Měření validity a reliability otázek v šetření European Social Survey a jeho využití*

Johana Chylíková**
Sociologický ústav AV ČR

Measurement and use of validity and reliability estimates of items in the European Social Survey

Abstract: *The aim of this article is to present a trend in research on measurement error in survey data and to suggest some problematic aspects of this approach. The article describes the Multitrait Multimethod experimental design and its modification into a 2 Split-ballot Multitrait Multimethod (2 SB MTMM), which is used for experimental data collection in the European Social Survey. The text shows how to analyze 2 SB MTMM data to obtain estimates of construct validity, reliability and common method variance for a single questionnaire item, and how to make use of these estimates. It also points to some problems encountered in 2 SB MTMM data analysis.*

Keywords: *survey, methodology, validity, reliability, data analysis, 2 SB MTMM*

Data a výzkum – SDA Info 2015, Vol. 9, No. 1: 5–29.

DOI: <http://dx.doi.org/10.13060/23362391.2015.9.1.234>

© Sociologický ústav AV ČR, v.v.i., Praha 2015

Úvod

V kvantitativním dotazníkovém výzkumu jsou získávány hodnoty proměnných, které analytici a analytičky považují za sice redukované, ale více méně objektivní a přesné ukazatele dané operacionalizace zkoumaného jevu. Pokud však klademe důraz na přesnost a objektivitu v kvantitativních datech, musíme zároveň mluvit i o šumu, odchylkách nebo chybách, které jsou v takových

* Práce na tomto článku byla financována z grantu Projekt CESSDA: vytvoření českého uzlu CESSDA a jeho zapojení do pan-evropské velké výzkumné infrastruktury datových služeb pro socioekonomický výzkum, sponzorovaného Ministerstvem školství, mládeže a tělovýchovy ČR s registračním číslem LM2010006.

** Korespondenci týkající se tohoto textu pošlejte na e-mail johana.chylikova@soc.cas.cz nebo na adresu Český sociálněvědní datový archiv, Sociologický ústav AV ČR, v.v.i., Jilská 1, Praha 1, 110 00.

datech obsaženy. Například víme, že v datech z výběrového šetření nejsou informace za všechny jednotky v populaci, ale pouze z výběrového souboru, a proto pracujeme s tzv. výběrovou chybou, jejíž princip stojí na inferenční statistické teorii, umožňující zobecnění výsledků analýzy na celou populaci. Výběrová chyba patří mezi tzv. chyby chybějícího pozorování [Groves 2004] společně s dalšími chybami, jako je chyba reprezentativity nebo chyba vyplývající z výpadku návratnosti. Tyto chyby jsou důsledkem toho, že ve výzkumu je šetřen pouze výběrový soubor, nikoliv celá populace, a že některé jednotky, které by ve výběrovém souboru být měly, nejsou zahrnuty. Dalším typem chyb jsou tzv. chyby měření, tedy chyby vyplývající z pozorování, které se nacházejí přímo v datech z dotazování vzorku populace. Chyby měření jsou ty složky rozptylu měřených proměnných, které mají náhodný či systematický charakter a nejsou podle tzv. *klasické testové teorie* [Traub 1997] reflexí existence latentní proměnné odpovídající měřené proměnné. Právě chyby měření jsou ústředním tématem tohoto článku.

Klasická testová teorie předpokládá rozdělení rozptylu měřené proměnné na část odpovídající měřenému konstruktovi a na část odpovídající chybě měření, přičemž chybová část rozptylu má negativní vliv na výsledky analýz. V analýze, kde není uvažována chyba měření, reprezentují v biviační nebo multivariační analýze odhady lineárních vztahů vztahy mezi měřenými proměnnými. V analýze s daty, ze kterých byla nevalidní část rozptylu odstraněna, však získáváme odhady lineárních vztahů mezi latentními proměnnými, které reprezentují zkoumaný konstrukt. Velikost a směr lineárních vztahů ve dvou naznačených typech analýzy se mohou lišit a podle řady autorů produkuje analýza dat bez chybových rozptylů výsledky, které lépe zachycují vztahy mezi zkoumanými fenomény. Klasická testová teorie zažila rozkvět v první polovině 20. století, kdy vznikl koncept náhodné chyby a některé metody pro odhad reliability [Traub 1997]. Na konci padesátých let se objevila idea důležitosti pro vznik experimentálního designu multitrait multimethod (MTMM) [Campbell, Fiske 1959], metody k identifikaci validního a nevalidního rozptylu měřené proměnné, která je stěžejní pro výklad v tomto článku. Zhruba od osmdesátých let se zájem o chyby měření začal rozšiřovat do oboru metodologie sociálních věd a stal se součástí přemýšlení o tzv. kvalitě dat z kvantitativních šetření. O validitě a reliabilitě dat se však v sociálních vědách ani dnes neuvažuje úplně běžně; analytici a analytičky s klasickou teorií měření nepracují a celý rozptyl měřených proměnných považují za relevantní data vstupující do analýzy. V běžné analýze bývají zpravidla zohledňovány jen některé chyby chybějícího pozorování, jako je výběrová chyba, chyba vyplývající z výpadku návratnosti (missing values analysis) nebo chyba reprezentativity (vážení). Chyby měření v analýze kvantitativních sociálněvědních dat nejsou zohledňovány téměř vůbec nebo se o nich uvažuje nesprávně. Např. někdy mezi výzkumníky dochází k mylnému chápání reliability, když za náhodnou chybu měření považují zbytekový rozptyl měřené proměnné v modelu lineárních vztahů mezi proměnnými. V tomto případě se však jedná o rozptyl, který není vysvětlen použitým mode-

lem vztahů mezi proměnnými, nikoliv o náhodnou chybu ve smyslu klasické testové teorie.

Analytickou praxi, která nepracuje s koncepty chyby měření, se snaží změnit výzkumný tým Willema Sarise z Research and Expertise Centre for Survey Methodology (RECSM) ze španělské Universitat Pompeu Fabra v Barceloně. Tým RECSM vyvinul software nazvaný Survey Quality Prediction 2.0 (SQP 2.0), který umožňuje badatelům využívajícím data z kvantitativních šetření získat informace o validitě a reliabilitě libovolné dotazníkové otázky měřící libovolnou proměnnou. Algoritmus programu SQP 2.0 čerpá ze stovek studií MTMM, realizovaných především v projektu mezinárodních dotazníkových šetření European Social Survey¹ (ESS). SQP 2.0 je bezplatně dostupný on-line; byl vytvořen se záměrem poskytnout uživatelům možnost testovat kvalitu dotazníkové položky ještě před tím, než tato otázka bude použita v dotazníkovém šetření. Informace z SQP 2.0 mohou analytici použít buď jako podnět ke zlepšení otázky a zvýšení kvality v následujícím empirickém výzkumu, nebo jako koeficienty validity a reliability, které užijí přímo v analýze sebraných dat k očištění dat od chybových rozptylů.

