresní vztah mezi proměnnými zvyšuje. V případě lineárního vztahu mezi ξ_3 -> x nastává jev opačný, regresní koeficient za předpokladu ideální reliability má hodnotu 0,064, při reliabilitě 0,5 už to je jen 0,017.

Tab. 7.1 Vliv složky náhodné chyby na regresní koeficienty; zdroj: (Bollen 1989)

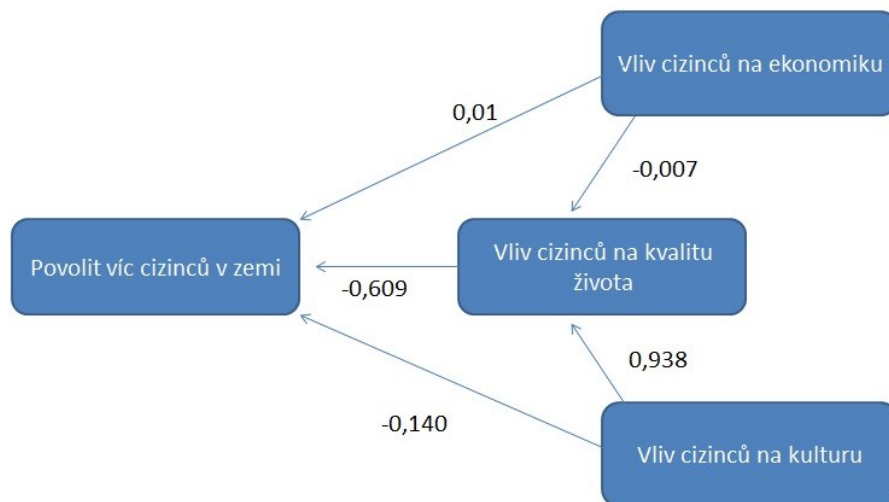
| Regresní koeficient | | | | | | | |
|---------------------|-------------------------|----------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Reliabilita | $\xi_{1 \rightarrow x}$ | $\xi_{2 \rightarrow \eta}$ | $\xi_{3 \rightarrow x}$ | $\xi_{4 \rightarrow x}$ | $\xi_{5 \rightarrow x}$ | $\xi_{6 \rightarrow x}$ | $\xi_{7 \rightarrow x}$ |
| 1 | 0,107 | 0,090 | 0,064 | 1,008 | 0,370 | -0,063 | -0,076 |
| 0,9 | 0,123 | 0,086 | 0,060 | 1,003 | 0,369 | -0,062 | -0,072 |
| 0,7 | 0,173 | 0,047 | 0,047 | 0,986 | 0,363 | -0,056 | -0,056 |
| 0,5 | 0,291 | 0,017 | 0,017 | 0,947 | 0,350 | -0,044 | -0,020 |

Další ukázkou vlivu chyby měření na vztahy v analýze zprostředkovává publikace Sarise a Gallhoferové (2014), kteří použili odhady reliability a systematické chyby měření získané prostřednictvím TS MTMM modelu z 2 SB MTMM dat z European Social Survey. Tito autoři ukazují ještě markantnější změny než Bollen. Vliv chyby ilustrují v lineárních vztazích mezi proměnnými na příkladu pěšinkové analýzy se čtyřmi proměnnými. Změny jsou v tomto případě patrnější, neboť autoři počítají s náhodnou i systematickou chybou, a to u všech proměnných v analýze. Rozdíly ve velikosti faktorových zátěží mezi proměnnými bez korekce o chybu měření a s korekcí jsou znázorněny na Obrázcích 7.2 a 7.3.

Obrázek 7.1 Kauzální model vztahů mezi proměnnými bez očištění dat o chybu měření; zdroj: (Saris a Galhofferová 2014: 288 – 291)



Obrázek 7.2 Kauzální model vztahů mezi proměnnými s očištěním dat o náhodnou a systematickou chybu měření; zdroj: (Sarıs a Galhofferová 2014: 288 – 291)



Z obrázků 7.1 a 7.2 vyplývá, že vztahy mezi proměnnými se po očištění rozptylů výrazně změnilo. Původní vztah mezi vlivem cizinců na ekonomiku a postojem k povolení většího množství cizinců v zemi o velikosti -0,133 v podstatě zmizel a např. vztah mezi názorem na vliv cizinců na kulturu na názor na vliv cizinců na kvalitu života v zemi se z původních 0,5 zvýšil na 0,938.

7.2 Očištění dat o chybový rozptyl

Ve čtvrté kapitole této dizertační práce prezentuji odhady reliability tří dotazníkových položek z šetření EU SILC, v šesté kapitole jsou uvedeny odhady reliability a validity pro několik položek použitých v šetření ESS. V Kapitole 6 čtenářky a čtenáři rovněž naleznou popis koeficientů, které lze získat z TS MTMM modelu, a výpočet dalších ukazatelů kvality, tj. společného rozptylu metody (CMV) a celkovou kvalitu položky q_{ij}^2 . Všechny tyto odhady a indikátory lze použít k očištění dat o chybové rozptyly. V případě, kdy máme k dispozici pouze odhady reliability (např. z kvazisimplexového modelu) je možné očišťovat data pouze o náhodnou chybu a existenci systematických chyb ignorovat. V situacích, kdy máme informace o validitě i reliabilitě měření si můžeme vypočítat celkovou kvalitu položky q_{ij}^2 a tu použít pro filtraci všech ne – substantiálních částí rozptylu proměnné.

Očištěná data mohou bez problému vstoupit do analýz, jako je regresní analýza, pěšinková analýza nebo strukturální modelování a lze s nimi provádět analýzu v běžných statistických paketech nebo v softwaru pro strukturální modelování jako je LISREL, Mplus nebo AMOS. Očištění dat o chybové rozptyly nelze udělat pro jeden případ ve výběrovém souboru zvlášť, je ale možné očistit korelace mezi měřenými proměnnými. Vstupními daty pro software je pak matice korelací, nikoliv primární datový soubor.

Nejjednodušším způsobem, jak očistit data od chybových rozptylů je vytvořit z vybraných položek korelační matici. Na diagonále takové matice se nacházejí korelace proměnné sama se sebou o hodnotě 1. Tato korelace je pro práci s očištěnými daty nahrazena hodnotou ukazatele celkové kvality položky q_{ij}^2 , případně odhadem reliability. Pokud takto upravenou matici vložíme jako vstupní data do programu pro SEM, program sám přepočítá ostatní korelace v matici podle hodnoty uvedené na diagonále a připraví tak matici očištěných korelací vhodnou k další analýze. V případě, kdy položky v matici sdílejí metodu měření, je nutné začlenit do očištěné korelační matice i CMV. Hodnoty CMV jsou umístěny do matice nad diagonálu obsahující hodnoty celkových kvalit položek na místo, kde se kříží položky sdílející metodu měření (Oberski, Gruner, Saris, 2011; Saris, Gallhofer, 2014).

V případě, že si nevíme rady se softwarem pro přepočet matic, můžeme matice korelací upravit i ručně, a to podle následujícího vzorce (Saris, Gallhofer, 2014:290):

Očištěná korelace $r_{ik} = (\text{pozorovaná korelace } r_{ik} - \text{CMV}) / q_i \cdot q_k$

kde r_{ik} je korelace položek i a k , q_i a q_k odpovídají odmocninám hodnot celkové kvality položek (q_{ij}^2) i a k . CMV odpovídá společnému rozptylu metody a využijeme jej pouze v případě, kdy korelující proměnné sdílejí stejnou metodu. Pokud korelující položky nebyly měřeny stejnou metodou, CMV v rovnici nemá opodstatnění a nepočítá se s ním. Všimněme si, že tento vzorec je velice podobný formuli *korekce náhodné chyby*, uvedené na začátku této kapitoly. Rozdíl spočívá v tom, že tento vzorec bere navíc v úvahu společný rozptyl metody a místo reliabilit jednotlivých proměnných obsahuje odmocniny jejich celkové kvality.

7. 3 Program pro odhad kvality dotazníkových položek Survey Quality Prediction

Mezinárodní program výběrových šetření European Social Survey je od svého počátku v roce 2002 pevně svázán s metodologickým výzkumem. Výzkumníci a výzkumnice, kteří

stáli u startu tohoto projektu, usilovali o to, aby vznikající sociálněvědní mezinárodní program vynikal mezi ostatními právě vysokou metodologickou úrovní. Byli si vědomi toho, že v mezinárodním výzkumu je leckdy odlišná kvalita dat z šetření v jednotlivých zemích příčinou toho, že analýzy založené na těchto datech přinášejí nepřesné výsledky, a chtěli se pokusit produkovat taková data, která budou k tomuto neduhu náchylná co nejméně (Harkness, van de Vijver, Mohler, 2002). ESS usilovalo o to, aby data z šetření měla co největší kvalitu, a zároveň chtělo vědět, s jakou přesností se výzkumníkům v jednotlivých zemích daří měřit. Proto byly do dotazníků ESS zaneseny experimenty 2 SB MTMM, které měly přinést přesné údaje o validitě a reliabilitě měření vybraných položek v každé zemi ESS. Kromě toho se však od začátků počítalo s tím, že se v rámci ESS během let vytvoří bohatá databáze odhadů kvality, kterou bude možné využít pro další účely. Koeficienty kvality stovek položek tak posloužily k vývoji algoritmu, který je schopen predikovat validitu a reliabilitu jakékoliv položky na základě jejích charakteristik, jako je počet slov, umístění v dotazníku, téma, kterého se položka týká, jazyk, ve kterém je otázka položena, náchylnost otázky k sociální desirabilitě a další. Tento algoritmus tvoří základ online softwaru nazvaného Survey Quality Prediction (SQP), v současné době operujícího ve verzi SQP 2.1.

Mezi analyticky a analyticky, kteří pracují s daty z kvantitativních šetření, dosud obecné povědomí o SQP příliš nepanuje, nicméně v oblasti kvantitativní sociálněvědní metodologie si projekt získal velkou pozornost. Autorský tým SQP, který tvoří především výzkumnice a výzkumníci RECSM, si za něj vysloužil uznání v podobě ocenění Americké asociace pro výzkum veřejného mínění (AAPOR), jež konstatovala, že se jedná o velmi ojedinělý a ambiciózní projekt. Rovněž jednotliví pracovníci a pracovnice RECSM jsou za svoji práci na SQP držiteli různých profesních a studentských ocenění. RECSM se snaží SQP všemožně popularizovat a dosáhnout toho, že bude používán v běžném výzkumu. Jednou z těchto popularizačních aktivit je např. program studijních návštěv RECSM hrazený z rozpočtu ESS, kde se návštěvníci dozvědí, proč a jak s SQP pracovat.

7. 3. 1 Popis programu SQP

Program SQP 2.1³⁵ má dvě hlavní funkce. První umožňuje najít v databázi odhadů reliability a validity údaje o kvalitě všech položek, které byly měřeny v 2 SB MTMM experimentech v ESS a pro které se podařilo v TS MTMM modelu získat odhad parametrů.

³⁵ Webová adresa <http://sqp.upf.edu/>

Tato funkce je velmi užitečná, neboť zjistit odhady reliability a validity ve vlastní analýze v 2 SB MTMM dat pro jednu zemi je ve většině případů nemožné. Analytičky a analytici, kteří chtějí ve své práci použít odhady kvality položek ESS, je mohou získat pouze prostřednictvím SQP.

Druhou a hlavní funkcí SQP je predikce reliability a validity jakékoliv dotazníkové položky bez ohledu na to, jestli tato položka již byla použita v nějakém šetření nebo ne. Predikční funkce používá databázi odhadů kvality z experimentů ESS, ale i z velkého množství dalších MTMM studií (Oberski, Gruner, Saris, 2011). Obrovské množství údajů, ze kterého algoritmus čerpá, má být zárukou důvěryhodnosti predikce. Algoritmus predikční funkce SQP vznikl na základě statistické analýzy vztahů mezi hodnotami koeficientů reliability a validity a charakteristikami položek a jeho popis poskytují publikace autorů programu (Oberski, Gruner, Saris, 2011; Saris et al., 2011).

Pro uživatele je postup získání predikcí z programu SQP v podstatě velice jednoduchý. Nejdříve je potřeba se registrovat na webové adrese SQP, poté lze s programem začít pracovat okamžitě. Do příslušné kolonky napíšeme znění položky, jejíž kvalitu chceme odhadnout, a škálu, na které je měřena odpověď. Následně program začne nabízet možné charakteristiky položky, ze kterých vybíráme ty, jež odpovídají naší položce. Těchto charakteristik je více než šedesát, takže proces kódování jedné položky trvá zhruba dvacet minut. Přesný popis procesu kódování je k dispozici v publikaci Sarise a Gallhoferové (2014), nicméně není třeba jej podrobně studovat, neboť program na webu během kódování u každého kroku podává nápovědu, která k pochopení příslušného kroku stačí. Po dokončení procesu kódování program vypočítá očekávanou validitu, reliabilitu a CMV dotazníkové položky.

Nabízejí se dvě cesty, jak odhady z SQP využít. Je možné je použít pro očištění dat tak, jak to ukazují v předchozím oddílu textu, nebo s nimi pracovat jako s informacemi o kvalitě navrhované dotazníkové položky. Pokud SQP odhadne validitu a reliabilitu položky nízkou nebo nedostačující, je možné uvažovat o tom, jak položku vylepšit. K případnému vylepšení SQP poskytuje návrhy na zlepšení formulace nebo na úpravu škály odpovědi. SQP je tak možné využít ve fázi přípravy dotazníku, neboť umí odhalit případné problémy ještě před tím, než je položka, resp. dotazník vyslán do terénu.

7. 4 Diskuze

Používání softwaru pro predikci kvality položek SQP vyvolává mnoho otázek. Je v sociálněvědním výzkumu a především ve výzkumu postojů a názorů vůbec možné něco predikovat? Lze z dat z šetření, která se uskutečnila v minulosti, provádět usuzování o budoucích chybách v datech? Je možné predikovat chybu v budoucím šetření, které proběhne na nějakém vzorku populace, na základě dat získaných na jiném vzorku populace? Na tyto otázky dosud neexistují přesvědčivé odpovědi. Jisté důkazy by mohl přinést experiment, ve kterém by byly použity položky, jež dosud nejsou v databázi SQP. Pro tyto položky bychom získali odhady z SQP a odhady ze skutečného MTMM měření. Odhady z obou zdrojů bychom porovnali, čímž bychom zjistili, do jak velké míry se shodují. V současné době ovšem nemám informace o tom, že by se takový experiment uskutečnil. Saris a Gallhoferová (2014:240) dokládají, že hodnoty odhadů z SQP jsou stejné nebo velice podobné těm, které tým získal z empirických měření z 2 SB MTMM. K této informaci však musím dodat, že tyto predikce kvality položek byly založeny na údajích z databáze, ve které se ty samé položky již nacházely. Získaný údaj tak byl zčásti založen na údajích, které již odhadovaný údaj obsahovaly a z toho důvodu nelze toto srovnání brát jako nezávislý test predikčních schopností SQP. Saris a Gallhoferová ještě upozorňují, že SQP se v odhadech mýlí v případech, kdy je reliabilita a validita položky ve skutečnosti nízká; v takovém případě vypočítá výrazně vyšší hodnoty koeficientů kvality a predikce je chybová (Saris, Gallhofer, 2014: 252). Dosavadní výsledky týmu RECSM jsou sice optimistické, přesvědčivé odpovědi na obecné otázky však nedávají. V současné době bohužel nejsou k dispozici jiné informace o důvěryhodnosti odhadů SQP než od týmu RECSM, tj. samotných autorů programu, které nelze považovat za nezávislé a nezaujaté kritiky. I z toho důvodu nelze přijmout konečné rozhodnutí o tom, zda predikční funkce SQP poskytuje správné odhady validity a reliability či nikoliv.

Tato dizertační práce obsahuje návody, jak získávat odhady chyb měření, a informuje o predikčním programu SQP, nedává však odpověď na to, zda je vhodné využívat tyto odhady pro korekci chyb měření v analýze substantiálních dat. Budoucnost korekce chyb měření s využitím odhadů reliability a validity ve výzkumné praxi stojí především na tom, zda ve vědecké komunitě zavládne konsensus o povaze chyb měření. V situaci, kdy neexistuje shoda o tom, zda je chyba měření rušivý rozptyl, který má negativní vliv na výsledky analýz, či zda se jedná o substantiální rozptyl, který přináší informaci svého druhu, nemůže se korekce pro chybu měření stát běžnou metodou úpravy dat, kterou je

v dnešní době např. vážení dat k zajištění reprezentativity. Ilustrace korekce pro chybu měření v této kapitole ukázaly, že analýza s očištěnými daty často dochází ke zcela jiným výsledkům, než analýza, která chyby měření nebere v úvahu. Korekce chyby měření tedy může výrazně změnit výsledky analýzy, takže rozhodnutí o tom, zda data čistit či nikoliv, je naprosto zásadní volba, která nestojí jen na osobním názoru analytika či analytičky, ale i na přijetí odborné veřejnosti.

V současné době v praktickém výzkumu mohou odhady reliability a validity sloužit především jako informace, na jejímž základě by se výzkumníci měli rozhodnout, zda data z konkrétní položky vůbec použít v analýze. Pokud analytik ví, že reliability položky, se kterou má v úmyslu pracovat, je velice nízká, měl by tuto položku nahradit položkou jinou, jež obsahuje menší proporcii chyby měření. Víme, že reliability měření může být různá a víme také, že je reálné, aby měření v sociálněvědních šetřeních bylo vysoce reliabilní. To je např. případ vybraných položek EU SILC, o jejichž vysoké reliability informovala čtvrtá kapitola této dizertační práce. Stejně tak v případě efektu použité metody měření nemusí docházet k velkým ovlivněním měření; např. souhlasný styl odpovídání nemusí měření ovlivnit tak silně, aby to mělo zásadní vliv na vztahy mezi měřenou a latentní proměnnou, jak ukazuje pátá kapitola této dizertace. Analýza kvality měření položek tedy nemusí nutně vést k odhalení katastrofální kvality dat, ale naopak může přinést pozitivní informace a učinit výsledky substanciální analýzy důvěryhodnější. Mé doporučení kvantitativním sociálním vědcům a vědkyním tedy je takové, aby se informovali o reliability a validity svých dat a na základě této informace činili rozhodnutí, zda data použít či nikoliv. V ideálním případě by měli výzkumníci údaje o reliability a validity měřených položek dodávat do svých publikací. Jsem přesvědčena, že právě tento způsob využití údajů o reliability a validity činí výzkum chyb měření opodstatněným, i přesto, že se korekce chyby měření v sociálněvědní kvantitativní analýze dosud neujala.

Tato dizertační práce představuje několik metod, resp. modelů a designů sběrů dat pro odhad chyby měření; jsou jimi kvazisimplexový model, MTMM model, resp. TS MTMM model a CF-RS model a MTMM design a split – ballot MTMM design. Všechny tyto modely aplikované na běžná data z šetření či na data z příslušného designu jsou definovány v teoreticko – analytickém rámci, který je de facto kombinací metody strukturálního modelování a rozšířenou Klasickou testovou teorií. Všechny tyto modely a designy tak fungují na principu, kdy je rozptyl měřené proměnné rozdělen na dvě či více částí; na část odpovídající měřenému konstruktovi či znaku a na část odpovídající náhodné chybě měření,

případně chybě, jež vzniká vlivem použité metody. Je důležité zdůraznit, že tyto prezentované metody plně vyčerpávají použitý princip, jinými slovy, nelze najít další, zcela odlišný model pro odhad chyb měření. Všechny ostatní modely odhadu chyb měření mohou být pouze variacemi na jeden z těchto tří modelů vztahů mezi měřenými proměnnými a latentními proměnnými, reprezentujícími náhodnou či systematickou chybu. Stejně tak nelze najít zcela nové a originální designy pro sběr dat, které lze uskutečnit v reálném výzkumu a jejichž data by bylo možné modelovat nějakým novým, inovativním způsobem. Metodologický výzkum dalších možností identifikace chyb měření prostřednictvím daného teoreticko – analytického rámce je tak velice omezený, neboť v současné době v podstatě nelze přijít s nějakou významnou a inovativní novinkou, jako je zcela nový model či design.

