
Odhad reliability vybraných položek z českých šetření EU SILC kvazisimplexovým modelem*

JOHANA CHYLÍKOVÁ**

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

Estimating the Reliability of Selected Items in Czech EU-SILC Data Using the Quasi Simplex Model Method

Abstract: The article presents estimates of the reliability of measurement in the Czech surveys carried out in the EU-SILC international longitudinal research project. The reliability estimates were obtained using the Quasi Simplex Model (QSM), which has never before been used in Czech research. An analysis was carried out on all the items in the EU-SILC questionnaire that fulfilled the criteria for the QSM analysis: PH010, the item that asks respondents about their subjective health, HS120, the item that asks about the household's financial situation, and HS130, which asks what the minimum sufficient income of a household is. The analysis drew on all available data from Czech EU-SILC surveys, that is, data from five rotating panel surveys carried out between 2005 and 2012. The QSM analysis showed that for the selected items EU-SILC data are highly reliable; the estimated reliability of each item was around 0.8, for HS130 it was even above 0.9. The steadiness of the results was confirmed by the high consistency of the reliability estimates across all the panels. A small difference was observed between the reliability of data collected using the PAPI mode and data collected using CAPI. Given the attributes of the QSM model, however, it was impossible to test statistically whether the reliability of PAPI and CAPI data differ significantly.

Keywords: quantitative data, reliability, structural equation modelling, Quasi simplex model, EU-SILC

Sociologický časopis / Czech Sociological Review 2016, Vol. 52, No. 1: 79–106

DOI: <http://dx.doi.org/10.13060/00380288.2016.52.1.243>

* 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 socio-ekonomický výzkum“ sponzorovaného Ministerstvem školství, mládeže a tělovýchovy ČR s registračním číslem LM2010006. Data EU-SILC využita k analýze v tomto článku byla použita na základě smlouvy č. EUSILC/2007/16, uzavřené 3. 8. 2007 mezi Evropskou komisí, Eurostatem a Sociologickým ústavem AV ČR, v.v.i. Autorka děkuje PhDr. Martině Mysíkové, Ph.D., za cenné rady, které jí při přípravě tohoto článku poskytla.

** Veškerou korespondenci posílejte na adresu: Mgr. Johana Chylíková, Český sociálněvědní datový archiv, Sociologický ústav AV ČR, v.v.i., Jilská 1, 110 00 Praha 1, e-mail: johana.chylikova@soc.cas.cz.

Úvod

Vysoká reliabilita 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. V běžné výzkumné praxi je ovšem nemožné informace o reliabilitě měřených položek získat. Je to z toho důvodu, že z jednoho jediného měření dotazníkové položky nelze reliabilitu měření zjistit. Každá metoda zjišťování reliability vyžaduje speciální design výběrového šetření, jako je např. test-retest metoda, tedy opakované měření té samé položky, nebo Multitrait-multimethod design (MTMM) analyzovaný jako tzv. True Score MTMM model [Saris, Gallhofer 2014], který předpokládá měření minimálně tří konstruktů minimálně třemi metodami. V tomto článku prezentovaná metoda kvazisimplexového modelu (z angl. Quasi Simplex Model; QSM) se od výše uvedených metod používajících speciální výzkumný design 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 substantiálních informací o zkoumaném fenoménu. QSM je 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 reliabilitu 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 substantiálních sociálněvědních témat.

QSM vyžaduje měření minimálně ve třech vlnách a data analyzuje v softwaru pro modelování strukturálních rovnic (z angl. Structural Equation Modeling, SEM) [Alwin 2007]. V této stati 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 pocítovaný 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 čeští 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 tomto článku 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. Problematiku biviační a multivariační analýzy dat bez složek chyby měření po-

pisuje např. Saris a Gallhoferová [2007], kteří mimo jiné ukazují, jak velký vliv na zjištěné lineární vztahy mezi proměnnými může mít chyba měření.

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 [1998], popis a ukázka použití však chybí, přestože metoda zažila svůj rozkvět již v sedmdesátých letech 20. století. Uvedení této v zahraniční odborné literatuře rozšířené metody do českého prostředí tedy považuji za důležitý přínos tohoto článku.

Kvazisimplexový model (QSM) a metoda 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/ka je dotázán/a 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 tzv. klasické testové teorie [ibid. 1968]. Tato teorie je vyjádřena v rovnicích:

$$\text{var } X = \text{var } T + \text{var } e \quad (1)$$

$$\text{rel } X = \frac{\text{var } T}{\text{var } X} \quad (2)$$

příčemž platí podmínky:

1. $E[e] = 0$, tedy že střední hodnota náhodné chyby měření e je nula.
2. $\text{COV}[T, e] = 0$, tedy že pravý skóre T je nezávislý na chybě měření e .
3. Při více měřeních jsou náhodné chyby e nekorelované. V metodě test-retest jsou tedy chyby e pro první a druhé měření nekorelované, to samé platí pro QSM.
4. Jednotlivá měření v test-retest, ale i v QSM jsou na sobě nezávislá.