V tomto textu popíšu funkce programu SQP 2.0 a naznačím, jak ho používat. Především se však zaměřím na popis metody odhadu validity a reliability, která byla použita k získání údajů, na jejichž základě SQP 2.0 pracuje. Nejdříve na dvou konkrétních případech ukážu, do jaké míry chyby měření ovlivňují lineární vztahy mezi proměnnými. Dále čtenáře seznámím s experimentálním designem MTMM, výchozí metodou pro stanovení validity a reliability, a s úpravou tohoto designu v tzv. 2 split ballot MTMM (2 SB MTMM), který je používán ke sběru dat v European Social Survey a ze kterého pocházejí odhady validity a reliability používané v programu SQP 2.0. Dále popíšu tzv. true score MTMM model, který tým RECSM používá pro odhad koeficientů validity a reliability z 2 SB MTMM dat. Závěr textu je věnován diskuzi problematických aspektů metody 2 SB MTMM a analýzy 2 SB MTMM dat. Tento text pokrývá mnoho témat, je poměrně členitý a v některých případech i méně přístupný v důsledku complexity popisovaných problémů. Proto prosím čtenářky a čtenáře, aby se nenechali odradit šíří témat, jež článek nabízí, složitostí některých částí textu a leckdy méně zjevnou návazností jednotlivých oddílů. Tento text je pečlivě logicky strukturován, byl napsán s důrazem na sdělnost a měl by čtenářům přinést ucelený přehled problematiky, jež je v současné době trendem v metodologickém kvantitativním výzkumu.

Chyby měření a jejich vliv na lineární vztahy mezi proměnnými

Klasická testová teorie za chybu měření považuje tu část rozptylu měřené proměnné, která neodpovídá měřené vlastnosti. Teorie předpokládá, že rozptyl měřené proměnné je vždy složen ze dvou částí; z části rozptylu odpovídající

1 www.europeansocialsurvey.org

Tabulka 1. Vliv složky náhodné chyby na korelace s jinými proměnnými

Reliabilita Y	Regresní koeficient						
	$X_1 \rightarrow Y$	$X_2 \rightarrow Y$	$X_3 \rightarrow Y$	$X_4 \rightarrow Y$	$X_5 \rightarrow Y$	$X_6 \rightarrow Y$	$X_7 \rightarrow Y$
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

Zdroj: [Bollen 1989]

měřené proměnné (substanciální rozptyl), pro který používá termín *pravý skór* (z angl. true score), a z náhodné chyby [Lord, Novick 1968]. Později začlenili další badatelé do teorie i systematickou složku, která je přisuzována vlivu použité metody měření [Andrews 1984]. Vliv systematické chyby je v teorii měření důvodem snížené tzv. konstruktové validity, náhodná chyba je pak důvodem snížené reliability. Systematická chyba ovlivňuje střední hodnotu měřené proměnné, ale i velikost lineárních vztahů mezi proměnnými; náhodná chyba pak ovlivňuje velikost lineárních vztahů mezi proměnnými [Andrews 1984]. Pokud jsou z naměřené proměnné eliminovány chybové rozptyly, zůstává substanciální rozptyl, přičemž analýza takto očištěných dat může přinášet jiné výsledky než analýza neočištěných dat.

V české odborné literatuře informoval o vlivu očištění dat o náhodnou chybu např. Řehák [1971, 1998a], v zahraniční literatuře byly prezentovány konkrétní příklady vlivu náhodné chyby na biviační analýzu např. v publikaci Bollena [1989]. Ten ve svém příkladu ukazuje, jaký vliv má snížená reliabilita na velikost regresních koeficientů v modelu s jednou závislou a sedmi nezávislými proměnnými. Závislou proměnnou Y je statistika úmrtnosti v různých lokalitách, nezávislými proměnnými X jsou různé faktory kvality ovzduší a kontrolní proměnné jako hustota zalidnění, stáří populace atd. Tabulka 1 ukazuje, jak snížená reliabilita závislé proměnné Y ovlivňuje pozorované vztahy mezi závislou a nezávislými proměnnými. Sníženou reliabilitu má v tomto případě pouze Y, o nezávislé proměnných X se předpokládá, že jsou bez chyby měření. V řádcích tabulky 1 je patrné, jak náhodná chyba měření mění velikost vztahů mezi proměnnými. Zatímco při reliabilitě rovné jedné, tedy v případě, kdy v rozptylu proměnné neexistuje náhodná chyba, je korelace např. mezi Y a X_1 rovna 0,107, při reliabilitě 0,5 je to 0,291. V tomto případě má náhodná chyba ten vliv, že pozorovanou korelaci mezi proměnnými zvyšuje. V případě lineárního vztahu mezi $X_3 \rightarrow Y$ nastává jev opačný, korelace za předpokladu ideální reliability má hodnotu 0,064, při reliabilitě 0,5 už to je jen 0,017.

Saris a Galhofferová [2014] ukazují ještě markantnější změny v nalezených lineárních vztazích mezi proměnnými na příkladu pěšinkové analýzy (path

Obrázek 1. Kauzální model vztahů mezi proměnnými bez očištění dat o chybu měření

Zdroj: [Saris a Galhofferová 2014: 288–291]

Obrázek 2. Kauzální model vztahů mezi proměnnými s očištěním dat o náhodnou a systematickou chybu měření

Zdroj: [Saris a Galhofferová 2014: 288–291]

analysis) se čtyřmi proměnnými. Změny jsou v tomto případě patrnější, neboť autoři počítají s náhodnou i systematickou chybou, a to u všech proměnných v analýze. Pro očištění dat z irského šetření ESS 3 použili odhady ukazatele kvality z experimentu 2 SB MTMM. Rozdíly ve velikosti faktorových zátěží mezi proměnnými bez korekce o chybu měření a s korekcí jsou znázorněny na obrázcích 1 a 2.