Odborná literatura informuje o upravených designech či modelech, které vycházejí z některé ze tří prezentovaných metod. Kvazisimplexový model byl použit jako předloha pro upravený Generalized Simplex Model Biemera, Christové a Wiesena (2009), kteří místo jedné položky v každé vlně použili data z baterie položek. Podobně Van der Veld a Saris (2005) použili v kvazisimplexovém modelu místo jednoho indikátoru pravého skóru opakovaná měření z test – retestu. V případě modelů s faktorem stylu odpovídání se nabízí varianty modelu s větším množstvím obsahových faktorů a jím odpovídajícím faktorům stylu, ve kterých lze pátrat po tom, zda byla měření všech konstruktů ovlivněna stejně silným faktorem stylu či zda bylo měření různých konstruktů ovlivněno stylem odpovídání v různé míře (Billiet, McClendon, 2000). Design MTMM je plně definován kombinováním metod a znaků a variabilní zde může být jen jejich počet, který však nesmí být nižší než tři. Andrewsův MTMM model vztahů (1984) byl inovován Sarisem a Andrewsem (1991), kteří stanovili True Score MTMM model, jenž je poslední významnou inovací v oboru modelování chyb měření. Stejně tak split – ballot MTMM design sběru dat (Saris, Sattora, Coenders, 2004) je poslední inovací v oblasti speciálních designů. Od té doby žádná podstatná inovace nepřišla a vzhledem k omezením, které mají SEM modely a praktický sběr dat ve výběrových šetřeních, asi už další nepřijde.

Období největších pokroků v modelování chyb měření v datech z výběrových šetření je zřejmě za námi, není se však třeba obávat, že metodologové a metodoložky, zabývající se měřeními v sociálních vědách, nebudou moci objevovat nové možnosti. Vzhledem k tomu, že kvantitativní sociální vědy již pomalu začaly využívat data ze sociálních sítí a další dostupné zdroje Velkých dat, je možné obrátit objevitelský zájem a uspokojení touhy po

inovacích tímto směrem. Jaké využití mají Velká data v sociálních vědách a jaké „chyby“ taková data obsahují? Je možné spojovat data z výběrových šetření s databázemi Velkých dat? Jaké chyby takové spojování generuje? Tyto a stovky dalších otázek budou muset metodologové a metodoložky v blízké budoucnosti řešit. Následující a poslední kapitola této dizertační práce poskytuje úvod k využití Velkých dat v sociálních vědách a naznačuje, jaké „chyby“ mohou Velká data obsahovat a jakým problémům budou muset čelit sociální vědci a vědkyně, které se rozhodnout s Velkými daty pracovat.

8. Velká data a internet jako zdroj dat pro výzkum: Exkurz do současnosti a budoucnosti kvantitativního sociálněvědního výzkumu

Předchozí kapitoly této práce se zabývají kvalitou dat z výběrových šetření; představují metody k identifikaci chyb měření a přinášejí informace o validitě a reliabilitě konkrétních dat, která čeští sociální vědci a vědkyně často používají ve svých studiích. Můj dlouhodobý zájem o problematiku chyb měření je motivován snahou o to, aby kvantitativní sociální věda přinášela co nejkvalitnější výsledky a aby nebyla jen hrou s čísly odtrženou od empirické reality současných společností. Považuji za nutné, aby kvantitativní sociologie a příbuzné vědy byly sebereflexivní a dodávaly informace, které v nejvyšší možné míře odpovídají skutečnému chování lidí, jejich postojům, názorům a hodnotám. Proto se zajímám o kvalitu dat z výběrových šetření, neboť právě ony jsou primární metodou kvantitativní sociologie a generují velké množství dat pro sociálněvědní analýzy.

Odbornice a odborníky z různých oblastí sociálních věd, ale i zástupce veřejnosti může napadnout, zda má výzkum kvality dat z šetření skutečný smysl. Víme přece, že metoda výběrových šetření je už poměrně stará a navíc se v poslední době ukazuje jako problematická. Nebylo by lepší využít technologického pokroku a poohlédnout se po nějaké jiné metodě? Není detailní zabývání se kvalitou dat z šetření uzavíráním se do slonovinové věže, ze které, symbolicky řečeno, není vidět na technologické inovace, které by mohli zlepšit výzkum? Proč nezačít využívat internet se všemi jeho výhodami, které klasické výběrové šetření nemá? S vývojem technologií a výpočetní síly a s rozšiřováním využívání mobilních služeb a přístrojů k připojení k internetu se výzkumníkům mohou otevřít mimořádné příležitosti k sociálnímu výzkumu. Je možné získat obrovské množství dat ze sociálních sítí jako je Twitter, Facebook nebo Instagram. Telefonní operátoři, banky, ale třeba i státní úřady disponují dříve nepředstavitelným množstvím dat, která obsahují informace o mnoha aspektech lidského chování. Všechny tyto datové zdroje by přece mohly způsobit v sociálních vědách revoluci. Navíc internet umožnil rychle a s malými náklady sbírat data prostřednictvím online platform pro vyplňování dotazníků; respondenti už nemusí nikam chodit a nikdo nemusí chodit za nimi. Proč se obtěžovat se školením tazatelů a výběrem domácností k rozhovoru?

Takové otázky a úvahy jsou na místě. Technologický pokrok opravdu přináší nové zdroje dat a metoda výběrových šetření se v současné době skutečně potýká s problémy. Prvním a zřejmě největším z nich jsou nízké návratnosti. I přesto, že se v akademickém

sociálněvědním výzkumu stále víc snažíme o nadstandardně kvalitní přípravu výběrových šetření a vytváříme nástroje pro evaluaci kvality sběru a i dat samotných, není v našich silách vyhrát boj se stále se snižujícími návratnostmi. Návratnosti klesají i navzdory úsilí investovanému do opakovaných oslovení a penězům na odměny pro respondenty. Lidé se už šetření zúčastňovat nechtějí; jsou často zaneprázdnění, nejsou k dostižení, častými žádostmi o rozhovor jsou unaveni, populace je přezkoumaná (Savage, Burrows, 2007; Couper, 2013). Na vině není jen akademický výzkum, ale především soukromé firmy a agentury, které ve velké míře zjišťují veřejné mínění či spotřebitelské chování. Tzv. průzkumy nedělají už pouze agentury, státní instituce nebo akademická sféra, ale i střední či menší firmy, které prostřednictvím krátkých dotazníků zjišťují spokojenost svých zákazníků. Snažit se v takovém prostředí o vysoké návratnosti z šetření je těžké. Další komplikací pro metodu šetření může být i to, že se mění etnické složení národních států. Ve společnostech s vysokou mírou migrace přestává mít podle Savage a Burrowse (2007) rámování populací podle fyzických hranic národních států smysl. Výběrové soubory rámované jako reprezentativní za populaci konkrétní země jsou podle nich spíše vědeckým konstruktem, než nějakým empiricky existujícím celkem. Třetím problémem, se kterým se výběrová šetření v současné době potýkají, je nepochybný fakt, že veřejnost ztrácí v tuto metodu důvěru. Velkým dílem se na tom podílely události roku 2016, britské referendum o vystoupení z Evropské unie a prezidentské volby v USA, nebo spíše způsob, jakým média referovala o výsledcích výzkumů veřejného mínění, které zjišťovaly preference lidí v těchto lidových hlasováních. Některé průzkumy, které těmto volbám předcházely, se v lecčem spletly, rozhodující vliv na veřejné mínění ale zřejmě měla média, která odmítala brát na vědomí existenci výběrové chyby a o každém procentu preference navíc pro jednu či druhou stranu referovala jako o významném rozdílu. Po volbách, které dopadly opačně, než média očekávala, nepřekvapivě přišlo snížení důvěry ve výzkumy veřejného mínění. Nízká důvěra v metodu je problém, který může mít v budoucnu vliv na samotnou její existenci; může způsobit další snižování návratností či ovlivnit rozhodování vědní politiky státu, zda vkládat peníze do výzkumu využívajícího výběrové šetření.

Nejen mezi veřejností, ale i v akademickém prostředí se v uplynulých deseti letech začaly zvedat kritické hlasy, které věští, že zlatý věk metody výběrových šetření je u konce. Takovým prvním výrazným šťouchnutím byl článek výzkumníků Mikea Savage a Rogera Burrowse, který vyšel v britském časopise *Sociology* v roce 2007. Jejich článek vzbudil

v komunitě kvantitativně zaměřených sociálních vědců a vědkyň velký zájem, byl od té doby mnohokrát citován (Burrows, Savage, 2014) a následovalo ho několik přímých reakcí, ať už pozitivních (např. Weber, 2009) či kritických (např. Crompton, 2008; Couper 2013). Množství a povaze reakcí se není třeba divit, neboť autoři článku formulovali své výhrady k výběrovým šetřením poměrně ostře. Výběrová šetření je podle nich potřeba opustit či výrazně upozadit a je nutné se přeorientovat na Velká data a sběr dat prostřednictvím internetu, neboť se jedná o mnohem lepší zdroje a metody. Od roku 2007 vyšlo v různých sociálněvědních odborných časopisech několik obecných pojednání o Velkých datech a internetových anketách a jejich vztahu k tradičnímu kvantitativnímu výzkumu. Některé texty se zaměřily na hlubší, epistemologické změny, které mohou především Velká data vyvolat nejen v sociálních vědách, ale ve vědě obecně (Kitchin, 2014; McFarland, Lewis, Goldberg, 2016; Chandler, 2015), jiné porovnávaly možnosti Velkých dat s tradičně využívanými daty a upozornily na to, že Velká data mají pro využití v sociálních vědách nezanedbatelná omezení (Couper, 2013; Boyd, Crawford, 2012; Murphy et al, 2014; Hargittai, 2015). Kromě problematiky Velkých dat zhodnocovali někteří autoři a autorky ve svých textech i celkový vliv internetu na metodu výběrových šetření, tedy de facto rozšíření metod, které spojuje fenomén dostupného internetu a možnost dosáhnout respondenta rychle, levně a s menším výdejem práce. Takovými metodami jsou např. online platformy pro vytváření dotazníků jako je Survey Monkey anebo internetová pracovní burza Mechanical Turk společnosti Amazon, kde lze za nízké poplatky získat online respondenty pro výzkumy a experimenty (Buhmester, Kwang, Gosling, 2011). Do společnosti těchto nových metod autoři začleňují i internetové opt – in panely, které využívají placené respondenty.

Metodologické debaty posledních deseti let naznačují, že se sociální vědy ocitly v nové historické situaci. V odborných textech se objevují slova o krizi a padají názory, že kvantitativní sociologové, politoložky, sociální geografové a další sociální vědci a vědkyně, které se drží kritizované metody výběrových šetření, by se brzo mohli ocitnout v pozici toho, který „chvíli stál a již stojí opodál“. Pokud se nástup internetu a Velkých dat do praxe sociálněvědního výzkumu v blízké budoucnosti ukáže jako revoluční činitel, jež změní podobu celého vědního oboru, nebyl by to v relativně krátké historii sociálních věd první významný obrat. V průběhu minulého století se objevily dvě zásadní metodologické změny, které formovaly dnešní podobu kvantitativního výzkumu. Tou první byla aplikace pravděpodobnostního výběru a úspěšné uvedení metody výběrových šetření v období po 2.

světové válce. Rozšíření metod využívajících statistické zpracování dat znamenalo do jisté míry odklon od filozofické tradice v sociologii, neboť šetření postupně vytlačovala jiné metody výzkumu (McFarland, Lewis, Goldberg, 2016). Nástup výběrových šetření do centra akademického výzkumu rozhodně neprobíhal k vůli všech soudobých sociologů, jak ukazuje např. slavné kritické pojednání Charlese Wrighta Millse o „abstraktním empiricismu“ v *Sociologické imaginaci* (Mills, 2000). Koncem sedmdesátých let minulého století došlo k další velké změně, neboť se objevily výkonné počítače, které umožnily dříve neuskutečnitelné výpočty. Rozvoj počítačového softwaru vedl k rozmachu statistické analýzy a leckdy složitému modelování dat. I tento obrat se stal předmětem kritiky; podle některých odborníků způsobilo extenzivní využívání složité statistické analýzy další odklon od teoretických konceptů a umožnilo vytváření samoúčelných studií, jejichž cílem byla spíše exhibice analytických schopností, než vědecké bádání (Schenk, 2013; McFarland, Lewis, Goldberg, 2016).

V současné době nelze Velká data a další, méně inovativní metody využívající internet, jako jsou ankety a opt – in panely, považovat za střední proud v sociálněvědním výzkumu. Pomalá adaptace sociálních věd na Velká data a další internetové zdroje sice již pomalu začala – sociální vědci a vědkyně se opatrně seznamují s daty ze sociálních sítí a uvažují o možnostech, jak využívat i jiná Velká data –, nicméně tyto zdroje ještě zdaleka nejsou v sociálních vědách využívány běžně. Akademická kvantitativní sociologie ještě stále stojí na datech z výběrových šetření a v České republice tato metoda ve své podstatě spíše vzkvétá, než aby upadala. O tom svědčí zapojení českých akademiků a akademiček do většiny velkých Evropských a světových komparativních projektů výběrových šetření (European Social Survey, International Social Survey Programme, Survey of Health, Aging and Retirement in Europe, EU SILC, European Values Study) či nedávný vznik velkého panelu domácností v rámci nákladného sociologického projektu Proměny české společnosti. Ale přesto, nebo možná právě proto, že si Velká data svou cestu do sociálněvědního mainstreamu ještě úplně neprorazila, je potřeba se jimi zabývat. Je možné a dokonce pravděpodobné, že tato data zmenší význam výběrových šetření v sociálních vědách, ale ještě před tím, než se tak stane, je nutné reflektovat jejich vlastnosti a identifikovat jejich „plusy a mínusy“. Stejně tak je potřeba zabývat se novou sociální situací, ve které se ocitli uživatelé internetu, sociálních sítí, věrnostních karet a dalších prostředků, přes která jsou Velká data sbírána, a tím, jak tato situace vytváří novou sociální realitu svého druhu. V následujícím textu se pokouším zprostředkovat čtenářům a

čtenářkám, které se s Velkými daty v sociálněvědním výzkumu dosud nesetkaly, základní informace o Velkých datech a možnostech a způsobech jejich využití ve výzkumu, stejně jako o jejich omezeních. Kromě jiného srovnávám možnosti Velkých dat s tradiční kvantitativní metodou výběrových šetření a upozorňuji na výhody a nevýhody Velkých dat vzhledem k datům z výběrových šetření. V textu se krátce zastavuji i u dalších alternativ sociálněvědního výzkumu, které umožnil internet a technologický rozvoj, tedy internetových anket a opt – in panelů. Cílem této poslední části mé dizertační práce je kromě poskytnutí informací o výzkumném využití Velkých dat i obhajoba metody výběrových šetření jako sociálněvědního datového zdroje, který má své nezastupitelné místo i v *báječném novém světě* počítačových technologií.

8. 1 Co jsou Velká data?

O Velkých datech (z angl. Big data) se v posledních letech mluví často a v podstatě všude. Název Velká data používají počítačové experti, akademici, marketéři, novináři, ale i veřejnost. V sociálních vědách se částečně ujal i alternativní název *organická data*, jehož autorem je významný sociálněvědní metodolog Robert Groves, tento termín se ale používá výjimečně (Couper, 2013; Habermann, Kennedy, Lahiri, 2016). Velká data bývají někdy definována jako soubory dat, které jsou tak velké, že je nedokáže zpracovávat běžný statistický software (Snijders, Matzat, Reips, 2012). Jak jsou tedy Velká data velká? Může to být např. objem dat a o velikosti 2,5 petabytů (peta = 10^{15}), která obsahují informace o jednom milionu nákupních transakcí amerického řetězce WalMart, jež byly zaznamenány během jedné hodiny. Nebo to může být 2 a půl miliardy komentářů, 2,7 miliardy lajků a 300 milionů fotek, které v roce 2012 nahráli uživatelé Facebooku za jeden den (Kitchin, 2014). Velká data však nejsou typická pouze svou velikostí, ale i kvalitou. Jsou to data vysoce granulovaná a detailní, někdy zachycující realitu v momentu od momentu (moment-to-moment data). Někteří datoví odborníci dokonce tvrdí, že Velká data jsou tak detailní, že mají ve své podstatě blízko k tomu zachycovat realitu jako takovou (Chandler, 2015).

Nejčastěji citovanou definicí Velkých dat je definice třemi V, z nichž každé odkazuje k jedné typické vlastnosti Velkých dat (Couper, 2013; Chandler 2015; Daas, Roos, van de Ven, Neroni, 2012). Objem (Volume) Velkých dat je tak velký, že překračuje kapacitu tradičních metod pro uchování a zpracování dat; Velocity vyjadřuje, že Velká data jsou získávána v reálném čase, tj. v okamžiku, kdy se událost skutečně děje, a různorodost či variabilita (Variety) znamená, že Velká data jsou syrová, neuspořádaná, nestrukturovaná a

nepřipravená k analýze. Další možnou definici nabízejí Boydová a Crawfordová (2012), podle kterých jsou Velká data „kulturní, technologický a vědecký fenomén“, jehož existence je dána kombinací několika činitelů. Tím prvním je *technologie*, konkrétně maximalizace výpočetní síly a algoritmické přesnosti získávat, analyzovat, spojovat a porovnávat velké soubory dat. Tím druhým je *analýza*, která umožňuje identifikovat ve velkých datových souborech struktury, což lze využít pro ekonomické, sociální a právní účely. Tím třetím je *mytologie*, konkrétně rozšíření přesvědčení, že velké datové soubory nabízejí vyšší formu poznání, jež přináší dosud neobjevené informace, a jež „má auru pravdy, objektivitu a přesnosti“.

Velká data vznikají z různých zdrojů a podle typu zdroje se obvykle rozdělují do tří skupin (Couper, 2013). Tou první jsou tzv. administrativní data, kterými disponují zpravidla státní či veřejné instituce a která vznikají pro účely veřejné či státní správy. Jsou to záznamy ministerstev, státních úřadů nebo třeba data z elektronické evidence tržeb. Druhou skupinou jsou tzv. transakční data, která jsou generována při elektronických transakcích nejrůznějšího typu. Typickými transakčními daty jsou údaje o platbách z debetních/kreditních karet, informace z věrnostních karet obchodů, záznamy telefonních operátorů o hovorech a přenosech dat, údaje z internetového vyhledávače či historie internetového prohlížeče. Transakční data jsou „automatický vedlejší produkt transakcí a aktivit“ (Couper, 2013) a jejich primárním účelem je vyřízení transakce nebo uživatelského požadavku na službu. Použití těchto dat např. pro marketingové účely je sekundárním využitím a příjemci služeb obchodníků, operátorů a bank si často vůbec nejsou vědomi toho, že data o jejich chování jsou ukládána a dále využívána (Couper, 2013). Do třetí skupiny patří data ze sociálních médií či sítí jako je Facebook, Instagram, Twitter, Tumblr a další. Tato média slouží primárně ke sdílení obsahu, jako jsou krátké texty, fotografie nebo internetové odkazy, a ke komunikaci s ostatními uživateli sítě (Murphy, Hill, Dean, 2013).

Případné využití Velkých dat v sociálněvědním výzkumu je specifické tím, že výzkumníci přímo nekontrolují jejich sběr. Velká data jsou „rutinně získávána jako vedlejší produkt institucionálních transakcí“ (Savage, Burrows, 2007) a jejich sběr není nijak designován. Tinati a kolegové (Tinati, Halford, Carr, Pope, 2014) tuto specifickou vlastnost popisují jako „sbírání dat v divočině“ (in the wild) a někdy se o Velkých datech mluví jako o „digitálních stopách“ (digital footprints), které za sebou zanechávají jednotlivci

v každodenním offline i online životě (Chandler, 2015; Golder, Macy, 2014; McFarland, Lewis, Goldberg, 2016).

Vzhledem k tomu, že již vzniklo určité množství studií využívajících Velká data a že jsou Velká data často zpracovávána pro marketingové účely, je možné ilustrovat jejich využití na reálných příkladech. Velká data např. mohou validizovat data z výběrových šetření pokud existuje podezření, že odpovědi respondentů neodpovídají realitě. Ukázkou takové studie poskytli Ansolabehere a Hersh (2012), kteří srovnávali výpovědi respondentů v internetovém dotazníkovém šetření o účasti ve volbách s administrativními daty, konkrétně s údaji z voličských registrů. Informace z registrů spojili s neanonymními daty ze šetření a porovnáním zjistili, že respondenti svou účast ve volbách nadhodnocovali. V sociálním, psychologickém či ekonomickém výzkumu nacházejí využití data ze sociálních sítí (např. Kosinski, Stillwell, Graepel, 2013; Bollen, Mao, Zeng, 2011; Lazer et al., 2014). Bollen a kolegové (2011) analyzovali vztah mezi náladou, kterou lidé vyjadřovali na sociální síti Twitter, a vývojem na burze. S pomocí tzv. mood tracking nástrojů měřili průměrnou náladu na Twitteru a zjistili, že zahrnutí této proměnné do modelu vývoje burzovního indexu Dow Jones, dokáže odhady vývoje zpřesnit. Podobně využili výzkumníci transakční data z prohlížeče Google a dokázali předurčit výskyt chřipkové epidemie (Lazer et al., 2014). Kosinski a kolegové (Kosinski, Stillwell, Graepel, 2013) ukázali, že z facebookových lajků lze predikovat charakteristiky jednotlivých uživatelů Facebooku jako je sexuální orientace, etnicita či politická orientace. Sociální sítě lze rovněž využít pro sociální a psychologické experimenty, které v tak velkém měřítku dosud nebylo možné realizovat. Např. Bond a kolegové (2012) manipulovali zobrazování určitého obsahu několika skupinám uživatelů sociální sítě, aby zkoumali vliv zobrazovaného obsahu na volební chování. Mediálně známým se stal experiment, který z vlastní iniciativy provedl Facebook v roce 2012 (Griffin, 2014) a jehož cílem bylo zjistit, zda má tato sociálních sítí schopnost ovlivňovat náladu uživatelů. Do experimentu vstoupilo celkem 700 tisíc uživatelů, kteří byli rozděleni do dvou skupin; první skupině Facebook zobrazoval pozitivní mediální obsah, druhé negativní. Výsledkem bylo zjištění, že Facebook dokáže prostřednictvím zobrazovaného obsahu uživatele činit buď smutnější, nebo šťastnější. Podobné experimenty mohou provádět pouze výzkumníci a výzkumnice, které jsou přímo navázané na provozovatele sociálních médií, takže akademické sféře jako takové je experimentování na sociálních sítích zatím nedostupné.

Velká data jsou darem z nebes pro firmy a jejich marketingové strategie. Transakční data, kterými disponují velké firmy, mohou bez výdajů na dotazování přinést validní informace o chování zákazníků. Např. americký maloobchodní řetězec analyzoval data o nákupech zákazníku za období dlouhé dvanáct let a našel v nich vztahy, kterých si do té doby nikdo nevšiml. Analýza odhalila, že některé výrobky zákazníci nakupují společně s konkrétními jinými výrobky, takže řetězec upravil umístění v zboží v regálech podle společných výskytů určitého zboží; produkt A, který se často vyskytoval v jednom nákupu s produktem B, umístil v regálu blízko produktu B, produkt C vedle D atd., čímž bylo dosaženo 16 procentního nárůstu tržeb (Kitchin, 2014). V podstatě stejný princip dlouhodobě využívá internetový obchod Amazon. Tento gigant mezi e-shopy uchovává informace o tom, co si jednotliví zákazníci v minulosti koupili, a při každé další návštěvě webu jim navrhuje, aby si koupili další zboží. Nabízí jim přesně to zboží, které si v minulosti koupili jiní lidé, kteří si mimo jiné už koupili to samé zboží, co zákazník, kterému je zobrazována nabídka. Marketingová strategie vyjadřovaná nedokončenou větou „Lidé, kteří si koupili stejnou věc jako vy, si rovněž koupili...“ se dlouhodobě vyplácí a převzali ji další webové obchody a stránky (např. IMDB, International Movie Database). Velká data k zajištění obchodního úspěchu použila i internetová televize Netflix, která produkuje vlastní televizní seriály. Při plánování natáčení seriálu House of Cards Netflix nejprve analyzoval ve své databázi preference svých předplatitelů, aby zjistil, které herce a režiséry mají nejvíce v oblibě. Na základě tohoto průzkumu obsadil do hlavní role Kevina Spaceyho a režii svěřil Davidovi Fincherovi, neboť právě tyto dva se ukázali jako neoblíbenější umělci předplatitelů Netflixu. Úspěch seriálu tak byl předem zajištěn (Barnes, 2013; Carr, 2013).

Anglický název Big Data odkazuje k fenoménu Velkého bratra (Big Brother) z Orwellova románu 1984 (Boyd, Crawford, 2012). Velká data podobně jako Velký bratr mají přehled o všem, co člověk v rámci určité služby dělá (placení kartou, prohlížení internetu, lajkování na Facebooku), a ti, kteří mají tato data k dispozici, získávají velké množství leckdy intimních informací o velkém množství lidí. Všechny tyto informace jsou navíc často získávány, aniž by si to uživatelé služeb vůbec uvědomovali. Srovnávání Velkých dat s Velkým bratrem nutí k otázkám týkajícím se ochrany osobních údajů; tento problém je ale velice komplexní, nemá jednoznačné řešení a představuje jeden z největších praktických, právních a etických problémů práce s Velkými daty (Habermann, Kennedy,

Lahiri, 2016; Couper, 2013; Boyd, Crawford, 2012; Burrows, Savage, 2009; Kalyvas, Overly, 2015; Murphy et al, 2014).

8. 2 Epistemologie Velkých dat

Vstup Velkých dat na scénu může mít vliv na podobu výzkumu v mnoha vědeckých oborech jako je biologie, genetika, lingvistika, psychologie či sociologie. V žádném z těchto oborů dosud Velká data nejsou široce rozšířeným nástrojem výzkumu, nicméně se už objevily úvahy o vlivu tohoto nového druhu dat na paradigmatické zákonitosti současné vědy. Podle některých komentátorů jsou Velká data přelomovým fenoménem, který zcela změní způsob vědecké práce. Velká data mají údajně potenciál způsobit vědeckou revoluci, neboť oplývají mimořádnými vlastnostmi, které dosud žádný vědecký nástroj či metoda neměli (Anderson, 2008; Golder, Macy, 2014; Kitchin, 2014; McFarland, Lewis, Goldberg, 2016; Chandler 2015). Ti, kteří vidí budoucnost velkých dat nad míru optimisticky, tvrdí, že ontologické vlastnosti Velkých dat umožní vědcům opustit teorie, hypotézy a přibližné modely empirické reality, a přinesou informace o světě takovém, jaký skutečně je (Anderson, 2008; Kitchin 2014; McFarland, Lewis, Goldberg, 2016).

Nejdůraznější manifest o revolučním významu Velkých dat nepřišel z akademického prostředí, ale z populárně – vědeckého internetového časopisu zabývajícího se technologiemi. Časopis Wired v roce 2008 publikoval komentář šéfredaktora Chrise Andersona, který vyjádřil velmi radikální zamyšlení nad blízkou budoucností vědy. Podle Andersona jsme svědky příchodu nového vědeckého věku, kdy technologie dokáží přinést naprosto věrný obraz reality. Tento obraz je své podstatě objektivním zachycením skutečnosti, jež má schopnost odhalit skutečnou povahu věcí. Pokud máme k dispozici tento obraz, již není zapotřebí žádná teorie, neboť data mluví sama za sebe; víme, jaké věci ve skutečnosti jsou, takže už nepotřebujeme vědecké aproximace. Že tento názor přišel z komerční, nikoliv akademické sféry je pro situaci Velkých dat příznačné; v posledních letech byla komerční sféra v adaptaci na nové technologické možnosti pružnější než akademie, je však také pravdou, že je náchylnější ke hledání senzací. Pokud na chvíli připustíme, že Velká data jsou objektivním obrazem skutečnosti a mluví sama za sebe, co to znamená pro epistemologii vědy? Podle mnoha autorů by to znamenalo de facto opuštění deduktivního způsobu poznávání, ve kterém zaujímá vrcholné místo teorie, ze které jsou odvozovány hypotézy a testovatelné modely (Kitchin, 2014; Chandler 2015). Chandler (2015) píše, že analytici, kteří pracují s Velkými daty, nepřicházejí k datům s přesnými představami o tom, co v datech hledat, ale prvotní informace o zkoumaném

fenoménu získávají nejprve z dat. Ti, kteří vidí opuštění deduktivního způsobu poznávání jako velkou šanci pro vědu, doufají, že se bude zvyšovat schopnost technologie zachycovat jevy, takže „svět začne mluvit sám za sebe bez zprostředkování omylným lidským interpretem“ (Chandler, 2015).

Andersonův článek vzbudil v akademickém prostředí velkou pozornost a přímo na něj reagovalo několik akademiček a akademiků (Kitchin; 2014; McFarland, Lewis, Goldberg, 2016; Boyd, Crawford, 2012; Chandler, 2015). Tito autoři a autorky upozornili především na to, že Anderson vychází z mylných předpokladů o ontologických a epistemologických vlastnostech dat. Velká data podle nich nejsou obrazem reality, nemluví „sama za sebe“ a nerodí objektivní, čisté vědění. I Velká data vznikají v určitých situacích a za určitých podmínek, které jsou plně dány lidskými aktéry. Prostředí, ve kterém Velká data vznikají, není sociální a kulturní vakuum a není to ani nově vytvořené „hřiště“, jehož zákonitosti a pravidla nebyla ovlivněna člověkem. Reagující akademici proto volají po tom, aby Velká data byla reflektována jako fenomén svého druhu (Kitchin, 2014; McFarland, Lewis, Goldberg, 2016; Boyd, Crawford, 2012). Podle nich jsou Velká data něčím, co skutečně mění mnohé ve vědě – mění povahu samotného vědění, ovlivňují proces výzkumu a povahu a kategorizaci reality – nejedná se ale o vědeckou revoluci, ale spíše o epistemologický posun (Kitchin, 2014; McFarland, Lewis, Goldberg, 2016; Boyd, Crawford, 2012).

V návaznosti na debatu o povaze Velkých dat vymezuje Kitchin (2014) dva možné způsoby, kterými je možné v rámci nededuktivních přístupů s Velkými daty pracovat. Ten první s nimi nakládá jako s věrným obrazem reality a analytik je v tomto případě objevitelem, který pouze dokumentuje silné vztahy mezi zachycenými jevy a zprostředkovává popisy, získané z dat. Tento přístup nazývá Kitchin *nový empiricismus* a nedává mu příliš nadějí na úspěch v akademickém výzkumu. Druhým způsobem je *data driven science*, který kombinuje přístupy abdukce, indukce a dedukce. Výzkumníci využívající tento přístup začínají práci s Velkými daty s nějakou teoretickou představou a dělají teoreticky informovaná rozhodnutí o tom, jak nejlépe nakládat s daty tak, aby odhalili informace, které jsou případně vhodné pro další výzkum. Místo toho, aby testovali všechny možné vztahy v datech, soustředí se na vztahy, které pravděpodobně budou nejužitečnější pro další výzkum a docílení výsledku analýzy. Analytici aplikují iterativní přístup, neboť se k datům vracejí a upřesňují rozpracované hypotézy, čímž se dostávají k porozumění zkoumaného fenoménu. Tento přístup je podle Kitchina možnou cestou pro

budoucnost vědeckého výzkumu a je to i jistý epistemologický posun, nikoliv však revoluce.

McFarland a kolegové (2016) poskytují popis technik, které jsou v současné době používány pro analýzu Velkých dat. Jejich anglické názvy nepřekládám do češtiny, neboť chci umožnit čtenářům, aby si je případně mohli vyhledat v zahraniční literatuře; rovněž nechci ustavovat či měnit terminologii v české počítačové vědě a IT. Popis techniky *computational ethnography* ve své podstatě odpovídá Kitchinově popisu přístupu *data driven science*, autoři dále poskytují informace o technikách *network analysis* a *machine learning*. Obě techniky využívají induktivní přístup k výzkumu, kdy „data mluví za sebe“ (network analysis) nebo kdy se algoritmy učí nacházet zákonitosti v datech do té doby, dokud nedokáží najít co nejpřesnější algoritmus pro popis vztahů v datech (machine learning). V sociálním výzkumu může být machine learning velmi užitečný pro výzkum vztahů, které nemají lineární povahu. Jeho využití by pomohlo sociálním vědám vymanit se z důrazu na analýzu lineárních vztahů (Golder, Macy, 2014).

8. 2. 1 Velká data v epistemologii sociálních věd

Kvantitativní sociálněvědní paradigma využívá deduktivní přístup poznávání, který formoval dnešní podobu sociálních věd a způsob, jakým sociální vědci přemýšlejí o sociální realitě. Sociální vědci si zcela přirozeně kladou otázky, na které hledají odpovědi deduktivním přístupem, a dá se očekávat, že se jej s příchodem Velkých dat jen tak nevzdají. Nededuktivní paradigmatické přístupy, které Kitchin (2014) označuje jako *nový empiricismus* a *data driven science*, jsou vysoce restriktivní vzhledem k množství a povaze výzkumných otázek, které lze s jejich využitím zkoumat. Jako největší omezení nededuktivního přístupu v kvantitativní analýze chápou sociální vědci především fakt, že s jeho využitím není možné pátrat po tom, proč se něco děje (Boyd, Crawford, 2012; Barnes, 2013). Velká data umožňují leckdy detailní popis chování lidí a lze s jejich využitím nacházet vztahy mezi jevy, zatím však není možné získat evidenci o tom, že se něco děje v důsledku něčeho (Boyd, Crawford 2012; Barnes, 2013). Přestože většina odborníků chápe tuto vlastnost Velkých dat jako problém, najdou se takoví, podle kterých to na obtíž není. Savage a Burrows (2007) explicitně uvádějí, že sociální vědy jsou v disciplíně zvané kauzální výzkum dlouhodobě neúspěšné a že je na čase ambice na vysvětlování příčin a následků vzdát a koncentrovat se jen na popis sociálních jevů. K těmto úkolům jsou Velká data často mnohem vhodnější, a tak je podle nich nutné se přeorientovat na jejich analýzu (Savage, Burrows, 2007).

Akademické debaty o epistemologii Velkých dat naznačují, že v sociálněvědním výzkumu budou tato data nahlížena ze dvou hledisek. Pro to první budou typické analytické přístupy data driven science, které opouštějí klasickou dedukci, a poznávání jejich prostřednictvím vychází především z dat. Ty umožní získávat popisy společnosti či lidského chování; příkladem jejich využití může být studie Webera (2009), který z obrovského množství údajů o křestních jménech a příjmení odhadl informace o etnicitě a příslušnosti ke společenské třídě jednotlivců. Velká data tak využil k získání informace, kterou by jinak mohlo přinést jen velmi rozsáhlé výběrové šetření. Tím druhým hlediskem je stávající, deduktivní přístup, který nachází využití především v analýze dat ze sociálních sítí. Již v současné době využívají analytici a analytičky klasické kvantitativní metody jako je obsahová analýza a popisná statistika k získávání informací o výskytu různých slov, klíčových slov a hashtagů, frekvencích sdílení různých obsahů či podobnosti jednotek v různých skupinách. Příklady takových prací nacházíme i v České republice (viz např. Karašáková, 2013; Hrdina 2016; Vochocová, Mazák, Štětka, 2016). Své místo ve výzkumu má i spojování dat ze sociálních sítí či transakčních dat s daty z neanonymních výběrových šetření či online dotazníků navázaných na facebookový účet a následná statistická analýza spojených dat (Kosinski, Stillwell, Graepel, 2013; Ansolabehere, Hersh, 2012).

8. 3 Velká data jako zdroj informací pro sociálněvědní výzkum

Pokud chceme Velká data používat obdobně, jako to děláme v kvantitativní sociologii dosud, tedy jako data z výběrových šetření nebo data z tisku a dalších tradičních zdrojů, musíme se zabývat všemi možnými vychýleními a nepřesnostmi, které tato nová data obsahují. Zabývat bychom se rovněž měli překážkami, které stojí v cestě tomu, aby byla Velká data používána v sociálněvědním výzkumu. Musíme věnovat pozornost všem charakteristikám, ve kterých se Velká data liší od dat z výběrových šetření, a zohlednit jejich netradiční vlastnosti ve statistické analýze a při interpretaci výsledků. Systematická analýza odchylek ve Velkých datech nebude jednoduchá, neboť tato data jsou komplexnější a komplikovanější než data z šetření. Metodologové by se ale neměli zastavit u technických záležitostí, ale měli by se zabývat i širšími souvislostmi používání Velkých dat pro výzkum, a to zejména ontologickou podstatou dat ze sociálních sítí a transakčních dat. Jaká jsou Velká data? V jakých podmínkách vznikají? Jak tyto podmínky ovlivňují chování lidí? Jaké sociální skupiny přednostně vytvářejí Velká data? V neposlední řadě by

výzkumnice a výzkumníci měli reflektovat, jak ovlivňuje využití Velkých dat samotný proces výzkumu; Velká data vznikají bez dohledu výzkumníků a jejich samotný charakter svádí výzkumníky k určité perspektivě. Jak upozorňují některé odbornice, povaha těchto dat formuje design výzkumu a ovlivňuje výzkumníky v tom, jaké si kladou otázky a výzkumné cíle (Boyd, Crawford, 2012). V následujícím textu přináším přehled zdrojů odchylek ve Velkých datech a upozornění na některé problematické aspekty Velkých dat. Postupně se zastavuji u problému náročnosti analýzy Velkých dat a jejich dostupnosti, dále diskutuji nereprezentativitu, vychýlení výběru jednotek k analýze, vliv sociálního okolí v datech a další problematické jevy.