Z obrázků je patrné, že vztahy mezi proměnnými se po implementaci očištění rozptýlily výrazně změnil. 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 $0,938$.

Uvedené příklady ukazují, že zahrnutí chyby měření do analýzy proměnných může zásadně ovlivnit velikost a směr nalezených lineárních vztahů. Analýza s neočištěnými daty dochází k jiným výsledkům než analýza dat, ve které je uvažována chyba měření, přičemž tyto výsledky se od sebe mohou výrazně lišit. Domnívám se, že ve světle tohoto poznatku je vhodné uvažovat o alternativním přístupu k analýze dat, který zohledňuje chyby měření, a seznámit se s jeho principy.

Multitrait multimethod (MTMM) design a split ballot MTMM design

Pro odhad či určení chyb měření existuje několik metod, jejichž popis zprostředkovává odborná literatura [Groves 2004, Řehák 1998ab, Urbánek et al. 2011, Bollen 1989]. Jednou z nich je experimentální design multitrait multimethod, který umožňuje získat taková data, v nichž je za použití multivariační analýzy možné rozdělit rozptyl měřené proměnné na substantiální složku, odpovídající měřené proměnné, složku tvořenou náhodnou chybou a složku tvořenou efektem metody. Princip metody MTMM představili Campbell a Fiske [1959]; jejich tzv. MTMM korelační matice sloužila jako prostředek k určení konstruktové validity vybraných dotazníkových otázek. Z idey MTMM matice vzešel MTMM design sběru dat, který byl testován v osmdesátých letech [Andrews 1984, Saris, Andrews 1991]. Nejčastěji se o MTMM designu uvažuje jako o devíti proměnných, měřících celkem tři latentní znaky (v angličtině zvané traits) třemi různými metodami (methods); znaků, jejich indikátorů a metod ale může být i více. V publikacích týmu pracoviště RECSM je explicitně formulována podmínka, že latentní znaky v MTMM by mezi sebou měly korelovat, tzn. být podkonstrukty jednoho zastřešujícího konstruktů vyššího řádu [Saris, Satorra, Coenders 2004], jiní autoři tuto podmínku nevyžadují. Všech devět proměnných/otázek v MTMM experimentu musí být zodpovězeno každým respondentem ve výběrovém souboru. Získaná matice korelací, resp. kovariancí mezi devíti proměnnými pak slouží jako vstupní data pro získání odhadů validity, reliability a tzv. společného rozptylu metody (common method variance – CMV). Data jsou analyzována jako strukturální rovnice, čímž lze dosáhnout informace o proporcí rozptylu odpovídajícího měřenému konstrukt, rozptylu odpovídajícího systematickému vlivu použité metody a rozptylu náhodné chyby měření [Andrews 1984].

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 European Social Survey. Tento MTMM 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 od sebe v tomto případě lišily formulací výroků a škálou, na které byl výrok měřen. Dotazníkové položky měřící znaky metodami jsou vypsány v rámečku 1.

Rámeček 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.

Sběr dat pro klasický MTMM design je náročný pro respondenty, protože design předpokládá, že dotázaní budou opakovaně odpovídat na velice podobné otázky. Například na otázku v příkladu MTMM designu v rámečku 1, zda by žena měla omezit placenou práci kvůli rodině, která reprezentuje jeden ze znaků, musí jeden respondent 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 znaky tři, 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. Prvním problémem pojícím se s vysokou zátěží na respondenta je ztráta motivace dostatečně se soustředit na rozhovor; tento jev je v anglicky psané odborné literatuře nazýván *satisficing* [Tourangeau, Rips, Rasinski 2000] a dá se volně přeložit jako *vyhovění požadavku na odpověď*. V takovém případě dotazováním unavený respondent odpovídá nedbale a jeho odpovědi nemusejí 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].

Výše popsané dva problémy řeší přístup 2 split ballot MTMM (2 SB MTMM) [Revilla, Saris 2013], který snižuje respondentovu zátěž z celkových devíti odpovědí v jednom experimentu na šest. Přístup je inspirován metodou *split ballot*, která se v metodologii sociálněvědních šetření používá desítky let ke zjišťování variability subjektivních, ale někdy i objektivních proměnných². Split ballot design je řešen tak, že je výběrový soubor rozdělen na minimálně dva podsoubory, do kterých jsou respondenti z původního souboru vybráni náhodně [Biemer 2004]. Tento design umožňuje testovat např. různé formulace otázek, efekt kontextu a další charakteristiky dotazníku a zjistit, jaký má změna různých variant vliv na distribuci odpovědí. V 2 SB MTMM experimentu je výběrový soubor rozdělen na dva podsoubory; díky této redukci každý respondent ve výběrovém souboru odpovídá pouze na šest otázek z jednoho MTMM experimentu a ne na devět, jak je tomu v klasickém MTMM. Redukce na šest otázek by podle týmu pracoviště RECSM [van Meurs, Saris 1990] měla výrazně snížit pravděpodobnost, že si respondent bude pamatovat své předchozí odpovědi na otázku zjišťující jeden ze znaků v MTMM. Zároveň má

2 Proměnné v sociálněvědních výběrových šetřeních mohou být buď „objektivní“, tedy takové, které lze objektivně určit, např. věk, nebo subjektivní, tj. postojové proměnné [Andrews 1984].

toto snížení zátěže zamezit efektu satisficing. Design 2 SB MTMM byl použit ve všech vlnách šetření ESS, kde první podsoubor obsahoval respondenty, kteří odpovíděli na MTMM otázky měřené metodou 1 a metodou 3, respondenti v druhém podsouboru odpovídali na otázky měřené metodou 2 a metodou 3.

Analýza 2 SB MTMM dat

Obecně se pro analýzu MTMM dat používá software pro modelování strukturálních rovnic (Structural Equation Modeling – SEM) a modely odpovídají modelům konfirmativní faktorové analýzy (confirmatory factor analysis – CFA) [Bollen 1989, Werts, Linn 1970, Kenny, Kashy 1992]. Prostřednictvím tohoto analytického přístupu je možné získat odhady faktorových zátěží mezi měřeními indikátory Y_{ij} a jejich latentními znaky F_i (viz obrázek 3) [Andrews 1984]. Konkrétní faktorové zátěže mezi měřeními a latentními proměnnými odpovídají koeficientu validity (faktorové zátěže mezi F_i a Y_{ij}) a systematickému vlivu metody (faktorové zátěže mezi M_j a Y_{ij}). V tomto modelu nelze přímo zjistit reliabilitu; zbytkový rozptyl měřených proměnných Y_{ij} reliabilitu naznačuje, neposkytuje však přímo její koeficient, ze kterého lze vypočítat reliabilitu.