8. 3. 1 Náročnost analýzy Velkých dat

Velkým omezením využití Velkých dat v sociálních vědách je fakt, že pro efektivní práci s nimi jsou potřeba programátorské schopnosti, které naprostá většina sociálních vědců a vědkyň nemá (Manovich, 2011; Golder, Macy, 2014; McFarland, Lewis, Goldberg, 2016; Kitchin, 2014). Aby mohla být Velká data použita pro výzkum, je potřeba je umět sbírat, skladovat a analyzovat, případně validizovat údaje z externích zdrojů. Velká data jsou navíc nestrukturována, resp. nejsou strukturována tak, aby mohla být instantně použita pro statistickou analýzu, a proto je nutné je upravovat, aby se s nimi dalo analyticky pracovat (Golder, Macy, 2014). Práce s Velkými daty rovněž v některých případech znamená i spojování souborů dat z různých zdrojů, což je komplikovaná technická disciplína. Náročnost práce s Velkými daty má za následek, že se jejich analýzou v současné době zabývají především programátoři a počítačová vědci; ti rovněž vyprodukovali většinu dosud publikovaných studií, využívajících Velká data (Golder, Macy, 2014; McFarland, Lewis, Goldberg, 2016). Angažmá IT expertů v hájemství sociálních věd podle některých autorů a autorek znamená velký problém. Většina počítačových expertů totiž nedisponuje znalostmi ze sociálních věd a jejich výzkum je proto často „redukcionistický, funkcionalistický a ignoruje vlivy kultury, politiky, veřejné politiky a kapitálu“ (Kitchin, 2014). Další autorky zase upozorňují na významnou genderovou disproporci v komunitě počítačových vědců, mezi kterými jsou nedostatečně zastoupeny ženy (Hill, 2010; Boyd, Crawford, 2012). Je známo, že gender výzkumníka nepřímo ovlivňuje tematické zaměření plánovaných studií i konkrétní výzkumné otázky a může mít vliv i na interpretaci zjištění z výzkumu (Harding, 2010; de Madariaga, 2012.). Silná převaha mužů mezi počítačovými experty, kteří analyzují Velká data pro účely sociálněvědních analýz, tak do velké míry

deformuje výzkumnou agendu směrem k tématům a problémům, které jako prioritní vidí především mužská část populace.

Na absenci sociálních vědců a vědkyň ve výzkumu s Velkými daty ukazuje i fakt, že v současné době nemáme k dispozici žádné učebnice, příručky či manuály, které by studujícím a výzkumníkům představovaly metody pro výzkum s Velkými daty. Významní výzkumníci a výzkumnice v oboru sociálních věd v uplynulých desítkách let publikovali množství nejrůznějších instruktážních textů, které vysvětlují zásady práce s daty z výběrových šetření a principy mnoha typů statistické analýzy (např. Babbie, 2015; Bollen, 1989; Saris, Gallhofer, 2014; Tabachnick, Fidell, Osterlind, 2001; Hox, Roberts, 2011), k analýze Velkých dat však oborové učebnice nejsou. A nejenže nemáme návody na analýzu Velkých dat, ale chybí i obecnější pojednání o tom, jaký typ úloh je možné s Velkými daty realizovat či jaký druh výzkumných otázek lze s jejich využitím řešit. Dosud vycházejí jen texty formou článků, které shrnují Velká data jako fenomén, který může na sociálněvědní výzkum v budoucnosti mít velký vliv (např. Boyd, Crawford, 2012; Couper, 2013; Murphy et al, 2014; McFarland, Lewis, Goldberg, 2016; Savage, Burrows, 2007).

8. 3. 2 Dostupnost Velkých dat pro sociálněvědní výzkum

Omezená dostupnost Velkých dat je pravděpodobně největší problém, kterému budou v blízké budoucnosti sociální vědci čelit. Společně s problematikou ochrany soukromí je omezená dostupnost Velkých dat považována za nejzávažnější překážku úspěšného uvedení Velkých dat do praxe sociálněvědního výzkumu (Haberman, Kennedy, Lahiri, 2016; Couper, 2013; Savage, Burrows, 2007; Boyd, Crawford, 2012; Manovich, 2011; McFarland, Lewis, Goldberg, 2016; Driscoll, 2014). Omezenost přístupu ke zdroji dat se týká transakčních dat i dat ze sociálních sítí, významně se však liší míra tohoto omezení. Data ze sociálních sítí jsou mnohem dostupnější než transakční data, což se odráží i v jejich častějším používání v sociálněvědním výzkumu (Murphy et al., 2014). Facebook, Twitter a další sociální sítě umožňují veřejnosti přístup do datového úložiště, ze kterého lze při znalosti programátorských postupů získat všechna aktuálně dostupná data. Tyto firmy však postupně mění politiku zveřejňování dat, neboť se snaží udržet si rozhodující část informací pro své účely. Informační giganty jako Facebook se v posledních letech staly datovými oligopoly, disponující obrovským množstvím informací, a uvědomují si výsadní postavení, které díky informacím získávají. Aby si toto postavení udržely, poskytují veřejnosti, tedy i výzkumníkům, stále menší objemy dat (Boyd, Crawford, 2012; Driscoll,

2014; McFarland, Lewis, Goldberg, 2016). Vědcům je údajně k dispozici pouze 1 až 10 procent veřejných tweetů (Boyd, Crawford, 2012; Driscoll, 2014), množství a typ dat, které je možné získat z Facebooku, je rovněž regulováno. V oblasti analýzy dat ze sociálních sítí tak získávají rozhodující konkurenční výhodu ti vědci a vědkyně, které jsou u firem jako je Facebook nebo Twitter přímo zaměstnány. Nejenže jim jsou dostupná všechna existující data, ale získávají i možnost realizovat na sociální síti vlastní experimenty. Jejich vědecká práce tedy může nad prací akademiků v mnohém vyniknout (Manovich, 2011; Savage, Burrows, 2007).

Administrativní a transakční data jsou sociálním vědkyním a vědcům prakticky nedostupná; administrativními daty disponuje stát a jeho úřady a podléhají přísným pravidlům ochrany soukromí, transakční data mají plně ve své moci soukromé firmy, jako jsou banky, telefonní operátoři či provozovatelé internetových vyhledavačů a prohlížečů. Pozice soukromých firem se tak oproti minulosti posiluje, neboť skutečně velké firmy disponují obrovským množstvím informací o jednotlivcích i společnosti jako celku. S tím se na druhou stranu oslabuje pozice sociálních věd, které ztrácejí monopol na popis a vysvětlování společenských jevů. Savage a Burrows (2007) pro tuto společenskou situaci používají termín převzatý od Thrifta (2005) *knowing capitalism*, který označuje kapitalistickou společnost, kde soukromé firmy mají prostředky k získávání vědění a disponují mocí toto vědění rekonstruovat a přetvářet ve svůj prospěch. V tomto společenském uspořádání ztrácejí svoji moc instituce jako je stát a na něj navázaná akademická sféra.

Neschopnost získávat Velká data pociťují sociální vědci jako obtíž, která může mít zásadně negativní vliv na budoucnost vědy a výzkumu. Metodolog Robert Groves vyzývá, aby věda v součinnosti se státem začala tento problém neprodleně řešit. Stát by měl akademické sféře nabídnout pomoc při vyjednávání s velkými firmami disponujícími Velkými daty a iniciovat vznik institucí na pomezí mezi akademickou, resp. státní a soukromou sférou, které budou spolupracovat s vlastníky Velkých dat a přesvědčovat je o nutnosti používání těchto dat ve výzkumu (Haberman, Kennedy, Lahiri, 2016). Čím konkrétně by mohli akademici velké firmy přesvědčit, aby jim data poskytovaly, však Groves neuvádí.

8. 3. 3 Nereprezentativita Velkých dat

Výzkumnice a výzkumníci, kteří běžně pracují s daty z výběrových šetření, berou za samozřejmost, že jejich data jsou reprezentativní nebo kvazireprezentativní za populaci národního státu či za jinou, specifickou populaci (studenti, ženy na rodičovské dovolené atp.). Většina z nich se proto hlasitě ohrazuje proti sociálněvědní analýze, která využívá data ze sociálních médií či jiná Velká data. Někteří akademici a akademičky, které ve svých textech reagovaly na návrhy začlenit Velká data do výzkumu, vyjádřily námitku, že „velikost není všechno“, a že obrovské objemy Velkých dat nejsou schopny vyvážit to, že zahrnutí jedinci proporcčně neodpovídají obecné populaci (Boyd, Crawford, 2012; Couper, 2013; Hargittai, 2015; McFarland, Lewis, Goldberg, 2016; Golder, Macy, 2014).

Problém reprezentativity se týká jak dat ze sociálních médií, tak transakčních dat. Transakční data firem obsahují jen údaje o lidech, kterým daná firma dodává služby; jsou to tedy např. uživatelé platebních karet jedné banky či klienti konkrétního telefonního operátora. V marketingovém výzkumu pro účely firmy nemusí být nereprezentativita problém, pokud jsou ale taková data využívána pro účely sociálněvědního výzkumu, musí být analýza doprovázena důrazným upozorněním, že výsledky nejsou zobecnitelné na populaci národního státu. Obdobně populace lidí využívajících sociální sítě není v současné době stejná jako obecná populace. Z analýz existujících profilů na Twitteru a na Facebooku vyplývá, že např. v USA jsou uživatelé sociálních sítí mladší, vzdělanější a bohatší (Golder, Macy, 2014); jednotlivci na síti se od obecné populace liší i v pohlaví a rase (Hargittai, 2015). Ani v Česku nezahrnují sociální sítě celou populaci; o demografickém složení profilů na Facebooku informuje např. Dočekal (2015).

Porušení zásady reprezentativity se z hlediska soudobé praxe zdá jako velké omezení Velkých dat. Mohlo by však být užitečné oprostít se od striktních požadavků a uvažovat o situacích, kdy je porušení reprezentativity vyváženo nějakou konkrétní a významnou výhodou. Velká data disponují některými vlastnostmi, které data z výběrových šetření nemají, a které lze využít ve výzkumu. Víme např. jak těžké je prostřednictvím výběrových šetření získávat od respondentů údaje o běžných činnostech, kterým se věnují v průběhu celého dne. Také např. víme, že respondenti vypovídají o svém chování s chybami, neboť si přesně nepamatují detaily jako v kolik hodin jedli nebo v kolik hodin šli spát. Velká data nám mohou pomoci právě při získávání tohoto typu informací, obzvláště pokud by bylo možné využít Velká data z několika zdrojů. Data z platebních a věrnostních karet umožňují zjistit, v jakou denní dobu dělají lidé běžné nákupy nebo v jaké dny nakupují speciální

zboží jako je nábytek nebo zahradnické náčiní; data z vyhledávačů, prohlížečů nebo navigačních aplikací dodávají např. informace o tom, v jakou dobu lidé zhruba začínají pracovat či jakým způsobem a za jakým účelem se pohybují ve veřejném prostoru; data z různých mobilních aplikací, které lidé využívají ke sportu, dietě nebo kontrole spánku a mnoha dalším činnostem, jsou rovněž zdrojem velice detailních informací o běžných činnostech a návycích. Kdyby sociální vědkyně a vědci měli přístup k různým zdrojům sociálních dat a dokázali data z různých zdrojů spojovat, naznačené výhody Velkých dat by rozhodně převážily nad problémem nereprezentativnosti. Otázkou však je, zda se sociální vědci k takovým datům někdy dostanou a mnohem palčivější otázkou je, jak nebezpečná je sociální situace, kdy existují spojené soubory dat, obsahující detailní informace o životech obrovského množství lidí.

8. 3. 4 Výběr jednotek z Velkých dat a odchylka výběru

Výběry dat ze sociálních sítí, ale i transakčních dat obsahují jednotky různého typu; mohou to být lidé, nákupy, platby kartou, tweety, fotografie, lajky, statusy a další jednotky informací (Murphy et al., 2014). Tyto výběry jsou vždy do jisté míry odchýlené od dané populace a od skutečných hodnot platných pro danou populaci. Obecně se o Velkých datech dá říci, že obsahují jen informace pocházející od jednotlivců, kteří učinili nějakou akci, případně informace o jednotkách obsahu, které mají nějakou charakteristiku. Do výběrů se dostávají jen lidé, kteří mají věrnostní kartu Tesca, lidé, kteří použili hashtag, lidé, kteří zaplatili platební kartou, lidé, kteří sdílí na Facebooku fotografie, případně hashtagy, které byly použity v daném období, fotografie, které byly zveřejněny v daném období na sociální síti atp. Couper (2013) tuto vlastnost vyjadřuje výstižnou větou „Big data tends to focus more on the “haves” and less on the “have-nots”.

V důsledku způsobu výběru dat k analýze vypadává z analyzovaných souborů Velkých dat podstatná část lidí, kterých se případný zkoumaný problém týká. Např. když používáme data ze sociálních sítí, musíme mít na paměti, že existuje více možností uživatelské aktivity, než je jen pouhá existence profilu na síti. Jak velká je proporce lidí, kteří sociální médium aktivně používají? A odpovídají jednotlivé profily skutečně existujícím lidem? Sociální sítě jsou plné „posluchačů“, tedy lidí, kteří především konzumují obsah ostatních uživatelů, ale sami vytvářejí velmi málo obsahu, případně se neprojevují vůbec (Boyd, Crawford, 2012). Jistou část profilů tvoří duplicitní účty a není ani známé množství falešných účtů, které jednotlivci zakládají z různých důvodů. Skutečností je i to, že lidé mají různou pravděpodobnost stát se aktivním uživatelem sítě (Murphy et al., 2014); např.

manuálně pracující mohou být v obsahu sociálních médií reprezentováni méně než lidé, kteří tráví naprostou většinu dne v kanceláři u počítače. Samostatný problém tvoří účty lidí, kteří jsou placeni za šíření propagandy. Tito lidé na sítích diskutují jako normální uživatelé, ovšem s jasným účelem šířit propagandu (King, Pan, Roberts, 2016; Murphy et al., 2014). Příbuzným problémem je aktivita robotů, kteří vytvářejí obsah, jež je na první pohled nerozeznatelný od obsahu, který vytváří skutečný uživatel (Boyd, Crawford, 2012; McFarland, Lewis, Goldberg, 2016; Bessi, Ferrara, 2016; Murphy et al., 2014). Roboty mohou být např. falešné twitterové účty, využívané pro účely politického boje s cílem ovlivnit diskurz na sociální síti. Existují sice metody, jak v datech identifikovat takové robotické účty, sociálněvědní výzkumníci je však běžně neovládají (Bessi, Ferrara, 2016).

Potenciální problém výzkumu dat ze sociálních sítí spočívá ve výběru, který je veden nějakým klíčem, jež přímo souvisí s tématem výzkumu. Hargittaiová (2015) tento jev nazývá „vybírání jednotek na základě závisle proměnné“ a uvádí příklad psychologické studie, která zkoumala vztah mezi typem osobnosti a způsobem používání jazyka. Do souboru dat byli zahrnuti lidé, kteří se přihlásili do facebookové aplikace nazvané My personality, jež byla opatřena popisem, že uživatel díky ní zjistí informace o své osobnosti. Do výběru se tak dostali pouze lidé, kteří měli iniciativu zjišťovat něco o svých psychologických vlastnostech. Taková iniciativa je však podle Hargittaiové explicitním projevem zájmu o znalost vlastní psychologie, která je vlastní určitému typu lidí. Tuto vlastnost vybraných jednotek však výzkumníci ve svém textu nediskutovali a neproblematizovali, což Hargittaiová považuje za chybu. Obdobný problém představuje výběr z dat Twitteru na základě hashtagu, klíčového slova, kterým uživatelé označují své tweety. Tweetuje-li např. někdo o amerických prezidentských volbách, přidá ke svému tweetu hashtag #presidentialelection nebo #USselection či nějaký jiný, s tématem spojovaný hashtag. Pro studii, jejímž cílem je analýza twitterových příspěvků týkajících se prezidentských voleb, jsou pak vybrány jen ty tweety, které byly označeny relevantními hashtagy. Do výběru se ale nedostanou relevantní tweety o prezidentských volbách, které hashtagem označeny nebyly. Výsledky takové studie proto nelze zobecňovat na populaci lidí, kteří v daném období tweetovali o volbách. V české prostředí studii se stejným druhem výběru dat z Twitteru realizovala např. Karašćáková (2016).

K odchylce výběru se váže i fakt, že lidé používají konkrétní sociální síť ke sdílení konkrétního typu obsahu (Hargittai, 2015). Např. Facebook lidé primárně používají k verbální, psané komunikaci nebo k odkazování na jiný internetový obsah, naopak na

Instagram dávají jenom fotky. Jednotliví uživatelé často neduplikují obsah na svých facebookových a instagramových účtech, takže obsahy těchto médií se nepřekrývají, ale doplňují. Pokud analyzujeme obsah na sociálních sítích a nemáme spojená data ze všech profilů uživatelů na různých sociálních sítích, nemůžeme mít celkový přehled o obsahu, který uživatelé sdílí přes všechny své profily na sociálních sítích. Zkoumá – li např. analytik vztah mezi lajkováním na Facebooku a fotografiemi, které uživatel nahrává na svůj FB profil, můžou mu unikát podstatné informace, neboť významná část uživatelových fotek se možná nachází na Instagramu. V analýzách je proto nutné explicitně uvádět, že usuzování z dat z jednoho média se týká pouze a výhradně jen účtů na konkrétním zkoumaném médiu.

8. 3. 5 Množství proměnných ve Velkých datech

Jeden ze základních rozdílů mezi Velkými daty a daty z výběrového šetření není jen v množství případů v jednom souboru, ale v i množství proměnných. „Velká data obsahují hodně případů a málo proměnných, zatímco data z šetření obsahují hodně proměnných, ale málo případů“ (Couper; 2013). Zásadním nedostatkem je přitom velmi malé množství demografických proměnných či jejich úplná absence. Např. internetové prohlížeče, které generují transakční data, často nemají údaje o demografii uživatele, a tak musí základní demografické údaje odhadovat z vlastních dat o vyhledávání. Tyto odhady ovšem obsahují spoustu chyb; Google odhaduje pro účely svých výzkumů gender uživatelů, je jej však schopen identifikovat správně pouze v 75 procentech případů (Couper, 2013). Ani na sociálních sítích není identifikace genderu a dalších demografických proměnných zcela přesná; uživatelé o sobě často demografické informace neudávají, případně poskytují nepravdivé údaje. Savage a Burrows (2007) navrhuje, aby sociální vědci tento problém řešili inspirací v marketingovém výzkumu, kde se prosadilo používání proměnné *místo bydliště*. Tato proměnná je údajně velmi silným prediktorem, schopným nahradit základní demografické proměnné jako je třída, etnicita či vzdělání (Weber, 2004). Místo bydliště by prý mohlo suplovat množství demografických proměnných, se kterými sociální věda pracuje dnes; proti tomuto nápadu se však již zvedly kritické hlasy (Crompton, 2008). Zajímavou ukázkou odvozování demografických proměnných přinesl Weber (2009), který ve Velkých datech z údajů o křestním jménu a příjmení dokázal s vysokou pravděpodobností odhadnout etnicitu a společenskou třídu jedince.

8. 3. 6 Nejistota ohledně motivací k akci

Společně s omezeným množstvím proměnných, kterými Velká data disponují, je užitečné upozornit i na kvalitativní omezení Velkých dat. Transakční data poskytují výhradně údaje o chování (např. platba kartou) a provedených akcích (např. vyhledávané heslo v Googlu), neříkají nám ale nic o motivacích k tomuto chování či o postojích a hodnotách, které člověka k chování či akci vedou. Data ze sociálních sítí dodávají o něco méně informací o chování, mohou ale sloužit jako zdroj informací o postojích a hodnotách (Couper, 2013). I dat ze sociálních sítí se však často týkají pochybnosti ohledně motivace uživatele k akci a nejistota ohledně toho, co pozorovaná akce vlastně znamená. Retweet na Twitteru může být problematický v tom, že nevíme, proč uživatel/ka retweetuje; retweetem může danému sdělení vyjadřovat podporu, ale rovněž jím upozorňovat na obsah, který je naopak v rozporu s jeho či jejím přesvědčením; typickým příkladem takového subverzivního tweetu je v současné době retweetování obsahu twitterového účtu Donalda Trumpa. Podobně si nemusíme být jisti motivacemi k lajkování facebookové stránky; lajkoval uživatel stránku protože s ní sympatizuje nebo proto, že má zájem o obsah této stránky bez ohledu na to, zda vyjadřuje jeho postoje a přesvědčení? Mohl ale také lajkovat jenom proto, že stránku lajkovali jeho facebookový přátelé a on nechtěl zůstat pozadu. Příkladem neznámé motivace k akci v transakčních datech je vyhledávání v internetovém vyhledávači. O tom, jak velkou roli hrají motivace k vyhledávání, se přesvědčili výzkumníci, kteří na Velkých datech realizovali studii zjišťující rozšířenost užívání drogy zvané *Salvia Divinorum* (Murphy, Hill, Richards, 2011). Domnívali se, že ke zjištění rozšířenosti drogy je možné použít data z vyhledavače Google a z Twitteru, indikátorem rozšířenosti měla být podle jejich předpokladu četnost vyhledávání a tweetování o této látce. Z dat Googlu a Twitteru zjistili frekvence výskytu klíčového slova *Salvia Divinorum* v určitém období a porovnali je s daty projektu National Survey of Drug Use and Health, kde respondenti vypovídali, zda tuto konkrétní látku užili či nikoliv. Porovnání obou datových zdrojů ukázalo, že ve Velkých datech byla četnost výskytu mnohem větší než v datech z šetření, což donutilo výzkumníky pátrat po příčinách. Zjistili, že bylinu ve zkoumaném období kouřila zpěvačka Miley Cyrus na oslavě svých osmnáctých narozenin a video, zachycující tuto událost, se dostalo na Youtube (jedná se mimo jiné o velice zábavné video plné smíchu a veselí). Příčinou velkého výskytu drogy ve vyhledávání Googlu a na Twitteru tedy nebyla rozšířenost užívání, ale drogový večírek známé zpěvačky.

8. 3. 7 Vliv sociálního okolí a závislost pozorování

Lidé, kteří svojí aktivitou v offline světě nebo na internetu vytvářejí digitální stopy, jsou pod velkým vlivem bezprostředního sociálního okolí. Na sociálních sítích se tento vliv může projevat jako cenzura, kdy sociální okolí aktivně potlačuje konkrétní projevy uživatelů na síti, nebo jako autocenzura svého druhu, kdy uživatel omezuje svůj projev z důvodu obav z reakce sociálního okolí. Následkem cenzury i autocenzury je, že některé jednotky pozorování jako jsou komentáře, lajky, tweety či jiný typ obsahu jsou buď vymazány cenzorem, nebo vůbec nevzniknou. Cenzura na sociálních sítích jako je Facebook či Instagram někdy probíhá ze strany provozovatele sociální sítě, kdy jsou vymazávány příspěvky, které odporují zásadám komunity (např. pornografie, otevřená nenávisť proti skupině obyvatel), je však také často uplatňována majiteli jednotlivých účtů či profilů. Mohou to být politické strany, osobnosti z veřejného života či celebrity, kdo si na svých profilech nepřejí příspěvky určitého typu, a vymazávají komentáře, které považují za nevhodné, či blokují uživatele, které chápou jako problematické. Cenzuru rovněž provádějí značky a firmy, které si nepřejí, aby jim uživatelé nevhodnými komentáři kazili pečlivě budovanou image značky. Obsah na sociální síti je tak často důsledně filtrován z hlediska zájmů firem, politiků a dalších významných lidí, kteří si často pořízují zaměstnance, aby likvidovali nevhodný obsah.

Autocenzura uživatelů se projevuje mnoha způsoby od budování vlastní image po chování v diskuzích a lajkování obsahu, a je důsledkem existence společenských norem, ale i přímého kontaktu s ostatními uživateli sítě. Právě přímý vliv sociálního okolí je činitelem, který je zodpovědný za jeden z několika hlavních rozdílů mezi daty z výběrových šetření a daty ze sociálních sítí. Data z výběrových šetření jsou statisticky zpracovávána a musí splňovat podmínky statistické analýzy, proto v nich musí být jednotlivá pozorování na sobě nezávislá. Nezávislost pozorování ovšem není vlastností, kterou disponují data ze sociálních sítí. Jednotky pozorování v datech ze sociálních médií, ať už jsou to uživatelé, komentáře, tweety, lajky atd., na sobě nezávislé nejsou, neboť vznikají v určitém kontextu, daném napojením uživatele na jiné uživatele. Sociální média obsahují obrovské množství menších „sociálních sítí“, tedy vzájemných či jednostranných propojení uživatelů, jakýchsi subpopulací, uvnitř kterých panují mocenská pravidla zčásti daná reálnými mocenskými silami ve společnosti, vtělenými v charakteristiky jako je pohlaví, vzdělání či etnicita, zčásti nově vytvářená v síti. Vzhledem k mocenskému rozvržení uvnitř subpopulací a faktu, že majitel profilu může jakéhokoliv uživatele okamžitě zablokovat, přizpůsobují

uživatelé své chování na síti subpopulaci, ve které se nacházejí, či subpopulaci, která má kontrolu nad profilem, kde aktuálně diskutují. Uživatelé upravují sdílený obsah tak, aby byl více či méně v souladu s pravidly subpopulace, a postupem času se učí, co si mohou a nemohou dovolit psát nejen na profily ostatních uživatelů, ale i na svůj profil; po určité době strávené na síti ví, jaký názor či postoj je vhodné vyjádřit, jaká slova je vhodné používat v určitých subpopulacích atd. Cílem přizpůsobení se pravidlům v subpopulacích je minimalizace sankcí za chování, postoje či názory, se kterými většina v subpopulaci nesouhlasí. Sankcemi na síti zpravidla bývají negativní komentáře ostatních uživatelů, či zablokování přístupu na profil, případně zablokování účtu. Ještě větší hrozbou však mohou být anticipované reakce extrémně agresivních uživatelů, kteří diskutéra či diskutérku, se kterou nesouhlasí, mohou krátkodobě i dlouhodobě šikanovat (Kopecký, Krejčí 2010).

V sociálněvědním výzkumu je potřeba brát vliv sociálního okolí na chování uživatelů sociálních sítí velmi vážně a zohledňovat jej v průběhu celé studie. Během doktorského studia jsem narazila na studentskou práci, na jejímž příkladu lze dobře ilustrovat, jak je zahrnutí vlivu sociálního okolí významné. V roce 2014 jsem oponovala bakalářskou práci nazvanou *Meze vtípnosti*, jejímž cílem bylo zodpovědět otázku „Jak jsou v současné společnosti stanovovány meze vtípnosti, tedy o čem je možné žertovat a o čem již nikoliv?“ (Halada, 2014). Autor pro výzkum použil data z facebookové stránky zaměřující se na černý humor; předmětem analýzy byly komentáře pod fotografiemi s vtipy na stránce. Autor studie se kromě jiného zaměřil na tzv. *superiorní humor*, tedy humor, který ponižuje ženy, příslušníky etnických skupin, příslušníky minoritní sexuální orientace, lidi s postižením atp., a sledoval reakce návštěvníků stránky na tento typ humoru. Analyzoval postoje a názory na konkrétní vtipy a sledoval množství komentářů, které humor chválily, a těch, které byly naopak proti němu. Výsledky výzkumu prezentoval jako obecná zjištění o lidech, kteří používají český Facebook. Kromě toho, že se autor dopustil té chyby, že interpretoval výsledky jako obecně platné a nezohlednil problematický výběr (ke vtípům se vyjadřovali jen návštěvníci stránky), dopustil se omylu v tom, že ignoroval vliv převládajícího diskurzu na facebookové stránce a *symbolického násilí* na příslušnicích diskriminovaných skupin. V analyzovaném diskurzivním prostoru měli převahu členové komunity facebookové stránky specializující se na humor ponižující příslušníky minorit, přičemž členy této komunity zjevně byli především bílí, deklarativně heterosexuální muži. V takovém diskusním prostoru určuje mocenskou převahu nejen příslušnost ke komunitě příznivců stránky, ale i reálné společenské vztahy. Uživatelé, kteří se vysmívají ženám,

Romům, černochům a homosexuálům, mají nad případnými protestními hlasy jiných uživatelů obrovskou převahu, neboť za nimi stojí masivní zásoba předsudků a existující sociální, kulturní a ekonomická nerovnost. Taková facebooková stránka v žádném případě nepředstavuje svobodné virtuální fórum, kde lze bez obav vyjádřit jakýkoliv názor, naopak, v takovém prostředí má hlas, který se vyslovuje proti prezentovanému humoru, velice slabou pozici. Je pravděpodobné, že uživatelé z řad diskriminovaných skupin či jejich zastánců, kteří se k obsahu stránky mohli dostat virálně, tedy přes aktivitu facebookového přítele, se v takovém prostoru raději zdrželi komentování, neboť se oprávněně obávali sankcí. Uživatel či uživatelka, která se v takovém prostředí rozhodne argumentovat proti prezentovanému obsahu, musí počítat s diskusní bitvou a palbou urážek, a v nejhorším případě riskuje, že mu či jí budou někteří uživatelé posílat do soukromé pošty výhrůžky zabitím či znásilněním (Soukalová, 2016). Rozvržení moci v diskusním prostoru tedy nevyhnutelně vede k tomu, že určité skupiny lidí některé svoje názory vůbec nevyjadřují.

Vliv sociálního okolí může aktivitu uživatelů nejen potlačovat, ale také jí podporovat či moderovat. Takový tlak může vytvořit např. vysoká popularita příspěvku na sociální síti vyjádřená množstvím lajků. Sama popularita příspěvku může ovlivnit postoj uživatele k zobrazovanému obsahu; nejenže se např. na Facebooku vysoce populární obsah zobrazuje více lidem, čímž může získat více lajků, ale i samotný fakt, že příspěvek je populární, může v uživatelích vzbudit sympatii k zobrazovanému obsahu, kterou uživatel vyjádří lajkem (Maybury et al., 2013). Aktivita ostatních uživatelů v síti hraje velkou roli v diskuzích, kde uživatelé přímo reagují jeden na druhého, a většina komentářů je přímo závislá na ostatních komentářích.

Je evidentní, že na sociálních sítích je vliv sociálního okolí silný v míře, která je nesrovnatelná s tlakem, který je uplatňován na respondenta v rozhovoru dotazníkového šetření s tazatelem či bez něj. Motivace uživatelů k projevení preference či názoru jsou na sociálních sítích zcela jiné než v šetření, neboť obsahují inklinaci k chování skupiny, ovlivnění ostatními uživateli, nebo obavy z projevení vlastního názoru. Naopak v dotazníkovém šetření je v momentě dotazování přítomno mnohem méně činitelů, které ovlivňují prezentovaný postoj či názor a panuje v něm často výrazně příjemnější klima než např. v diskuzi na Facebooku, kde je uživatel neustále obklopen dalšími lidmi. Ve výzkumném rozhovoru se respondent do značné míry vyjadřuje nezávisle na mínění ostatních lidí, neboť neví, jak odpovídají ostatní respondenti ve výběru, a navíc se neobává,

že prezentovaný názor vzbudí hlasitou kritiku. U facebookových lajků a komentářů je však vliv sociálního okolí přímý a uživatelé se nevyjadřují nezávisle na ostatních. Nezávislost pozorování v datech z výběrových šetření je kvalita, které nemohou data ze sociálních sítí konkurovat.

Vliv sociálního okolí hraje svoji roli i v transakčních datech, zde má však spíše charakter odchylky vzniklé v důsledku sociální desirability, než přímého ovlivnění ostatními lidmi. Z transakčních dat mohou vypadávat údaje reflektující chování, jež lidé považují za společensky nežádoucí, jako je např. nákup alkoholu či erotických pomůcek. Nákupy tohoto zboží mohou být výrazně častěji placeny v hotovosti, nikoliv platební kartou či bankovním převodem (Couper, 2013). Pokročilejší uživatelé internetu mohou rovněž maskovat svůj pohyb v online prostoru, pokud mají pocit, že se dopouštějí sociálně nevhodného chování. Dalším příkladem vlivu sociální desirability v transakčních datech může být selektivní používání mobilních aplikací zaměřených na fitness a zdravou výživu, které svým uživatelům poskytují informace o kalorických hodnotách jídel, množství spotřebované energie při určitých typech pohybu atd. (např. www.kaloricketabulky.cz). Uživatelé těchto aplikací si zde vedou dietní deníky a provozovatelé aplikace získávají informace, které lze využít pro marketingové účely. Na data z těchto aplikací však nelze plně spoléhat, neboť uživatelé nemusí být v reportování o své dietě zcela upřímní.

8. 3. 8 Online vs. Offline život

Sociální vědci a vědkyně často upozorňují, že uživatelé se na sociálních sítích nechovají úplně stejným způsobem jako v běžném životě (Boyd, Crawford, 2012; Golder, Macy, 2014). V prostředí, kde se většinou komunikuje psaným textem, kde je jistá míra anonymity a uživatelé k sobě někdy přicházejí bez předchozí vzájemné znalosti z reálného světa, se u lidí mohou projevit vlastnosti, které v offline světě neuplatňují, např. agresivita. Podle Goldera a Macyho (2014) si však uživatelé do virtuálního prostředí přenášejí především své skutečné charakterové vlastnosti a z chování na síti se dá o psychologických a sociálních vlastnostech jednotlivce usuzovat leccos, např. člověk, který vyhledává ve svém životě úspěch a obdiv ostatních se tak bude chovat i na sociální síti.

Sociální síť poskytuje relativně velkou kontrolu nad sebezprezentací, čímž lidem umožňuje budovat si image způsobem, který byl před vznikem internetu nemožný (Pospíšilová, 2016). Sociální média mají oproti reálnému světu tu výhodu, že na pomyslné Goffmanovské jeviště sociálního jednání, se většinou může dostávat jen to, co o sobě sám

uživatel zveřejní (pomineme-li extrémní případy, kdy uživatelé zveřejňují citlivý obsah týkající se jiných uživatelů). Kontroly nad zveřejňovaným obsahem využívá v menší či větší míře každý uživatel sociálních sítí, a buduje tak vlastní obraz, který o sobě chce šířit na veřejnosti (Couper, 2013; Manovich, 2011). V tom uživatelům pomáhají i funkce sociálních sítí, které umožňují skrýt před veřejností jakýkoliv obsah, který vznikl v emotivní chvíli či v důsledku snížené pozornosti. V jazyce výběrových šetření se jedná o jakýsi haló efekt, kdy si jednotlivci jsou vědomi toho, že jsou sledováni a přizpůsobují tomu své chování; nezveřejňují fotografie, kde jim to nesluší, a vědomě sdílí jen informace, o kterých se domnívají, že je v očích ostatních příliš nepoškodí.

Chování lidí na síti dále ovlivňují uživatelé, kteří jsou na síti a často i v offline světě silnými hráči. Těmito silnými hráči jsou facebookové či instagramové účty firem a značek, které ovlivňují uživatele přímo prostřednictvím svých účtů nebo zprostředkovaně přes celebrity a bloggery, kteří dané značky propagují (Miles, 2015). Uživatelé jsou prostřednictvím účtů vyzýváni, aby ke svým fotografiím přidávali konkrétní hashtagy či se fotili s daným produktem, např. za účelem účasti v soutěži. Často jsou také nabádáni, aby komentovali určitý obsah, a je jim sděleno, o čem konkrétně mají v komentáři psát; mohou být rovněž požádáni, aby lajkovali určitý obsah či ho sdíleli na svých profilech. Některé velké značky jako Apple nebo Starbucks se již dostaly do tak silné pozice, že uživatelé používají jejich produkty jako statusový symbol a sdílejí fotografie těchto produktů na sociálních sítích sami od sebe, aniž by je k tomu musel kdokoliv vyzývat (Neal, 2015).

Kromě výše uvedených specialit online světa, které je nutné mít na paměti, když používáme data ze sociálních sítí k výzkumu, je potřeba vzít v úvahu i fakt, že analýza dat ze sociálních sítí může přinášet zkreslený obraz sociálních vztahů mezi lidmi. V prostředí sociálních sítí může docházet ke vzniku vzájemných, leckdy silných vazeb mezi lidmi, kteří se ve skutečném světě nikdy nepotkali. Ze silné vazby identifikované na sociální síti tak nelze se stoprocentní jistotou usuzovat na silnou vazbu v reálném světě, ze které vyplývají sociálním vědám známé benefity jako je např. pomoc při získání zaměstnání. Nevalidní informace o sociálních vazbách se však netýkají jen sociálních médií, matoucí mohou být i údaje z transakčních dat. Např. GPS data z mobilních telefonů lze použít k rekonstrukci vzájemné prostorové blízkosti s ostatními lidmi, nemělo by z nich ale být automaticky usuzováno na kvalitu a sílu vztahu mezi lidmi, kteří spolu tráví hodně času. Vezměme si příklad zaměstnankyně, která tráví devět hodin denně s kolegy, se kterými sedí v kanceláři, a jen dvě hodiny týdně se svou matkou. S kolegy sice tráví mnohem více

času, její skutečná vazba k nim je však výrazně slabší, než vazba k vlastní matce (Ruppert, Law, Savage, 2013). Ruppert a kolegové (2013) upozorňují, že než začneme používat nové technologie, jako je sběr Velkých dat prostřednictvím chytrých telefonů, je nutné analyzovat a reflektovat novou sociální situaci samu o sobě a „konceptuálně uchopit specifické vlastnosti digitálních přístrojů a dat, která generují“.

8. 4 Internetové ankety

Rozšíření internetu do většiny domácností i pracovišť umožnilo soukromým firmám nejen prodávat prostřednictvím e-shopů, ale i levně a rychle získávat informace o svých zákaznících prostřednictvím elektronických dotazníků. Výzkumné prostředí tak velice rychle zaplevelily internetové ankety obchodů, zpravodajských a dalších médií či různých zprostředkovatelů služeb. V současné době však internetové ankety nevyužívají jen soukromé firmy pro marketingové účely. Kromě velkého množství anket, které připravují studenti vysokých škol jako zdroj dat pro studentské práce, využívají internetové ankety dokonce i akademičtí výzkumníci. Příkladem akademické internetové ankety v sociálních vědách je společný projekt České televize a Sociologického ústavu AV ČR nazvaný Generation What? (Červenka, 2016). V rámci této kooperace vznikla rozsáhlá internetová anketa, které se zúčastnilo 45 tisíc respondentů. Cílem výzkumu bylo zmapovat postoje a názory dospělých Čechů a Češek ve věku cca do 35 let a prezentovat je ve vysílání České televize. K tomuto ojedinělému využití ankety v akademickém sociologickém výzkumu je však nutno dodat, že výzkumný tým Sociologického ústavu pro účely projektu realizoval navíc výběrové šetření, které de facto sloužilo jako skutečný zdroj informací o dané věkové skupině, a data z ankety byla jen zpestřením televizního vysílání.

Velmi podobný projekt využívající internetovou anketu pro sociologický výzkum se uskutečnil ve Velké Británii. Akademici Burrows a Savage (2014) se podíleli na projektu televize BBC, která prostřednictvím svého vysílání a webových stránek zmobilizovala přes 300 tisíc lidí, kteří vyplnili online dotazník. Někteří britští akademici kritizovali využití ankety v akademickém výzkumu a upozorňovali, že z dat pocházejících ze samovýběru nelze činit validní inference. Podle Burrowse a Savage sice mají data z internetových anket tu nevýhodu, že nejsou reprezentativní, mohou však přinést informace, které klasické výběrové šetření poskytnout neumí. Uvádějí příklad využití metadat o přesném čase, kdy respondenti vyplňovali internetový dotazník. Z těchto dat výzkumníci zjistili, že příslušníci prekariátu měli čas vyplňovat především ráno a dopoledne, zatímco movitější příslušníci elit vyplňovali buď velmi brzo ráno, nebo až večer. Z tohoto zjištění Burrows a Savage

usoudili, že elity jsou zaměstnané již od časného rána a na rozdíl od nižších sociálních tříd nemají přes den čas brouzdat po internetu. Autoři upozorňují, že nemá smysl ztrácovat data z internetových anket jako taková a že je nutné vždy přemýšlet nad tím, jaké výhody mohou určitá data přinést. Potom je s takovými daty potřeba poučeně pracovat při vědomí všech omezení, které mají.