Tým RECSM používá pro analýzu 2 SB MTMM dat model nazvaný true score MTMM model (dále jen TS MTMM model) [Saris, Andrews 1991, Saris 2009]. „True score“ v názvu značí tzv. pravý skór, čímž odkazuje k ústřednímu konceptu klasické testové teorie a naznačuje rozdíl oproti dříve navrženému modelu, prezentovaném v obrázku 3. TS MTMM model totiž oproti předchozímu řešení umožňuje získat přímo vyjádření koeficientu reliability, a to zavedením latentní proměnné T_{ij} reprezentující pravý skór. V TS MTMM

Obrázek 3. Multitrait multimethod (MTMM): znázornění vztahů mezi měřeními a latentními proměnnými [Andrews 1984]

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. TS MTMM model je výchozí model pro všechny analýzy týmu RECSM a jsou podle něj počítány všechny odhady reliability a validity z šetření ESS, které tvoří základ databáze programu SQP 2.0.

TS MTMM model lze ve formě rovnic vyjádřit jako [Saris, Sattora, Coenders 2004]:

$$[1] \quad Y_{ij} = r_{ij} T_{ij} + e_{ij}, \quad \text{kde } i = 1;3 \text{ a } j = 1; 3,$$

$$[2] \quad T_{ij} = v_{ij} F_i + m_{ij} M_j, \quad \text{kde } i = 1;3 \text{ a } j = 1; 3.$$

- Y_{ij} je měřená proměnná/indikátor. 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 4.

V analýze 2 SB MTMM dat v softwaru pro strukturální modelování, jako je např. LISREL, Mplus nebo AMOS, je na TS MTMM model nutná aplikace některých restrikcí [Saris, Sattora, Coenders 2004]. Nejpodstatnější z nich je fixace faktorové zátěže pro latentní proměnné metod tak, aby byly stejné pro všechny proměnné Y_{ij} , tj. že:

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

Tato restrikce vyjadřuje, že o každé metodě použité v MTMM designu se předpokládá, že působí stejně na všechny tři indikátory, které jsou touto meto-

Obrázek 4. Znáznornění true score MTMM modelu; zdroj: [Saris a Galhofferová 2014]

dou měřeny. Aplikace této restrikce navíc zvyšuje stupně volnosti modelu, což zvyšuje pravděpodobnost, že bude v SEM analýze získáno řádné konvergující řešení (z angl. *proper solution*). To je takové řešení, kdy je výsledkem analýzy výstup s informacemi o korespondenci modelu s daty (z angl. *model fit*) a s odhady strukturních parametrů bez tzv. Heywoodových případů (*Heywood cases* – HC)³.

Další restrikce, které je nutné aplikovat na TS MTMM model, 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 restrikce zvyšují stupně volnosti modelu a zlepšují šance na získání řádného konvergujícího řešení, všechny jsou zároveň teoreticky opodstatněné.

3 HC je nežádoucí výsledek SEM analýzy, kdy jsou v modelu některé rozptyly odhadnuty jako menší než nula nebo některé korelace větší než $+/-1$ [Kolenikov, Bollen 2008].

Odhady validity a reliability a analýza s očištěnými daty

Výsledky analýzy dat s použitím TS MTMM modelu přináší informace pro určení proporce rozptylu proměnné se substantiálním významem a proporce chybového rozptylu. Koefficient validity (v_{ij}) je faktorová zátěž mezi latentní proměnnou T_{ij} a odpovídajícím latentním znakem F_i . Validita (v_{ij}^2) je koefficient validity umocněný na druhou [Saris, Satorra, Coenders 2004] a je vyjádřením vlivu latentního znaku F_i na pravý skór T_{ij} , tedy na latentní proměnnou měřené položky y_{ij} .

Faktorová zátěž mezi latentní proměnnou T_{ij} a latentní proměnnou metody M_j vyjadřuje vliv metody na pravý skór. Druhá mocnina tohoto koeficientu je komplementární k validitě, neboť platí, že:

$$v_{ij}^2 = 1 - m_{ij}^2 \text{ [Saris, Gallhofer 2014: 201].}$$

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 společného rozptylu metody (CMV), který mají dvě a více proměnných, které byly měřeny stejnou metodou měření (např. stejnou škálou). CMV se vypočítá podle:

$$\text{CMV} = r_i m_i m_j r_k,$$

kde r_i a r_k jsou koeficienty reliability položky i a položky k , které byly měřeny metodou M_j [Saris, Gallhofer 2014: 290]⁴. 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. Právě použití CMV dokáže „umělé“ zvýšení korelace mezi dvěma proměnnými vzniklé vlivem použité metody odstranit.

Koefficient 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.

Kromě validity, reliability a CMV pracuje tým pracoviště RECSM se souhrnným ukazatelem nazvaným *total quality of a measure* (q_{ij}^2) [Saris, Gallhofer 2014: 294], pro který volím český jazykový ekvivalent *celková kvalita položky*. Ukazatel číselně vyjadřuje, jak velká část rozptylu měřené proměnné odpovídá výhradně měřenému konceptu. Celková kvalita položky q_{ij}^2 se vypočítá podle: $q_{ij}^2 = (r_{ij} \cdot v_{ij})^2$. Ukazatel q_{ij}^2 je ten substantiální rozptyl, který v příkladu v tomto článku výše „T1M1 Žena by měla být připravena omezit svou place-

4 Saris a Gallhoferová 2014: 290 uvádějí vzorec $\text{CMV} = r_i m_i m_j r_k$ 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.

nou práci kvůli rodině“ odpovídá latentní proměnné názoru respondentů na omezení práce žen kvůli rodině. Celková kvalita položky q_{ij}^2 je o všechny chyby očištěný rozptyl, který reflektuje pouze měřený koncept, a je to ten koeficient, s jehož využitím se dají data z výběrového šetření očistit o chybové náhodné a systematické rozptyly. S očištěnými daty lze provádět bivariační i multivariační analýzu v běžných statistických paketech, jako je SPSS, nebo v softwaru pro strukturální modelování, jako je LISREL, Mplus nebo AMOS. Taková data mohou bez problému vstoupit do analýz, jako je regresní analýza, pěšinková analýza nebo strukturální modelování (SEM).