8. 4. 1 Internetové ankety a neanonymní výzkumník

Agenda metodologie výběrových šetření obsahuje zájem o anonymitu respondenta, jejímž účelem je ochrana osobních údajů, ale i snaha o zvýšení návratnosti (výzkumníci nechtějí, aby se respondenti báli účastnit výzkumů kvůli obavám o své soukromí) a potřeba získávat validní hodnoty proměnných (respondent se nesmí bát říct pravdu). Se vstupem internetových anket na scénu akademického výzkumu se ale objevil nový rozměr anonymity ve výzkumu; rozšíření internetu umožnilo vznik neanonymního výzkumníka, který je celebritou svého druhu, jež si prostřednictvím internetu vytváří silnou fanouškovskou základnu, kterou dlouhodobě využívá ke sběru dat. Výzkumník už tedy nutně není někdo, o kom respondenti nic nevědí. V běžných výběrových šetřeních nebo i v menších sběrech dat, jako jsou třeba psychologické experimenty, mají respondenti či probandi buď omezené, nebo žádné informace o tom, kdo výzkum vede, jaké jsou jeho či její východiska a hypotézy a jaké výsledky výzkumnice či výzkumník publikuje.

V současné době však existuje možnost se s výzkumníkem prostřednictvím internetu seznámit, sledovat jeho výsledky a zároveň se stát jeho respondentem či probandem. V českém prostředí je takovým výzkumníkem – celebritou Prof. RNDr. Jaroslav Flégr, CSc., který si svoji popularitu vydobyl blogem na serveru lidovky.cz, vystupováním v dalších médiích jako je např. Česká televize a aktivitou na sociálních sítích. Flégr svůj výzkum realizuje prostřednictvím webové stránky (www.pokusnikralici.cz), kde jsou umístěny dotazníky a informace o výzkumných projektech. Dotazníky na webu mohou respondenti vyplňovat volně, i bez registrace, a vypovídají v nich o svém chování a zdravotních indispozicích, v některých případech prezentují i své postoje a názory, vypovídají o své politické orientaci, hodnotí různé situace, rozhodují o morálních dilematech nebo hodnotí přitažlivost mužů a žen na fotografiích. Data z těchto dotazníků jsou využívána pro studie, které vycházejí v odborných časopisech z oboru sociobiologie (např. Flégr, Kuba, 2016). Prostřednictvím facebookové stránky Pokusní králičí vyzývá Flégr a jeho výzkumný tým své fanoušky, aby vyplňovali dotazníky, a průběžně je informuje o výzkumech a jejich výsledcích. Dále podává informace o nových publikacích,

kteře byly vytvořeny s použitím dat z veřejných dotazníků, a stručně shrnuje hlavní zjištění publikovaných studií. Facebooková stránka Pokusní králíci měla na začátku roku 2017 přes 14 tisíc fanoušků a daří se jí získávat další; Flégr získává nové potenciální respondenty prostřednictvím vystoupení v médiích, např. po jeho posledním vystoupení v České televizi v lednu 2017 přibylo na facebookové stránce 250 fanoušků (viz Příloha 8. 1).

Profesor Flégr svou mediální aktivitou a využíváním fanoušků – respondentů vytváří pozoruhodný výzkumný kruh; své fanoušky, kteří jsou zároveň jeho respondenty, zjednodušeně a popularizační formou průběžně informuje o teorii, ze které vychází, a o výsledcích, které získal, a takto informované je nechává odpovídat na dotazníkové položky. Poté je znovu informuje o výsledcích předchozího výzkumu, a znovu je vyzve k účasti v dalším výzkumu. Není přehnané tvrdit, že tento „fanouškovský“ způsob výzkumu může významně ovlivnit výsledky. Jistá část Flégrových respondentů se ve zkoumané problematice alespoň částečně orientuje; někteří respondenti, kteří se výzkumu účastní dlouhodobě, o zkoumané problematice určitě hlouběji přemýšleli a jejich výpovědi, ale možná i jejich chování tím může být ovlivněno. Sám profesor Flégr však toto potenciální ovlivnění nezohledňuje a možnou odchylku způsobenou tím, že se část respondentů ve zkoumaném problému orientuje a přemýšlí o něm, ve svých publikacích neproblematizuje.

8. 5 Internetové opt – in panely

Další výzkumnou metodou, kterou umožnilo rozšíření internetu, jsou výzkumy prostřednictvím online panelu. Ke vzniku tohoto druhu výběrového souboru přispěla ve velké míře zvyšující se neochota populace participovat ve výběrových šetřeních. Tzv. opt – in panel je tvořen dobrovolníky, kteří jsou dlouhodobě k dispozici a prostřednictvím online aplikací vyplňují různé dotazníky, zaměřené na spotřebitelské chování či veřejné mínění. Účast panelistů v jednotlivých výzkumech si provozovatelé panelu zajišťují tím, že respondenty za vyplnění dotazníku platí. Panely respondentů hrají velkou roli v komerčním výzkumu např. v USA, kde mají pravděpodobností výběry v současné době velice nízké návratnosti pohybující se pod deseti procenty (Pew research Center, 2012). Couper (2013) upozorňuje, že existence opt – in panelů zapříčinila stav, kdy je „poměrně velká proporce“ dotazníků v komerčním výzkumu vyplňována relativně úzkou skupinou lidí. Jeden placený panelista je často členem několika panelů a je ve své podstatě profesionálním respondentem. Výsledky z těchto panelů tak vycházejí z vychýleného výběru a i manifestované chování a postoje profesionálních respondentů se mohou lišit od chování a

postojů běžné populace. Dalším problémem je, že panelisté jsou zpravidla placeni za vyplnění dotazníku, nikoliv za členství v panelu. Respondenti, jejichž hlavní motivací pro účast v panelu je vydělávání peněz, jsou tak motivováni k vyplnění co největšího množství dotazníků. Z toho důvodu mohou někteří nepravdivě vypovídat o svých vlastnostech, motivech, chování a postojích. Některé dotazníky spotřebitelského chování např. filtrují respondenty hned v úvodní otázce, aby získali odpovědi jen od skupiny respondentů, která zajímá klienta. Dotazník např. může začínat otázkou „Plánujete si v blízké budoucnosti koupit auto?“ a respondenti, kteří odpoví negativně, jsou z dalšího odpovídání vyloučeni a nedostanou odměnu za vyplnění dotazníku. Některí zkušenější respondenti, kteří chtějí odměnu za dotazník, na tuto otázku odpoví souhlasem, i přesto, že nákup auta neplánují, čímž se v datech objevují nevalidní hodnoty.

8. 6 Společná budoucnost Velkých dat, internetu a výběrových šetření

V současné době se zdá, že se budoucnost sociálněvědního výzkumu bez Velkých dat a dalších nástrojů využívajících internet neobejde. Velká data lze již dnes považovat za potenciálně převratný fenomén, který může do výzkumu přinést mnoho inovací. Některí nevítají Velká data s optimismem a domnívají se, že obrovské, nestrukturované datové soubory nemají pro sociální vědy velké využití a že obsahují především šum. Analýza těchto dat bude údajně přinášet především nevalidní výsledky (Barnes, 2013). Pro část výzkumníků však Velká data představují výzvu a vzácný zdroj informací o lidských společnostech, který může sociálním vědám pomoci dále se vyvíjet. Americký metodolog Robert Groves vidí zapojení velkých dat do výzkumu jako nevyhnutelné; podle něj bude v následujících deseti letech potřeba začít propojovat data z výběrových šetření s Velkými daty a přizpůsobit této nové praxi designy výběrových šetření (Habermann, Kennedy, Lahiri, 2016).

Pokud jsou Velká data využívána pro kvantitativní sociálněvědní výzkum a jsou statisticky zpracovávána, je potřeba reflektovat všechny jejich chyby a odchylky, jakožto i speciální sociální situaci, ve které tato data vznikají a která se liší od situace v offline světě. Tato reflexe se týká nejen dat ze sociálních sítí, ale i různých typů transakčních dat jako jsou data z vyhledávačů, platebních karet či GPS dat z mobilních telefonů. Všechny Velkých dat se týkají problémy jako je nereprezentativita, problematické vytváření výběru, nedostatek proměnných a nutnost proměnné odhadovat na základě zaznamenaného chování, silný vliv sociálního okolí, který je významný především v datech ze sociálních sítí, a mnoho dalších vlastností, které tato data odlišují od dat z výběrových šetření. V datech ze sociálních sítí

rovněž musíme brát ohled na závislost pozorování, která znemožňuje používání běžné inferenční statistiky.

Kromě odchylek a chyb, které vznikají v důsledku zmíněných problémů, existují ve Velkých datech další chyby, které jsou důsledkem procesu získávání dat ze zdroje, editování a spojování s dalšími datovými zdroji. Velké množství odchylek a chyb ve Velkých datech zmiňuje jako vysoce problematický aspekt nových dat Robert Groves, podle kterého se však s většinou těchto chyb bude možné vypořádat prostřednictvím již známých modelů pro odhadování chyb měření v datech. Z hlediska nových příležitostí, které umožňují Velká data, oceňuje dlouhodobou snahu metodologů v oblasti výzkumu chyb měření v datech z šetření, jež přinesla řadu postupů, které bude možné využít právě v analýze validity Velkých dat (Habermann, Kennedy, Lahiri, 2016).

Společně s Grovesem, Savagem, Burrowsem a dalšími se i já domnívám, že Velká data mohou být navzdory všem chybám a nedostatkům pro kvantitativní sociálněvědní výzkum v mnohém přínosná. V první řadě mohou sociálním vědám umožnit získávat druh informací, ke kterým vědci dosud neměli přístup nebo je mohli získávat jen s vysokými náklady a s velkým množstvím chyb (Savage, Burrows, 2007). McFarland, Lewis a Goldberg (2016) zdůrazňují, že ideální vytěžení více zdrojů Velkých dat umožní v budoucnosti získávat informace na úrovni sociálních systémů, takže bude možné zjišťovat v dříve nemyslitelném rozsahu, jak se vzájemně ovlivňují a proplétají ekonomická, sociální a kulturní sféra společnosti. Tyto nové znalosti podle nich způsobí významný posun v sociologické, politologické a další teorii a zasadí se o rozvoj těchto věd; umožní rovněž zefektivnění veřejných politik. Podle Goldera a Macyho (2014) najdou data ze sociálních médií a transakční data jako jsou záznamy z telefonních rozhovorů či emaily velké využití ve výzkumu sociálních sítí a interakcí a pomohou významně přispět k oživení těchto oblastí výzkumu. Velká data rovněž nepochybně rozšíří agendu metodologie kvantitativních sociálních věd, neboť bude potřeba najít řešení pro spojování různých druhů dat a identifikaci různých chyb v těchto datech.

Dalším přínosem Velkých dat v sociálních vědách může být jejich vliv na změnu organizace vědecké práce. Tradičně existuje v sociologii a v příbuzných oborech poměrně malá dělba práce. Je běžné, že jeden výzkumník či výzkumnice obstará veškeré úkoly v procesu zpracování jedné studie od stanovení výzkumných cílů, přes rešerši literatury a statistickou analýzu až po grafickou úpravu obrázků v textu publikovaného článku. Malá či

žádná dělba práce se odráží v množství autorů publikovaných studií; autorství v sociálních vědách se většinou připisuje jednomu či dvěma autorům, vícečetné autorství je výjimkou (McFarland, Lewis, Goldberg, 2016). Stanou – li se Velká data běžným datovým zdrojem sociálních věd, bude nutné začít stavět výzkumné týmy, jejichž členové budou zodpovědní za různé druhy analýz. Již teď jsou nároky na schopnosti úspěšných výzkumníků poměrně vysoké – ideální kvantitativní sociolog či socioložka by měl/a mít nadprůměrné nadání v humanitní, ale i technické oblasti – a přibude-li mezi vyžadované schopnosti ještě zvládnutí pokročilého programování, nebude lehké takové superženy a supermuže pro sociálněvědní akademický výzkum získat. V budoucnosti by tedy vedoucí výzkumníci měli začít vytvářet heterogenní týmy, ve kterých budou kromě sociologů, politoložek a dalších sociálních vědců a vědkyň datoví odborníci, zodpovědní za programování databází a analýzu Velkých dat. Takové zvýšení dělby práce by se navíc mohlo velmi pozitivně projevit v kvalitě výzkumu a publikací. Domnívám se, že pokud výzkumné týmy nezačnou angažovat odborníky na práci s Velkými daty, k rozšíření Velkých dat v sociálních vědách dojde jen těžko.

Třetím možným pozitivem rozšíření analýzy Velkých dat by mohlo být osvěžení přezkoumané populace, unavené častými žádostmi o vyplnění dotazníku či účast ve výzkumném rozhovoru. Pokud by soukromé firmy a agentury pro výzkum trhu a veřejného mínění začaly ve větší míře využívat Velká data, snížila by se poptávka po respondentech. Pro zjišťování spotřebitelského chování jsou Velká data stejně či možná více vhodná, než data ze šetření, takže pokud k nim budou mít agentury a firmy přístup, sníží se množství realizovaných výběrových šetření. Lidé by dostávali méně nabídek k účasti ve výzkumném rozhovoru, což by se mohlo pozitivně promítnout do návratnosti akademických výběrových šetření, a tím do kvality dat ze šetření. Pokud nás v budoucnosti čeká tento scénář, nebudou nakonec Velká data něčím, co výběrová šetření zničí, ale naopak fenoménem, který zvýší jejich kvalitu a důvěryhodnost.

Svou roli v sociálněvědním a komerčním výzkumu hrají kromě Velkých dat i další inovace, které přinesl internet, a to především internetové ankety či opt – in panely. V současné době, kdy může kdokoliv vyvěsit dotazník či anketu o jedné otázce na internet a získat odpovědi od stovek, tisíců či desetitisíců lidí, se stávají ankety silnou a ve své podstatě nekalou konkurencí výsledků výzkumů veřejného mínění či akademických šetření. Akademická sféra by měla upozorňovat veřejnost na to, že data z internetových anket nejsou reprezentativní, a to ani v případě, kdy obsahují údaje od desetitisíců lidí.

Měla by také počítat s tím, že ani důrazné upozornění na nereprezentativitu dat nemusí k příjemcům sdělení dojít. Obezřetnost je na místě především v případech, kdy nereprezentativní data využívají akademici. Využívání nereprezentativních dat ve výzkumu s sebou nese to riziko, že se upozornění na nereprezentativitu dat a implikace nereprezentativity ztratí v tiché poště, tvořené řetězcem *výzkumníci – média – veřejnost*. Akademici využívající nereprezentativní data by měli být na takové profesní úrovni, že upozornění na nereprezentativitu samozřejmě umísťují do svých publikací, a čtenáři z řad odborné veřejnosti by neměli mít problém vzít tento fakt na vědomí. Ovšem v situaci, kdy jsou výsledky nereprezentativní studie sdělovány veřejnosti, ztrácí se důležité informace o metodologii již v době, kdy novináři informace zpracovávají, a pokud se přece jen zpráva o nereprezentativitě dostane do mediálního sdělení, bere ji na zřetel jen část publika. S výsledky je pak ve veřejném diskurzu zacházeno jako s poznáním reflektujícím celou populaci a zkreslené výsledky se pak např. mohou stát základem pro vytváření veřejných politik.

Jsem přesvědčena, že ve věku internetových anket a opt – in panelů by akademická sféra měla být garantem kvality a spolehlivosti. Měla by investovat více úsilí a peněz do přípravy kvalitních výběrových šetření a ručit za vysokou kvalitu získaných dat. Rovněž by si měla vážit svých respondentů a vycházet jim vstříc; v době, kdy se výběrová šetření potýkají s kritikou veřejnosti a nízkými návratnostmi, by v dotaznících nemělo být místo pro nadbytečné otázky, nesrozumitelné či složité formulace a nereálné požadavky. Nadále bychom měli klást silný důraz na kvalitu výzkumných nástrojů a situace dotazování, nadále bychom měli prostřednictvím statistických analýz kontrolovat kvalitu dat ze šetření. Akademici by si měli být vědomi, že je mnohem lepší mít menší množství kvalitních dat, než velké množství nekvalitních dat, a s tímto vědomím přistupovat k plánování výzkumných projektů, ve kterých by měla mít své místo sekundární analýza vysoce kvalitních dat jakými jsou např. data projektu European Social Survey.

Velmi těžko se odhaduje, jak bude vypadat sociálněvědní kvantitativní výzkum v budoucnosti, ve které zřejmě bude hrát technologie stejnou nebo ještě větší roli než v současnosti. Kvantitativní vědci a vědkyně dosud ještě neví, jak kombinovat tradiční zdroje dat s Velkými daty a nemají ani jasnou představu o tom, k čemu a jak Velká data využívat. Výhled navíc zastírá fakt, že internet se velice rychle mění; mění se nejsilnější internetové firmy (kdo si dnes vzpomene na prohlížeč Yahoo či síť MySpace, služby, jejichž slávu dávno odvál čas?), mění se rovněž prostředí a pravidla sociálních sítí jako je

Facebook, a mění se populace, která využívá nové technologie. Toto dynamické prostředí má za následek, že každé pojednání o Velkých datech rychle zestárne. V tuto chvíli lze s určitostí říct jen to, že v současné době nemohou Velká data v sociálněvědním výzkumu nahradit výběrová šetření, neboť jejich povaha neumožňuje zodpovídat podstatnou část výzkumných otázek. Proto by si sociální vědy měly výběrová šetření ponechat, minimálně do té doby, kdy nás technologie opanuje natolik, že bude zaznamenávat veškeré dění v našich životech vteřinu po vteřině, a výpočetní síla bude natolik silná a samostatná, že bude ze spleti yottabytů extrahovat smysluplné proměnné a nalézat mezi nimi spojitosti. Otázkou je, zda ve společnosti s tak extrémním stupněm kontroly bude vůbec povolena existence sociálních věd jako takových. Přinejmenším sociologie, často kritická a politickými i podnikatelskými elitami zpochybňovaná, ve společnosti Velkého bratra asi příliš vítána nebude.

9. Resumé

This dissertation aims at the domain of measurement error in social science survey data. To conceptualize and estimate measurement errors it employs the analytical and theoretical framework that stems from the analytical method of structural equation modeling (SEM) and Classical Test Theory (CTT), extended with the component of the systematic measurement error. This thesis has two goals that may contribute to development and extension of Czech social science methodology. The first goal is to illustrate methods of measurement error estimation, which has not been used for analysis of Czech data yet, and to point out to some problematical aspects of these methods. The second goal is to employ presented methods to obtain new findings regarding the quality of data from Czech surveys.

The dissertation presents three empirical studies, each of which uses one of the methods defined within the presented theoretical and analytical framework. First study presents an analysis of reliability of measurement with the Quasi Simplex Model (QSM). It illustrates how to use the model and brings optimistic results regarding the reliability of the Czech EU SILC panel data. In the second study the confirmatory factor analysis model, operationally called CF-RS model, is applied. This model specifies a systematic measurement error that emerges as a consequence of the acquiescent response style (ARS). The analysis revealed that the Czech acquiescent response style is obviously different from the ARS identified in data from Western European countries. The CF-RS model did not confirm a theoretically expected correlation between the ARS and age, resp. education in Czech data. Since this correlation has been previously reported in Western Europe, the lack of such correlation in the Czech data is considered a Czech cultural specificity. The third study aims at the experimental design called 2 split – ballot MTMM, repeatedly used in the European Social Survey (ESS), and analysis of this data in the True Score MTMM model. Scholars involved in the ESS methodological research specified this experimental design and the model as an analytical tool to estimate the reliability and the systematic measurement error that originates as a consequence of a measurement method. In the analysis it is shown that such data are uneasily analyzed and that it is very hard to obtain estimates of the quality of data. The text aims at the problematical aspects of the analysis and explains why the problems emerge. The following chapter illustrates how to employ estimates of reliability and systematic measurement error in the substantial analysis. The text encourages survey data analysts to be concerned in reliability and validity of

measurement and to aspire to employ survey items of known and high quality. Concluding chapter of the thesis considers the future of surveys in the era of Big Data and the internet and evaluates the present situation, when Big data has not pervaded the academic research yet. The chapter also calls attention to different errors and bias in different kinds of Big data and data from non – representative internet polls and opt – in panels.

10. Bibliografie:

Achen, Ch. H. (1975). Mass Political Attitudes and the Survey Response. *American Political Science Review* 69 (4), 1218 – 1231.

Alwin, D. F. (2007). *Margins of Error. A Study of Reliability in Survey Measurement*. Wiley New Jersey.

Alwin, D. F., J. A. Krosnick. (1991). The Reliability of Survey Attitude Measurement. The Influence of Question and Response Attributes. *Sociological Methods and Research* 20 (1), 139 – 181.

Alwin, D. F., J. A. Krosnick. (1989). Aging and Susceptibility to Attitude Change. *Journal of Personality and Social Psychology* 57, 416 – 425.

Anderson, Ch. (2008). The End of theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*.

Andrews, F. (1984). Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. *Public Opinion Quarterly* 46, 409 – 442.

Ansolabehere, S., E. Hersh. (2012). Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate. *Political Analysis*, 20 (4), 437 – 459.

Babbie, E. R. (2015). *The Practice of Social Research*. Nelson Education.

Barnes, T. J. (2013). Big Data, Little History. *Dialogues in Human Geography*, 3(3) 297 – 302.

Bentler, P. M., CH. Chou. (1987). Practical Issues in Structural Modeling. *Sociological Methods Research* 16 (1), 78 – 117.

Bessi, A., E. Ferrara. (2016). Social Bots Distort the 2016 US Presidential election Online Discussion. *First Monday*, 21 (11).

Biemer, P. B., S. L. Christ, Ch. A. Wiesen. (2009). A General Approach for Estimating Scale Score Reliability for Panel Survey Data. *Psychological Methods* 14 (4), 400 – 412.

Biemer, P. B., R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, S. Sudman. (2004). *Measurement Errors in Surveys*. New Jersey: Wiley.

- Billiet, J. B., M. J. McClendon. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling* 7 (4), 608 – 628.
- Billiet J. B., B. Cambré, J. Welkenhuyseb-Gybels. (2002). Equivalence of Measurement Instruments for Attitude Variables in Comparative Surveys, Taking Method Effects into Account: The Case of Ethnocentrism. In *Developments in Social Science Methodology*, Anuška Ferligoj and Andrej Mrvar (Editors) Metodološki zvezki, 18, Ljubljana: FDV.
- Billiet, J. B., E. Davidov. (2008). Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design. *Sociological Methods & Research* 36 (4), 542 – 562.
- Blok, H., W. E. Saris. (1983). Using Longitudinal Data to Estimate Reliability. *Applied Psychological Measurement* 7 (3), 295 – 301.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, J, H. Mao, X. Zeng. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science* 2 (1): 1 – 8.
- Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, J. H. Fowler. (2012). A 61 Million Person Experiment in Social Influence and Political Mobilization. *Nature*, 489 (7415) 295 – 298.
- Boyd, D., K. Crawford. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, communication & society*, 15 (5), 662 – 679.
- Buhrmester, M., T. Kwang, S. D. Gosling. (2011). Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives of Psychological Science* 6(1).
- Burrows, R., M. Savage. (2014). After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology. *Big Data & Society*, 1 (1).
- Byrne, B. (2009). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. New Jersey.

- Campbell, D. T., D. W. Fiske. (1959). Convergent and Discriminant Validation by the Multitrait – Multimethod Matrix. *Psychological Bulletin* 6, 81 – 105.
- Carr, D. (2013). Giving Viewers What They Want. *The New York Times*
 Online: <http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html>
- Cernat, A., P. Lugtig, S. C. N. Uhrig, N. Watson.(2014). Assessing and Relaxing Assumptions in Quasi-simplex Models. *ISER Working Paper Series*.
- Červenka, J. (2016). Generation What? Závěrečná zpráva z výsledků ankety a reprezentativního šetření.
 Online: <http://www.ceskatelevize.cz/program/porady/11467315381/download/zaverecnazprava.pdf>
- Chandler, D. (2015). A World without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium-Journal of International Studies*, 43(3) 833 – 851.
- Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. *Ideology and discontent*. London: Free Press of Glencoe, pp. 206 – 261.
- Crompton, R. (2008). Forty Years of Sociology: Some comments. *Sociology*, 42(6) 1218 – 1227.
- Daas, P., M. Roos, M. van de Ven, J. Neroni. (2012). Twitter as a Potential Source for Official Statistics in the Netherlands. *CBS Discussion Paper*.
- de Leeuw, E. D., J. Hox, G. Snijkers. (1998). The effect of Computer-Assisted Interviewing on Data Quality. A Review, pp. 173 – 198 in Blyth, B. *Market Research and Information Technology*. ESOMAR Monograph.
- de Madariaga, S. (2012). Structural Change in Research Institutions: Enhancing Excellence, Gender Equality and Efficiency in Research and Innovation. Report of the Expert Group on Structural Change. European Commission.
 Online: https://ec.europa.eu/research/science-society/document_library/pdf_06/structural-changes-final-report_en.pdf

- DiStefano, Ch, R. W. Motl. (2009). Self-Esteem and Method Effects Associated With Negatively Worded Items: Investigating Factorial Invariance by Sex. *Structural Equation Modeling* 16, 134 – 146.
- Ding, L., W. F. Velicer, L. L. Harlow. (1995). Effects of Estimation Methods, Number of Indicators per Factor, and Improper Solutions on Structural Equation Modeling Fit Indices. *Structural equation Modelling* 2 (2), 119 – 144.
- Dillman, D. A., L. M. Christian. (2005). Survey Mode as a Source of Instability in Responses across Surveys. *Field Methods* 17 (1), 30 – 52.
- Dočekal, D. (2015). Velký pohled Facebooku na české uživatele: co mají nejvíce v oblibě? *Lupa*. Online: <http://www.lupa.cz/clanky/velky-pohled-facebooku-na-ceske-uzivatele-co-maji-nejvice-v-oblibe/>
- Driscoll, K., S. Walker. (2014). Big Data, Big Questions/ Working within a Black Box: Transparency in the Collection and Production of Big Twitter data. *International Journal of Communication* 8.
- Ehrhardt, J. J., W. E. Saris, R. Veenhoven. (2000). Stability of Life-Satisfaction Over Time. Analysis of Change in Ranks in a National Population. *Journal of Happiness Studies* 1, 177 – 205.
- Erikson, R. S. (1979). The SRC Panel Data and Mass Political Attitudes. *British Journal of Political Science* 9 (1), 89 – 114.
- Feldman, S. (1989). Measuring issue preferences: The problem of response instability. *Political Analysis*, pp. 25 – 60.
- Ferjenčík, J. (2000). *Úvod do metodologie psychologického výzkumu*. Praha: Portál.
- Flegr, J., Kuba, R. (2016). The Relation of Toxoplasma Infection and Sexual Attraction to Fear, Danger, Pain, and Submissiveness. *Evolutionary Psychology*, 14 (3).
- Golder, S. A., M. W. Macy. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, 40, 129 –152.

Green, D. P., B. Palmquist. (1990). Of Artifacts and Partisan Instability. *American Journal of Political Science* 34 (3), 872 – 902.

Greenleaf, E. A. (1992a). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research* 29, 176 – 188.

Greenleaf, E. A. (1992b). Measuring Extreme response style. *Public Opinion Quarterly* 56, 328 – 351.

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. *PS: Political Science & Politics*, 48 (1), 80 –83.

Griffin, A. (2014). Facebook Manipulated Users' Moods in Secret Experiment. Independent. Online: <http://www.independent.co.uk/life-style/gadgets-and-tech/facebook-manipulated-users-moods-in-secret-experiment-9571004.html>

Groves, R. M. (2004). *Survey Error and Survey Costs*. New Jersey: Wiley.

Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, R. Tourangeau. (2004). *Survey Methodology*. Wiley New Jersey

Groves, R. M., L. Lyberg. (2010). Total Survey Error: Past, Present and Future. *Public Opinion Quarterly* 74 (5), 849 – 879.

Habermann, H., C. Kennedy, P. Lahiri. (2016). A Conversation with Robert Groves. *Statistical Science*.

Harding, S. (2010). Feminism, Science and the anti-Enlightenment Critiques, in *Women, Knowledge and Reality: Explorations in Feminist Philosophy*, eds A. Garry, M. Pearsall, U. Hyman, Boston, MA, 298 – 320.

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The Annals of the American Academy of Political and Social Science*, 659 (1) 63 – 76.

Harkness, J. A., F. J. R. van de Vijver, P. P. Mohler. (2002). *Cross – Cultural Survey Methods*. New Jersey: Wiley

- Harzing, A. (2006). Response styles in cross-national survey research: a 26 – country study. *International Journal of Cross Cultural Management* 6 (2), 243 – 266.
- He, J., F. J. R. Van de Vijver. (2015). Effects of a General Response Style on Cross Cultural Comparisons: Evidence from the Teaching and Learning International Survey. *Public Opinion Quarterly* 79, 267 – 290.
- Heise, D. R. (1969). Separating Reliability and Stability in Test-retest Correlation. *American Sociological Review* 34, 93 – 101. Hooper, D., J. Coughlan, M. Mullen. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods* 6 (1), 53 – 60.
- Hidalgo, C. A. (2014). Saving Big Data from Big Mouths. *Scientific American*.
- Holada, M. (2014). *Meze vtípnosti*. Bakalářská práce. ISS FSV UK.
- Hox, J., de Leeuw, E, and Snijkers, G. (1998). Fighting Nonresponse on Telephone Interviews; Successful Interviewer Tactics In A. Koch, R. Porst (eds) *Nonresponse in survey research*. Zuma Mannheim.
- Hox, J., J. K. Roberts (eds.). (2011). *Handbook of advanced multilevel analysis*. Psychology Press.
- Hrdina, M. (2016). Identity, Activism and Hatred: Hate Speech against Migrants on Facebook in the Czech Republic in 2015. *Naše společnost*.
- Cheung, G. W., R. B. Rensvold. (2000). Assessing Extreme and Acquiescence Response Sets in Cross Cultural Research Using Structural Equations Modeling. *Journal of Cross Cultural Psychology* 31 (2), 187 – 212.
- Chylíková, J., M. Buchtík. (2017). Validity of the Construct of Right – Wing Authoritarianism and its Measure in Post – socialistic Region: A Case of the Czech Republic. *Journal of Social Research and Policy* (v tisku).
- Chylíková, J. (2016). Odhad reliability vybraných položek z českých šetření EU SILC kvazisimplexovým modelem. *Sociologický časopis / Czech Sociological Review* 52 (1), 79 –106.

Chylíková, J. (2015). Měření validity a reliability otázek v šetření European Social Survey a jeho využití. *Data a výzkum - SDA Info* 9 (1), 5 – 29.

Chylíková, J. (2011). Úvod do problematiky výzkumu citlivých témat ve výběrových šetřeních. *Data a výzkum - SDA Info* 5 (2), 185 – 203.

Jagodzinski, W., S. M. Kuhnel. (1987). Estimation of Reliability and Stability in Single-Indicator Multiple-wave Models. *Sociological Methods and Research* 15 (3), 219 – 258.

Johnson, T., P.Kulesa, Y. I. Cho, S. Shavitt. (2005). The Relation Between Culture and Response Styles. Evidence from 19 Countries. *Journal of cross-cultural psychology* 36 (2), 264 – 277.

Jöreskog, K. G., D. Sörbom. (2004). LISREL 8.7 for Windows (Computer software). Skokie, IL: Scientific Software International, Inc.

Jöreskog, K. G., D. Sörbom. (1996). *Structural equation modeling*. Workshop presented for the NORC Social Science Research Professional Development Training Sessions. Chicago.

Jöreskog, K. G. (1970). Estimation and Testing of Simplex Models. *British Journal of Mathematical and statistical Psychology* 23 (2), 121 – 145.

Joye, D., M. Sapin. (2016). Report to the ISSP General Assembly on monitoring work undertaken for the ISSP by the Methodology Committee.

Judd, Ch. M., M. A. Milburn. (1980). The Structure of Attitude Systems in the General Public: Comparisons of a Structural Equation Model. *American Sociological Review* 45 (4), 627 – 643.

Kaminska, O., A. L. McCutcheon, J. B. Billiet. (2010). Satisficing Among Reluctant Respondents in a Cross-National Context. *Public Opinion Quarterly* 74 (5), 956 – 984.

Karašáková, Z. (2013). Prvá priama voľba prezidenta v Českej republike na Twitteri. *Naše spoločnosť* 2 (11), 41 – 52.

Kenny, D. A. (2015). Citováno z webovej stránky Measuring Model Fit <http://davidakenny.net/cm/fit.htm>, 12. 5. 2015.

- Kenny D. A., D. A. Kashy. (1992). Analysis of the Multitrait – Multimethod Matrix by Confirmatory Factor Analysis. *Psychological Bulletin* 112 (1), 165 – 172.
- Kenny, D. A., D. T. Campbell. (1989). On the Measurement of Stability in Over-time Data. *Journal of Personality* 57 (2), 445 – 481.
- King, G., J. Pan, M. E. Roberts. (2016). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. *Harvard University*.
- Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1 (1).
- Kline, R. (2005). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.
- Kolenikov, S., K. A. Bollen. (2008.) Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification?. *Sociological Methods and Research* 41 (1), 124 – 167.
- Kosinski, M., D. Stillwell, T. Graepel. (2013). Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences*, 110 (15), 5802 – 5805.
- Knauper, B. (1999). The Impact of Age and Education on Response Order Effects in Attitude Measurement. *Public Opinion Quarterly* 63, 347 –370.
- Knowles, E. S., Condon, C. A. (1999). Why People Say “Yes”: A Dual – Process Theory of Acquiescence. *Journal of Personality and Social Psychology* 77, 379 –386.
- Kopecký, K., V. Krejčí. (2010). *Rizika virtuální komunikace. Příručka pro učitele*. Olomouc: Net.
- Kreidl, M. (2005). Zhodnocení vlivu práce výzkumných agentur na konstruktovou validitu škál. *Sociologický časopis / Czech Sociological Review* 41 (2), 103 –123.
- Krejčí, J. (2008). *Kvalita sociálněvědních výběrových šetření v České republice*. Praha: SLON.

- Kroh, Martin. (2006). An Experimental Evaluation of Popular Well-Being Measures.“ DIW-Diskussionspapiere, No. 546.
- Krosnick, J. A. (1991). ‘Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5, 213 – 236.
- Lazer, D., R. Kennedy, G. King, G., A. Vespignani. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343, 1203 –1205.
- Lord, F. M., M. R. Novick. (1968). *Statistical Theories of Mental Test Scores*. Reading MA: Addison –Welsley Publishing Company.
- Lucas, R. E., B. Donnellan. (2011). Personality Development Across the Life Span: Longitudinal Analyses with a National Sample from Germany.“ *Journal of Personality and Social Psychology* 101, 847 – 861.
- Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data. *Debates in the digital humanities*, 2, 460 – 475.
- Matějů, P. (1989). Metoda strukturního modelování: Přehled základních problémů. *Sociologický časopis/ Czech Sociological Review* 25 (4), 399 – 417.
- Maybury, K., L. Tatosky, A. Mallonee, A. DeCarlo. (2013). The Influence of the Number of ‘likes’ and Priming Information on Observers’ Evaluations of Social Media Status Updates. Online: <http://www.coplac.org/teagle/projects/DeCarloFinalPaper.pdf>
- McClendon, M. J. (1991). Acquiescence and Recency Response-Order Effects in Interview Surveys. *Sociological Methods and Research* 20, 60 – 103.
- McFarland, D. A., K. Lewis, A. Goldberg. (2016). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*, 47 (1), 12 – 35.
- Mills, C. W. (2000). *The Sociological Imagination*. Oxford University Press.
- Mirowsky, J., C. E. Ross. (1991). Eliminating Defense and Agreement Bias from Measures of the Sense of Control: A 2 x 2 Index. *Social Psychology Quarterly* 54 (2), 127 – 145.

- Meisenberg, G., A. Williams – Shillingford. (2008). Are Acquiescent and Extreme Response Styles Related to Low Intelligence and Education? *Personality and Individual Differences* 44, 1539 – 1550.
- Moors, G. (2008). Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement. *Quality & Quantity* 42, 779 – 794.
- Murphy, J., E. Dean, C. A. Hill, A. Richards. (2011). Social Media, New Technologies, and the Future of Health Survey Research. In *National Center for Health Statistics (Ed.), Proceedings of the 10th conference on health survey research methods* (pp. 231 –241).
- Murphy, Joe, Craig A. Hill, and Elizabeth Dean. (2013). Social Media, Sociality, and Survey Research. In *Social Media, Sociality, and Survey Research*, edited by Craig A. Hill, Elizabeth Dean, and Joe Murphy, 1 – 34. Hoboken, NJ: John Wiley and Sons.
- Murphy, J., M. W. Link, J. H. Childs, C. L. Tesfaye, E. Dean, M., Stern, P. Harwood. (2014). Social Media in Public Opinion Research Executive Summary of the Aapor Task Force on Emerging Technologies in Public Opinion Research. *Public Opinion Quarterly*, 78 (4), 788 –794.
- Mysíková, M. (2011). EU-SILC a jeho metodologická úskalí: mezinárodní srovnatelnost a příjmové proměnné. *Data a výzkum - SDA Info* 5 (2), 147 – 170.
- Neal, C. (2015). Starbucks' Marketing Power Comes from its Commitment to Social Media. Online: <http://www.brittonmdg.com/the-britton-blog/starbucks-commitment-to-social-media-is-their-marketing-power>
- Novick, M. R. (1965). The Axioms and Principal Results of Classical Test Theory. *Research Bulletin* 2.
- Oberski, D., T. Gruner, W. Saris. (2011). The Program SQP 2.0 for Prediction of Quality of Questions and Its Applications. *RECSM Working paper* 24, chapter 7.
- Olson, K., I. Bilgen. (2011). The Role of Interviewer Experience on Acquiescence. *Public Opinion Quarterly* 75, 99 – 114.
- Ones, D., S. Ch. Viswesvaran, A. D. Reiss. (1996). Role of Social Desirability in Personality Testing for Personnel Selection: The Red Herring. *Journal of Applied Psychology* 81 (6), 660 – 79.

Peter, J. P. (1979). Reliability: A Review of Psychometric Basics and Recent Marketing Practices. *Journal of Marketing Research* 16, 6 – 17.

Pew Research Center. (2012). Assessing the Representativeness of Public Opinion Surveys. Online: <http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/>

Pospíšilová, M. (2016). *Facebooková (ne)závislost. Identita, interakce a uživatelská kariéra na Facebooku*. Karolinum.

Revilla, M. A., W. E. Saris, J. A. Krosnick. (2014). Choosing the Number of Categories in Agree – Disagree Scales. *Sociological Methods & Research* 43 (1), 73 – 97.