Je logické, že očištění dat o chybové rozptyly nelze udělat pro jeden případ ve výběrovém souboru zvlášť; je možné očistit pouze korelace mezi měřenými proměnnými. Nezpůsobuje to však žádný problém, neboť statistická analýza lineárních vztahů si vystačí s údaji o korelacích, případně kovariancích. Proces očišťování dat od chybových rozptylů probíhá tak, že vybereme proměnné, které chceme mít v analýze, a vytvoříme korelační matici všech těchto proměnných. 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 [Oberski, Gruner, Saris 2011]. 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 [Saris, Gallhofer 2014: 290]. 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 (více viz Saris, Gallhofer [2014: 290]).

Vzorec pro přepočet korelací mezi proměnnými, podle kterého lze získat očištěná data stejně jako ze SEM softwaru, je [Saris, Gallhofer 2014: 290]:

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

kde r_{ik} je korelace položek i a k , q_i a q_k odpovídají odmocninám hodnot celkové kvality položek 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.

Survey Quality Prediction 2.0

Dlouhodobý sběr dat metodou 2 SB MTMM v mezinárodním projektu ESS a následné vytvoření on-line softwaru SQP 2.0 byly motivovány snahou o kultivaci mezinárodního komparativního výzkumu. Leckdy odlišná kvalita dat z šetření v jednotlivých zemích vede k mezinárodním srovnáním, která kromě substantiálních rozdílů v měřených proměnných nevědomě reflektují i různé velké chyby měření v datech [Harkness et al. 2002]. Cílem výzkumníků

a výzkumníci pracující na metodologických šetřeních ESS bylo získat více či méně jednoduchý nástroj, který analytikům umožní odhadnout, jak velkou chybou je zanesena konkrétní proměnná měřená v konkrétní zemi, a použít tento odhad pro očištění dat pro substanciální analýzu. Tak se zrodil nápad na vytvoření softwaru, jehož algoritmus je založen na koeficientech validity a reliability z MTMM měření a který je schopen predikovat validitu a reliabilitu jakékoliv dotazníkové otázky.

Povědomí o projektu SQP 2.0 dosud není příliš rozšířeno mezi analytiky pracujícími s daty z kvantitativních šetření, nicméně v oboru kvantitativní metodologie se jedná o ojedinělý a ambiciózní výzkum, který budí pozornost. Výzkumný tým pracující na vývoji SQP 2.0 byl v roce 2014 oceněn cenou Warren J. Mitofsky Innovators Award prestižní Americké asociace pro výzkum veřejného mínění (AAPOR) a jednotliví pracovníci a pracovnice týmu jsou držiteli různých profesních a studentských ocenění právě za svoji práci v projektu SQP 2.0. Samo pracoviště se snaží SQP 2.0 popularizovat mezi analytiky a analytičkami například projektem studijních návštěv RECSM hrazených z rozpočtu ESS, kde se návštěvníci dozvědí, co je SQP 2.0, jak se používá a jak lze očistit data od chyb měření.

Princip SQP 2.0 stojí na myšlence, že z koeficientů validity a reliability otázek z již proběhlých měření MTMM lze získat algoritmus, který dokáže odhadnout validitu a reliabilitu jiné dotazníkové otázky podle jejích charakteristik, jako je např. téma, na které se otázka ptá, použitá škála odpovědi, jazyk, ve kterém je otázka položena, náchylnost otázky k sociální desirabilitě nebo množství slov či slabik v otázce⁵. Databáze odhadů validity a reliability, ze které algoritmus SQP 2.0 čerpá, aktuálně obsahuje údaje o validitě a reliabilitě více než 3000 otázek z experimentů 2 SB MTMM realizovaných v ESS 1 až 3, ale i z desítek dalších studií [Oberski, Gruner, Saris 2011]. V současné době dosud nejsou k dispozici výpočty validity a reliability z 2 SB MTMM experimentů z vln ESS 4 až 6, ale v nejbližší době by měly být do SQP 2.0 dodány [Saris, Gallhofer 2014: 245]. Dodání dalších údajů a kvalitě dotazníkových položek by mělo podle týmu RECSM vyústit ve zlepšení kvality odhadů z SQP 2.0, neboť algoritmus bude mít k dispozici větší množství informací pro predikci validity a reliability.

Pro názornost je třeba explicitně rozdělit dvě hlavní funkce SQP 2.0, které mohou analytičky a analytici využít. První z nich přímo nesouvisí s odhadovací funkcí programu, neboť spočívá pouze ve vyhledání ukazatele validity a reliability položky z některého konkrétního 2 SB MTMM experimentu, který v minulosti proběhl v šetřeních ESS. Tato funkce je však velice užitečná, neboť získání odhadů reliability a validity ve vlastní analýze je ve většině případů ne-

5 Vysvětlení predikčního algoritmu programu SQP 2.0 je vysoce technicky náročné, jeho výpočet je založen na pokročilém typu regresní analýzy a podrobný popis toho, jak byl vytvořen, dalece přesahuje možnosti tohoto článku. Zájemce o vysvětlení algoritmu odkazují na [Saris, Gallhofer 2014] a [Saris, W. E. et al. 2011].

možné. Jak ukážu dále v tomto textu, získat odhady validity a reliability v SEM analýze dat z nějakého 2 SB MTMM experimentu realizovaného na výběru z jedné země v mnoha případech vůbec nelze, a pokud chceme znát hodnoty ukazatelů validity a reliability z nějakého takového experimentu, musíme je vyhledat právě v databázi SQP 2.0, kde jsou uvedeny odhady pocházející ze složitých výpočtů týmu RECSM.

Druhou a tou podstatnou funkcí programu je predikce, tedy získání odhadu validity a reliability jakékoliv dotazníkové položky bez ohledu na to, jestli tato položka již byla použita v nějakém šetření, nebo ne. Pokud chceme v SQP 2.0 získat predikci validity a reliability pro nějakou dotazníkovou položku, musíme se registrovat na webové adrese www.sqp.nl, kde můžeme v on-line prostředí začít otázku hned kódovat. Znění otázky, škálu, na které je odpovídána, a další text, který může otázku doplňovat, vepíšeme do příslušné kolonky na stránce a následně položce přiřazujeme relevantní kódy, které software nabízí. Položku ručně kódujeme podle více než šedesáti charakteristik. Přesný popis procesu kódování je k dispozici v [Saris, Gallhofer 2014], nicméně není třeba jej podrobně studovat, neboť program na webu během kódování u každého kroku podává nápovědu a otázku je možné zakódovat i bez předchozího studia manuálu. Po dokončení procesu kódování program pomocí algoritmu vypočítá očekávanou validitu, reliabilitu a CMV dotazníkové položky.