Revilla, M., W. E. Saris. (2013a). The Split-Ballot Multitrait-Multimethod Approach: Implementation and Problems. *Structural Equation Modeling* 20, 27 – 46.

Revilla, M. A., W. E. Saris.(2013b). A Comparison of the Quality of Questions in a Face-to-face and a Web Survey. *International Journal of Public Opinion Research* 25 (2), 242 – 253.

Revilla, M., W. E. Saris. (2011a). The Split – Ballot MTMM Approach: Implementation and Problems. *RECSM Working paper* 19.

Revilla, M., W. E. Saris.(2011b). Estimation Problems and Solutions.“ *RECSM Working paper* 24.

Ruppert, E., J. Law, M. Savage. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, culture & society*, 30 (4), 22 – 46.

Řehák, J. (1971). Definice měření ve společenských vědách.“ *Sociologický časopis* 7 (6), 638 – 647.

Řehák, J. (1998a). Kvalita dat I. Klasický model měření reliability a jeho praktický aplikační význam. *Sociologický časopis* 34 (1), 51 – 60.

Řehák, J. (1998b). Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur.“ *Sociologický časopis* 34 (2), 195 –204.

Řehák, J., I. Bártová, J. Hamanová. (1998). Kvalita dat III. Empirické výsledky měření reliability pro vybrané míry a stupnice. *Sociologický časopis* 34 (3), 363 – 372.

- Říčan, Pavel. (1980). Úvod do Psychometrie. Psychodiagnostické a didaktické testy, n. p. Bratislava
- Rorer, L. G. (1965). The Great Response – Style Myth. *Psychological Bulletin* 63, 129 – 156.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48 (2), 1 – 36. URL <http://www.jstatsoft.org/v48/i02/>.
- Ross, C. J. Mirowsky. (1984). Socially Desirable Response and Acquiescence in a Cross – Cultural Survey of Mental Health. *Journal of Health and Social Behavior* 25, 189 –197.
- RStudio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- Saris, W. E., I. Gallhofer. (2014). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Wiley New Jersey.
- Saris, W. E., D. Oberski, M. Revilla, D. Zavala, L. Lilleoja, I. Gallhofer, T. Gruner. (2011). The Development of the Program SQP 2.0 for the Prediction of the Quality of Survey Questions. *RECSM Working paper 24*.
- Saris, W. E. (2009). The MTMM Approach to Coping with Measurement Errors in Survey Research. *RECSM Working Paper 2*.
- Saris, W. E., I. Gallhofer. (2007). Estimation of the Effects of Measurement Characteristics on the Quality of Survey Questions. *Survey Research Methods* 1 (1), 29 – 43.
- Saris, W. E., A. Sattora, G. Coenders. (2004). A New Approach to Evaluating the Quality of Measurement Instruments: the Split – Ballot MTMM design. *Sociological Methodology* 34 (1), 311 – 347.
- Saris, W. E., I.Gallhofer. (2003). Report on the MTMM Experiments in the Pilot Studies and Proposals for Round 1 of the ESS.
- Saris, W. E., F. M. Andrews. (1991). Evaluation of Measurement Instruments Using a Structural Modeling Approach. Pp. 575 – 597 in P. P. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz, S. Sudman (Eds.). *Measurement errors in surveys*. New York: Wiley.

- Savage, M., R. Burrows. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, 41(5), 885 – 899.
- Savage, M., R. Burrows. (2009). Some Further Reflections on the Coming Crisis of Empirical Sociology. *Sociology*, 43 (4) 762 – 772.
- Schilling, O. (2006). Development of Life Satisfaction in Old Age: Another View on the Paradox. *Social Indicators Research* 75, 241 – 271.
- Scholz, E., J. Harkness, T. Faab. (2005). ISSP Study Monitoring Report to the ISSP General Assembly on Monitoring Work Undertaken for the ISSP by ZUMA, Germany.
- Schubert, J. (2010). Klasická testová teorie reliability v metodologii výběrových šetření. *Data a výzkum - SDA Info*. 4 (2), 105 – 122.
- Schräpler, J. P., J. Schupp, G. G. Wagner. (2006). Changing from PAPI to CAPI: A Longitudinal Study of Mode – Effects Based on an Experimental Design. *Discussion Papers*. Berlin: German Institute for Economic Research.
- Schwartz, N., F. Strack, H. J. Hippler, G. Bishop. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology* 5, 193 – 212.
- Snijders, C., U. Matzat, U. D. Reips. (2012). Big Data: Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*, 7 (1), 1 – 5.
- Soukalová, D. (2016). Česka sháněla bydlení pro Afghánce. Hrozili jí za to znásilněním i smrtí. Blesk.cz, zveřejněno 9. 12. 2016. Online: http://www.blesk.cz/clanek/zpravy-udalosti/436030/ceska-shanela-bydleni-pro-afghance-hrozili-ji-za-to-znasilnenim-i-smrti.html?utm_source=blesk.cz&utm_medium=copy
- Štětka, V., J. Mazák. (2014). Whither Slacktivism? Political Engagement and Social Media Use in the 2013 Czech Parliamentary Elections.“ *Cyberpsychology: Journal of Psychosocial Research on Cyberspace (online)* 8 (3).
- Tabachnick, B. G., L. S. Fidell, S. J. Osterlind. (2001). *Using multivariate statistics*.
- Thrift, N. (2005). *Knowing capitalism*. Sage.

- Tinati, R., S. Halford, L. Carr, C. Pope. (2014). Big data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology* 48 (4), 663 – 681.
- Tourangeau, R., L. J. Rips, K. Rasinski. (2000). *The Psychology of Survey Response*. Cambridge University Press.
- Traub, R., E. (1997). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice* 16 (4), 8 – 14.
- Trusínová, R. (2014). No Two Ageisms Are the Same: Testing Measurement Invariance in Ageism Experience Across Europe. *International Journal of Social Research Methodology* 17 (6), 659 – 675.
- Van der Veld, W. M., W. E. Saris, A. Satorra. (2008). Judgment Aid Rule Software.
- van der Veld, W. M., W. E. Saris. (2005). The Nature of Measurement Error in Panel Data. Estimating Opinion Stability in Panel Surveys. Paper prepared for the first meeting of the European Association for Survey Research, Barcelona, 2005, July 18 – 21.
- Van Herk, H., Y. H. Poortinga, T. M. M. Verhallen. (2004). Response Styles in Rating Scales: Evidence of Method Bias in Data from Six EU Countries. *Journal of Cross-Cultural Psychology* 35, 346 – 360.
- van Meurs, A., W. E. Saris. (1990). Memory Effects in MTMM Studies. Pp. 134–46 in Saris, W. E., van Meurs. *An Evaluation of Measurement Instruments by Meta-analysis of Multitrait-Multimethod Studies*. Amsterdam: North Holland.
- Van Vaerenbergh, Y., T. D. Thomas. (2013). Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies”. *International Journal of Public Opinion Research* 25 (3), 195 –217.
- Vinopal, J. (2008). *Kognitivní přístupy v metodologii výzkumných šetření – metoda okamžité validizace*. Sociologický ústav AV ČR. Praha.
- Vochocová, L., J. Mazák, V. Štětka. (2016). Nic pro holky? Genderové nerovnosti v politické participaci na sociálních sítích. *Gender, rovné příležitosti, výzkum* 17 (2), 64 – 75.

- Watson, D. (1992). Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness. *Sociological Methods & Research* 21 (5), 2 – 88.
- Webber, R. (2009). Response to The Coming Crisis of Empirical Sociology: An Outline of the Research Potential of Administrative and Transactional Data. *Sociology*, 43 (1), 169 – 178.
- Weijters, B., M. Geuens, N. Schillewaert. (2010a). The Individual Consistency of Acquiescence and Extreme Response Style in Self – Report Questionnaires. *Applied Psychological Measurement* 34, 105 – 121.
- Weijters, B., Geuens, M., N. Schillewaert. (2010b). The Stability of Individual Response Styles. *Psychological Methods* 15, 96 – 110.
- Welkenhuysen-Gybels, J., J. B. Billiet, B. Cambré. (2003). Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-type Score Items. *Journal of Cross-Cultural Psychology* 34 (6), 702 – 722.
- Werts, C. E., R. L. Linn. (1970). Path Analysis: Psychological examples. *Psychological Bulletin* 74, 194 – 212.
- Wetzel, E., O. Lüdtke, I. Zettler, J. R. Böhnke. (2016). The Stability of Extreme Response Style and Acquiescence Over 8 Years. *Assessment* 23 (3), 279 – 91.
- Wiley, D., J. Wiley. (1970). The Estimation of Measurement Error in Panel Data. *American Sociological Review* 35 (1), 112–117.

11. Přílohy

Příloha 4.1 Syntaxe programu LISREL použití k analýze QS modelů podle 1) Heiseho (1969) a 2) Wileyeho a Wileyeho (1970).

1) Syntax LISREL Heise (1969) QSM model

model ny=4 ne=4 ps=sy,fi be=fu, fi ly=fu,fi te= sy,fi

va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4

fr be 2 1 be 3 2 be 4 3

fr ly 1 1

eq ly 1 1 ly 2 2 ly 3 3 ly 4 4

fr te 1 1 te 2 2 te 3 3 te 4 4

2) Syntax LISREL Wiley Wiley (1970) QSM model:

model ny=4 ne=4 ps=sy,fi be=fu, fi ly=fu,fi te= sy,fi

fr ps 1 1 ps 2 2 ps 3 3 ps 4 4

fr be 2 1 be 3 2 be 4 3

va 1 ly 1 1 ly 2 2 ly 3 3 ly 4 4

fr te 1 1

eq te 1 1 te 2 2 te 3 3 te 4 4

Příloha 4.2 Stability β v kvazisimplexových modelech pro položky *PH010 Zdraví*, *HS120*

Příjem a HS130 Odhad příjmu:

| Položka PH010 | Mód | $\beta (\eta_1;\eta_2)$ | $\beta (\eta_2;\eta_3)$ | $\beta (\eta_3;\eta_4)$ |
|------------------|------|-------------------------|-------------------------|-------------------------|
| Řada 2005 - 2008 | PAPI | 0,92 | 0,94 | 0,96 |
| Řada 2006- 2009 | PAPI | 0,94 | 0,96 | 0,96 |
| Řada 2007 - 2010 | PAPI | 0,94 | 0,93 | 0,94 |
| Řada 2008 - 2011 | CAPI | 0,93 | 0,96 | 0,95 |
| Řada 2009 - 2012 | PAPI | 0,92 | 0,96 | 0,94 |
| Řada 2009 - 2012 | CAPI | 0,94 | 0,94 | 0,99 |
| Položka HS120 | Mód | $\beta (\eta_1;\eta_2)$ | $\beta (\eta_2;\eta_3)$ | $\beta (\eta_3;\eta_4)$ |
| Řada 2005 - 2008 | PAPI | 0,87 | 0,9 | 0,91 |
| Řada 2006- 2009 | PAPI | 0,9 | 0,9 | 0,9 |
| Řada 2007 - 2010 | PAPI | 0,89 | 0,91 | 0,93 |
| Řada 2008 - 2011 | CAPI | 0,84 | 0,88 | 0,93 |
| Řada 2009 - 2012 | PAPI | 0,89 | 0,93 | 0,94 |
| Řada 2009 - 2012 | CAPI | 0,91 | 0,98 | 0,94 |
| Položka HS130 | Mód | $\beta (\eta_1;\eta_2)$ | $\beta (\eta_2;\eta_3)$ | $\beta (\eta_3;\eta_4)$ |
| Řada 2005 - 2008 | PAPI | 0,9 | 0,94 | 0,95 |
| Řada 2006- 2009 | PAPI | 0,92 | 0,95 | 0,97 |
| Řada 2007 - 2010 | PAPI | 0,91 | 0,95 | 0,96 |
| Řada 2008 - 2011 | CAPI | 0,88 | 0,94 | 0,98 |
| Řada 2009 - 2012 | PAPI | 0,9 | 0,95 | 0,95 |
| Řada 2009 - 2012 | CAPI | 0,93 | 0,94 | 0,99 |

Příloha 5.1 Syntaxe Modelu CF RS; program lavaan v prostředí R Studio

```
ModelCFRS <- '
```

```
Baterie =~ v50 + v51 + v52 + v53 + v54 + v61 + v62
```

```
Style =~ 1*v50 + 1*v51 + 1*v52 + 1*v53 + 1*v54 + 1*v61 + 1*v62
```

```
Baterie ~~ 0*Style
```

```
,
```

```
fit <- cfa (ModelCFRS, data=CZ_ISSP2003_rec, estimator = "MLM")
```

```
summary (fit, fit.measures=TRUE, standardized=TRUE)
```

Příloha 5.2 Syntaxe Modelu CF RS s korelací s indexem souhlasu; program lavaan
v prostředí R Studio

```
ModelCFRS_sumagree <- '
```

```
Baterie =~ v50 + v51 + v52 + v53 + v54 + v61 + v62
```

```
Style =~ 1*v50 + 1*v51 + 1*v52 + 1*v53 + 1*v54 + 1*v61 + 1*v62
```

```
Sumagree =~ v50_v62_ABSsumagree
```

```
Baterie ~~ 0*Style
```

```
Baterie ~~ Sumagree
```

```
'
```

```
fit <- cfa (ModelCFRS_sumagree, data=CZ_ISSP2003_rec, estimator = "MLM")
```

```
summary (fit, fit.measures=TRUE, standardized=TRUE)
```

Příloha 6.1 Syntax programu LISREL k odhadu parametrů TS MTMM modelu 2 SB
MTMM experimentu Role mužů a žen ve společnosti z ESS2.

Analyza ESS 2 data experiment gender group 1

Data ng=2 ni=9 no=1033 ma=cm

Km

*

1.00

.046 1.00

.424 -.102 1.00

.103 -.116 .079 1.00

-.263 .315 -.362 -.067 1.00

.109 -.164 .293 .312 -.117 1.00

.000 .000 .000 .000 .000 .000 1.00

.000 .000 .000 .000 .000 .000 .000 1.00

.000 .000 .000 .000 .000 .000 .000 .000 1.00

Me

*

2.53 1.97 3.12 2.65 2.91 3.00 0.00 0.00 0.00

sd

*

1.08 .91 1.25 1.10 1.19 1.26 1.00 1.00 1.00

model ny=9 ne=9 nk=6 te=di,fr ly=fu,fi ps=di,fi be=fu,fi ga=fu,fi ph=sy,fi

value 1 ly 1 1 ly 2 2 ly 3 3 ly 4 4 ly 5 5 ly 6 6

value 1 te 7 7 te 8 8 te 9 9

value 0 ly 7 7 ly 8 8 ly 9 9

free ga 1 1 ga 4 1 ga 7 1 ga 2 2 ga 5 2 ga 8 2 ga 3 3 ga 6 3 ga 9 3

value 1 ga 1 4 ga 2 4 ga 3 4 ga 4 5 ga 5 5 ga 6 5 ga 7 6 ga 8 6 ga 9 6

free ph 1 2 ph 2 3 ph 1 3 ph 4 4 ph 5 5 ph 6 6

value 1 ph 1 1 ph 2 2 ph 3 3

start .5 all

value .10 ph 5 5 ph 6 6

out iter=200 adm=off sc ec

Analyza ESS 2 data experiment gender group 2

Data ni=9 no=827 ma=cm

Km

*

1.00

.041 1.00

.390 -.087 1.00

.000 .000 .000 1.00

.000 .000 .000 .000 1.00

.000 .000 .000 .000 .000 1.00

.588 -.067 .325 .000 .000 .000 1.00

-.061 .442 -.148 .000 .000 .000 -.060 1.00

.322 -.133 .670 .000 .000 .000 .402 -.151 1.00

me

*

2.57 1.84 3.00 .00 .00 .00 2.70 2.20 3.14

sd

*

1.13 .89 1.30 1.00 1.00 1.00 1.22 1.15 1.33

model ny=9 ne=9 nk=6 te=di,fr ly=fu,fi ps=in be=in ga=in ph=in

value 1 ly 1 1 ly 2 2 ly 3 3 ly 7 7 ly 8 8 ly 9 9

value 1 te 4 4 te 5 5 te 6 6

value 0 ly 4 4 ly 5 5 ly 6 6

out iter= 200 adm=off sc ec

Příloha 8. 1

Pokusní králíci (Facebooková stránka). Datum zveřejnění: 16. ledna 2017. Text statusu:

„#vitejte Víťáme více než 250 nových Pokusných králíků, kteří se k nám přidali po zhlédnutí pořadu Václava Moravce, ve kterém o pravdě a nepravdě hovořil i náš prof.

Jaroslav Flegr!“

12. Použitá data

Kapitola 4:

EUSILC LONGITUDINAL UDB 2008 – version-3 of August 2011

EUSILC LONGITUDINAL UDB 2009 – version-4 of March 2013

EUSILC LONGITUDINAL UDB 2010 – version-5 of August 2014

EUSILC LONGITUDINAL UDB 2011 – version-1 of August 2013

EUSILC LONGITUDINAL UDB 2012 – version-1 of August 2014

Kapitola 5:

Sociologický ústav (Akademie věd ČR). Národní identita (ISSP 2003) - Česká republika [datový soubor] [online]. Ver. 1.0. Praha: Český sociálněvědní datový archiv, 2007 [citováno 9. 9. 2016]. DOI 10.14473/ISSP00012

Sociologický ústav (Akademie věd ČR). Národní identita (ISSP 2013) - Česká republika [datový soubor] [online]. Ver. 1.0. Praha: Český sociálněvědní datový archiv, 2013 [citováno 9. 9. 2016]. DOI 10.14473/ISSP00021

Kapitola 6:

ESS Round 1: European Social Survey Round 1 Data (2002). Data file edition 6.3 NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3.3 NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

13. Seznam zkratek:

(uvedeny jsou pouze zkratky, které se v textu vyskytují opakovaně):

2 SB MTMM... 2 split – ballot Multitrait Multimethod; design sběru dat

ARS... Acquiescence Response Style; souhlasný styl odpovídání

CAPI... Computer Assisted Personal Interview

CF-RS model ... Content Factor – Response style model

CFA... Confirmatory Factor Analysis; konfirmativní faktorová analýza

CFI ... Comparative Fit Index; ukazatel shody modelu s daty

CMV... Common Method Variance; společný rozptyl metody

df ... degree of freedom; stupně volnosti v modelu

DRS... Disagreeing Response Style; nesouhlasný styl odpovídání

ERS ... Extreme Response Style; extrémní styl odpovídání

ESS... European Social Survey

EU SILC... European Union Statistics on Income and Living Conditions; šetření Životní podmínky

HC... Heywood cases; Heywoodovy případy

ISSP... International Social Survey Programme

KTT... Klasická testová teorie

ML... Maximum Likelihood

MLM... Satorra – Bentler Mean Adjusted Maximum Likelihood

MRS... Middle Response Style; střední styl odpovídání

MTMM ... Multitrait Multimethod

MG ML... Multigroup Maximum Likelihood

MG SEM... Multigroup Structural Equation Modeling

PAPI ... Paper and Pencil Interview

QSM... Quasi Simplex Model, kvazisimplexový model

RECSM... Research and Expertise Centre for Survey Methodology

RIRS... Representative Indicators of Response Style

RMSEA... Root Mean Square Error of Approximation; ukazatel shody modelu s daty

RS... Response Style; styl odpovídání

SAQ... Self – administered Questionnaire; dotazník vyplňovaný respondentem

SQP... Survey Quality Prediction

SEM... Structural equation modeling, modelování strukturálních rovnic

SRMR ... Standardized Root Mean Square Residual; ukazatel shody modelu s daty

TS MTMM model... True Score Multitrait Multimethod model