Nabízejí se dvě cesty, jak odhady z SQP 2.0 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 2.0 odhadne validitu a reliabilitu položky nízkou nebo nedostačující, je možné vzít si tyto údaje jako základ pro uvažování nad vylepšením otázky. Program dokonce sám umí navrhnout zlepšení ve formulaci položky. SQP 2.0 se tak stává užitečným pomocníkem ve fázi přípravy výzkumu, neboť umí odhalit případné problémy ještě před tím, než je položka, resp. dotazník vyslán do terénu.

Problematické aspekty analýzy 2 SB MTMM dat

Metoda sběru dat 2 SB MTMM a analýza těchto dat není bez problémů. Ty se dají rozdělit na dvě skupiny: prvními jsou technické problémy, z nichž ty nejpodstatnější budu diskutovat v textu dále. V druhé skupině jsou problémy spíše epistemologické a vyplývající z omezení, které obecně má kvantitativní sociální věda a inferenční statistická analýza. Problémem tohoto druhu je např. fakt, že do každého modelu vstupuje omezené množství informací o sledovaném jevu a že v případě, kdy do analýzy dodáme další relevantní informace k odhadu vztahů mezi proměnnými, mohou se vztahy mezi proměnnými, jež byly zahrnuty do původního, prvního modelu, více či méně změnit. Můžeme tak získat dvě konkurující řešení, nikoliv jednu, objektivní hodnotu platnou pro vztah mezi proměnnými. Podrobné vysvětlení tohoto problému je ale látkou k samostatnému článku, proto se mu v tomto textu nemohu věnovat. Zde se soustředím pouze na technické problémy, které provázejí analýzu dat z 2 SB MTMM designu.

Důvody pro implementaci designu 2SB MTMM již byly v tomto textu představeny. Design 2SB MTMM sice řeší problémy spojené s designem MTMM, přináší však obtíže jiné, které se projevují v analýze dat. První z nich je obtížnost analýzy datových souborů o velikosti 1000 až 3000 případů, tedy souborů odpovídajících jedné zemi v šetření ESS, z nichž často není možné získat odhady kvality. Další problém vyplývá ze způsobu, jakým tým RECSM musí 2 SB MTMM data analyzovat, aby byl vůbec schopen získat odhady validity a reliability. Tento postup spočívá v zahrnutí dat ze všech zemí ESS do jedné analýzy a je velmi technicky náročný, z čehož plyne jistá netransparentnost výsledků RECSM. Posledním a poměrně zásadním bodem je fakt, že v některých případech nelze z 2 SB MTMM dat získat odhady kvality vůbec. Všechny naznačené problémy detailně popisují v následující části textu.

Nekonvergence modelu v datovém souboru s 1000 až 3000 případy

2 SB MTMM design sběru dat 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 že se v důsledku častého opakování podobné otázky unaví a odpovídá nedbale. Protože 2SB 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 ballot zásadně snižuje, a proto je přínosem a metodologickou inovací. Aplikace metody s sebou nicméně přináší komplikace; data z 2SB MTMM lze analyzovat metodou strukturálního modelování stejně jako MTMM data pocházející z úplného designu, ovšem taková analýza vede častěji k nekonvergujícímu řešení nebo k častému výskytu Heywoodových případů [Revilla, Saris 2011b]. 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 2013, Saris, Gallhofer 2014: 220]. V mnoha případech nelze pro tyto relativně malé datové soubory vůbec získat řádné konvergující řešení, a tedy ani odhady validity a reliability. Analytici a analytičky, kteří se rozhodnou začlenit chyby měření do své analýzy dat z jedné země, tak v mnoha případech nemohou odhady reliability a validity získat sami a musí použít ty, které obsahuje SQP 2.o.

Nekonvergence a častý výskyt HC v 2SB MTMM datech jsou důsledkem toho, že v matici korelací devíti MTMM proměnných sebraných designem 2SB MTMM na rozdíl od dat z kompletního MTMM chybí korelace mezi indiká-

tory měřenými metodami 1 a 2, což vede ke snížení počtu stupňů volnosti TS MTMM modelu. Řešením problému je využití designu 3 SB MTMM, tedy split ballot designu se třemi podsoubory, který uchovává výhody split ballot MTMM přístupu. 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. V datech jsou tak přítomny korelace mezi všemi devíti proměnnými v experimentu, takže se v analýze téměř nevyskytuje nekonvergence a HC [Revilla, Saris 2011]. Data z designu 3 SB MTMM však až na výjimku několika studií nejsou k dispozici; design se skoro nepoužívá, neboť je náročný na organizaci výběrového šetření. A jeho největším nedostatkem je to, že neposkytuje úplná data za všechny jednotky v souboru pro minimálně jednu proměnnou z trojice měřeného konceptu. To je důvod, proč tento design nebyl použit ani v jednom šetření ESS.

Nekonvergence a HC v českých souborech 2 SB MTMM dat

Vzhledem k tomu, že RECSM upozorňuje na obtížnost analýzy souborů dat z jedné země, zajímalo mě, zda je, či není možné získat odhady kvality položek z českých dat. Analyzovala jsem data pro Českou republiku z vln ESS 1 a ESS2, dílčím cílem bylo porovnat mé odhady reliability a validity s odhady, které jsou k dispozici v SQP 2.0, který pro česká data poskytuje jen odhady z první a druhé vlny ESS. Analýzu jsem prováděla s použitím výše popsaného TS MTMM modelu v SEM softwaru LISREL [Jöreskog, Sörbom LISREL 8.72], k odhadu parametrů modelu byla použita metoda multigroup maximum likelihood.

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 HC. Tento výsledek je z části důsledkem toho, že jsem v analýze na rozdíl od týmu RECSM neuplatňovala jiné restriktce, než které jsou obsaženy v popisu TS MTMM modelu v tomto článku. Po obdržení nekonvergujícího řešení nebo řešení s HC jsem neimplementovala úpravy, jako je např. fixace některých faktorových zátěží, které by mohly vést k získání řádného konvergujícího řešení; nevyužila jsem tedy strategii, kterou uplatňuje RECSM (viz dále).

V datech z ESS1 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 ESS 2 jsem získala tři řešení s HC, jedno nekonvergující a jedno řádné konvergující řešení. Tím jediným konvergujícím řešením byl experiment s genderovou tematikou nazvaný Role mužů a žen ve společnosti popsán v tomto článku výše. Syntax k tomuto modelu je uveden v příloze tohoto článku⁶.

6 Syntaxe ke všem modelům, které jsem použila na data z 2 SB MTMM experimentů z ESS1 a ESS2, jsou stejné. Liší se pouze zadání dat, tj. korelační matice, průměry a směrodatné odchylky.

Tabulka 2. Srovnání odhadů validity a reliability položek z experimentu *Role mužů a žen ve společnosti*

Zdroj:	ν^2		Γ^2	
	Model	SQP	Model	SQP
T1M1	0,846	0,962	0,640	0,619
T2M1	0,624	0,976	0,372	0,249
T3M1	0,902	0,972	0,723	0,612
T1M2	0,240	NA	0,185	NA
T2M2	0,864	NA	0,830	NA
T3M2	0,593	0,907	0,292	0,421
T1M3	0,922	0,908	0,846	0,648
T2M3	0,903	0,889	0,656	0,633
T3M3	0,941	0,930	1, 00	0,734

NA – Not available. SQP 2.0 z blíže nespecifikovaného důvodu neposkytuje pro tyto položky odhady kvality. Týmu RECSM se pravděpodobně nepodařilo získat v TS MTMM modelu odhady pro tyto položky.

Zdroj dat: European Social Survey, vlna 2. Vlastní analýza českých dat ESS2 a využití databáze programu SQP 2.0 (hodnoty z SQP 2.0 jsou rovněž platné pouze pro česká data).

Pro zajímavost jsem porovnávala odhady validity a reliability ze své analýzy českých dat z experimentu Role mužů a žen ve společnosti s odhady, které obsahuje SQP 2.0 (viz tabulka 2). Je evidentní, že mé odhady validity a reliability se od těch z programu SQP 2.0 v mnoha případech liší. Příčinou může být rozdílná analytická strategie, kterou jsem použila já, od strategie, kterou používá tým RECSM. Španělští výzkumníci a výzkumnice v SEM analýzách vyžadují co nejlepší ukazatele vhodnosti modelu a pro dosažení vynikajících ukazatelů implementují množství úprav modelu. Já jsem v analýze dat experimentu Role mužů a žen ve společnosti obdržela ukazatele vhodnosti modelu o hodnotách Chi2: 121.3, d.f.: 57, RMSEA: 0.035, CFI = 0,97, které jsou podle odborné literatury dobré až vynikající [Kenny 2015]. Tým RECSM by však s těmito hodnotami nebyl spokojen a implementoval by další restriktce k tomu, aby získal lepší ukazatele vhodnosti modelu. Z dále upravovaného modelu tak mohl získat odhady parametrů, které se lišily od těch v mém modelu. Dále mohly být rozdíly mezi mými odhady a těmi z SQP 2.0 způsobeny i tím, že RECSM analyzuje ESS data pro všechny země pohromadě (viz dále), a proto jsou jeho analýzy robustnější a přesnější.

Analýza 2 SB MTMM dat prováděná týmem RECSM

Tým RECSM se musel nějak vypořádat s problémem obtížně realizovatelné analýzy relativně malých datových souborů. Výsledky jeho dosavadní práce naznačovaly, že řádná konvergence modelů na 2 SB MTMM datech je tím pravděpodobnější, čím je analyzovaný soubor dat větší [Revilla, Saris 2013]. Vyvinul tedy strategii hromadné analýzy 2 SB MTMM dat, do které vstupují data ze všech zemí, které se v dané vlně účastnily ESS. Tato strategie je poměrně časově a technicky náročná, umožňuje však získat odhady validity a reliability pro naprostou většinu otázek z 2 SB MTMM experimentů.

RECSM analyzuje najednou 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]. Postupně upravuje restriktce modelu, až najde unikátní řešení pro každou zemi zvlášť. Analýzu provádějí dva na sobě nezávislí analytici, kteří porovnávají své výsledky po každém analytickém kroku. Analýza končí, když je nalezen nejlepší model fit, tedy takové řešení, kdy data co největší možnou měrou odpovídají použitému modelu. K určení nejlepšího modelu z řady tzv. zahnízđených modelů (*nested models*) používá RECSM program JRule [Van der Veld et al. 2008], který kontroluje chybu 1. a 2. druhu.

Přestože je TS MTMM model modelem konfirmativní faktorové analýzy, postupuje tým RECSM při hledání nejlepšího modelu více či méně exploračně, protože upravuje restriktce modelu tak, aby získal model, který maximálně odpovídá datům vloženým do analýzy. Tato úprava modelu však nejde za hranice předem stanovených, logicky odůvodnitelných vztahů mezi proměnnými naznačených v TS MTMM modelu. Cílem analýzy 2 SB MTMM dat není konfirmace teoretických předpokladů, ale nalezení co nejpřesnějších odhadů koeficientů kvality; z toho důvodu je více méně explorační přístup týmu RECSM v CFA analýze legitimní.

Náročnost procedury používané RECSM má za následek, že analytičky a analytici mimo tým RECSM se ve výsledcích týmu hůře orientují. 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 nakládat s chybějícími hodnotami v analýze s využitím TS MTMM modelu. Vzhledem k tomu, že autoři doporučují pro analýzu metodu odhadu multigroup maximum likelihood (group je v případě analýzy 2 SB MTMM dat z jedné země proměnná příslušnost k podsouboru split ballot s hodnotami 1 nebo 2) v programu LISREL, kde musí být vstupními daty dvě matice korelací mezi proměnnými, musíme se rozhodnout, zda budeme používat metodu *listwise deletion*, nebo *parwise deletion*, a pokud budeme používat pairwise deletion, musíme se nějak vypořádat s faktem, že máme různé počty případů pro korelace dvojic proměnných z MTMM. To je problém, neboť v syntaxi programu LISREL můžeme zadat na jednu matici korelací pouze jedno

číslo značící počet případů v analýze. Pokud nebudeme tým RECSM kontaktovat a rozhodneme se na základě vlastní úvahy, můžeme získat jiné odhady. Pracovníci a pracovnice týmu RECSM jsou však velmi vstřícní a na dotazy zájemců odpovídají emailem ochotně a velmi rychle. Není tedy překážkou jim napsat a získat odpověď. Výše popsany problém například RECSM řeší tak, že používá pairwise deletion a pro počet chybějících hodnot volí průměr ze všech případů pro všechny dvojice korelací⁷. Z ochoty týmu odpovídat na otázky usuzuji, že určitá míra netransparentnosti výsledků RECSM není úmyslná, ale vyplývá z toho, že projekt, který řeší, je technicky a časově vysoce náročný a tým není schopen zanášet všechny detaily procesu výpočtu do svých publikací a umisťovat dodatečné materiály na web.

Nemožnost získat odhady pro některé experimenty

Největším problémem metody 2 SB MTMM je to, že některá 2 SB MTMM data nelze úspěšně analyzovat vůbec, tedy ani při analýze dat ze všech ESS zemí najednou. Jedním z těchto případů je experiment 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. Odhady kvality pro tyto otázky tedy nejsou k dohledání v SQP 2.0 a jejich kvalita je neznámá.

Nekonvergence je v případě 2 SB MTMM experimentu Media způsobena tím, že skutečná korelace mezi latentními znaky (F_i) je blízka nule [Saris, Gallhofer 2014: 218]. 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 rozptýl 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 [Saris, Satorra, Coenders 2004] a v analýze není možné dojít k odhadům validity a reliability.

7 Zdroj: e-mailová komunikace se členkou týmu RECSM Melanií Revillovou (Melanie Revilla), Ph.D., v červnu 2014.

Závěr

Cílem tohoto textu bylo poskytnout základní uvedení do problematiky sběru a analýzy 2 SB MTMM dat v ESS a naznačit využití odhadů validity a reliability. Článek byl psán s úmyslem usnadnit českým kvantitativně zaměřeným sociologům a socioložkám orientaci v oblasti chyb měření v sociálněvědních kvantitativních datech a poskytnout jim základní informace pro úvahy o vhodnosti používání konceptů validity a reliability v analytické praxi. V tomto textu nepodávám odpověď na to, zda je lepší odstraňovat z dat nevalidní rozptyl, nebo zda má větší opodstatnění používat celý rozptyl měřených proměnných, jak velí zažitá praxe. Chtěla jsem pouze zprostředkovat informace, které mohou analytiku a analytičky dosud neseznámené s problematikou chyb měření uvést do problému, a inspirovat je k přemýšlení o chybách měření.

Text se ani nevyslovuje o důvěryhodnosti predikčních schopností programu SQP 2.0. Čtenářky a čtenáře možná při četbě o predikční funkci programu napadly různé otázky. Je ve výzkumu postojů a názorů vůbec možné něco predikovat? Lze z dat z výběrových 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 odpovědi. Analýzám věrohodnosti programu se dosud věnuje pouze tým RECSM, který se zabývá dílčími otázkami. Saris a Gallhoferová [2014: 240] například dokládají, že hodnoty odhadů z SQP 2.0 jsou stejné nebo velice podobné těm, které tým získal z empirických měření z 2 SB MTMM. Výjimkou jsou případy, kdy je validita a reliability položky získaná v modelu MTMM nízká; v takovém případě SQP 2.0 vypočítá výrazně vyšší hodnoty koeficientů kvality a predikce je chybová [Saris, Gallhofer 2014: 252].

V současné době nelze přijmout konečné rozhodnutí o tom, zda predikční funkce SQP 2.0 poskytuje důvěryhodné odhady validity a reliability. Dosavadní výsledky týmu RECSM jsou optimistické, přesvědčivé odpovědi na obecné otázky však nedávají. Zda bude použití programu SQP 2.0 někdy běžnou součástí analýzy kvantitativních dat, jako je dnes např. vážení dat k zajištění reprezentativity, se teprve uvidí.

Mgr. JOHANA CHYLÍKOVÁ je doktorandkou v Sociologickém ústavu AV ČR, v.v.i. Pracuje v oddělení Českého sociálněvědního datového archivu a zaměřuje se na studium kvality dat ze sociálněvědních výběrových šetření.

Literatura:

- Andrews, F. 1984. „Construct validity and error components of survey measures: A structural modeling approach.“ *Public Opinion Quarterly* 46: 409–442.
- Biemer, P. B. et al. 2004. *Measurement Errors in Surveys*. New Jersey: Wiley.

- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Campbell, D. T., D. W. Fiske. 1959. „Convergent and discriminant validation by the multitrait-multimethod matrix.“ *Psychological Bulletin* 6: 81–105.
- Groves, R. M. 2004. *Survey Error and Survey Costs*. New Jersey: Wiley.
- Harkness, J. A., F. J. R. van de Vijver, P. P. Mohler. 2002. *Cross-Cultural Survey Methods*. New Jersey: Wiley.
- Jöreskog, K. G., D. Sörbom LISREL 8.72.
- Kenny, David A. 2015. citováno z webové 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.
- 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.
- Lord, Frederick, M., Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley Publishing Company.
- 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.
- Revilla, Melanie, Saris, Willem E. 2013. „The Split-Ballot Multitrait-Multimethod Approach: Implementation and Problems.“ *Structural Equation Modeling* 20:27–46.
- Revilla, M., W. 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.
- 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, Jan. 1998b. „Kvalita dat II. Přístupy ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur.“ *Sociologický časopis* 34 (2): 195–204.
- Saris, W. E., I. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Wiley New Jersey.
- Saris, W. E. et al. 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., 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., Andrews, F. M. 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.
- 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.
- Urbánek, T., D. Denglerová, J. Širůček. 2011. *Psychometrika. Měření v psychologii*. Praha: Portál.
- Van der Veld, W. M., W. E. Saris, A. Satorra. 2008. Judgment Aid Rule Software.
- van Meurs, A., W. E. Saris. 1990. „Memory Effects in MTMM Studies.“ Pp. 134–46 in Saris, Willem E., van Meurs. *An Evaluation of Measurement Instruments by Meta-analysis of Multitrait-Multimethod Studies*. Amsterdam: North Holland.
- Werts, C. E., R. L. Linn. 1970. „Path analysis: Psychological examples.“ *Psychological Bulletin* 74, 194–212.

**Příloha 1. Syntax programu LISREL k odhadu parametrů
TS MTMM modelu 2 SB MTMM experimentu Role mužů a žen
ve společnosti z ESS2.**

Analýza 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