Rovnice 1 říká, že rozptyl měřené proměnné X se rovná rozptylu teoreticky uvažovaného tzv. pravého skóre (T), tedy latentní proměnné stojící za měřenou proměnnou, a rozptylu tzv. náhodné chyby (e), která je definována právě tím, že její průměr se rovná nule a že není závislá na pravém skóre (viz podmínky 1 a 2). Rovnice 2 ukazuje, že reliabilita měření je dána jako poměr mezi rozptylem pravého skóre T a měřenou proměnnou X . Z těchto podmínek vyplývá, že reliabilita nabývá hodnot od 0 do 1, přičemž vyšší hodnota znamená vyšší reli-

abilitu. V metodě test-retest je za reliabilitu měření dané dotazníkové položky považována korelace dvou opakovaných měření. Pro odečítání reliability z modelu strukturálních rovnic, což je případ QSM, platí, že reliabilita X je mocnina faktorové zátěže mezi měřenou proměnnou X a jejím pravým skórem T , tedy, že $\text{rel } X = r^2(X, T)$.

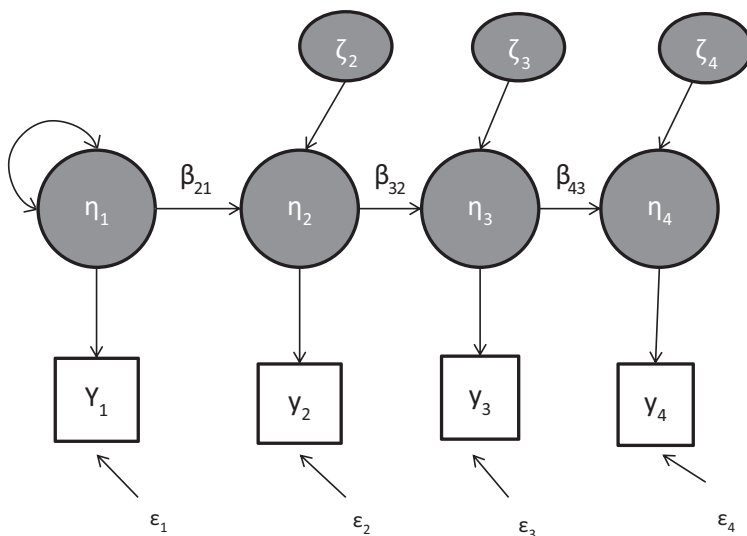
Metoda test-retest má problémy zásadního charakteru. Zaprvé nelze vyloučit, že respondenti, kteří na tu samou otázku odpovídají dvakrát, si svoji první odpověď pamatují. V případě, kdy je druhá odpověď závislá na té první, je porušen předpoklad o nezávislosti jednotlivých měření (podmínka 4) a zjištěná 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 nevalidní buď z prvního, nebo z druhého uvedeného důvodu.

Kvazisimplexový model problémy metody test-retest řeší 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 skóreů v čase odpovídá tzv. Markovovým řetězcům, tedy že pravý skóre v čase i může být přímo ovlivněn pravým skóreem v čase $i-1$, ale nikoli pravými skóreem v čase $i-2$ a staršími – jejich vliv může být pouze zprostředkovaný hodnotou v čase $i-1$ [Alwin 2007]. Očekávaná změna mezi pravými skóreem tedy vyplývá ze stochastického modelu, nikoliv z nějakého empiricky pozorovaného trendu.

V úvodu tohoto článku jsem uvedla, že metoda QSM je použitelná pro panelová data pocházející z minimálně tří vln měření, taková data však mají analytické omezení. Pro data ze tří vln není možné prostřednictvím modelování strukturálních rovnic (SEM) ověřit, zda jsou získané odhady reliability spolehlivé, tzn. nelze zjistit ukazatele shody modelu s daty (z angl. model fit). Kvazisimplexový model pro data ze tří vln je tzv. saturovaný (angl. just identified) a jeho počet stupňů volnosti (z angl. degrees of freedom) 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 šetření.¹ Teprve pak je možné získat ukazatele shody modelu s daty a stanovit, nakolik lze odhadnutým parametřům důvěřovat.

¹ Přidáním další, v pořadí čtvrté vlny se počet stupňů volnosti modelu zvýší na dva.

Obrázek 1. Kvazisimplexový model se čtyřmi indikátory



Poznámka: Obrázek 1 používá klasické maticové značení používané v SEM analýze, znaky η_i jsou latentní proměnné pravého skóru T_i , $\beta_{i,i-1}$ jsou tzv. stability, ζ_i reprezentuje zbytkový rozptyl latentních proměnných pravého skóru; ϵ_i je rozptyl měřené proměnné y_i odpovídající náhodné chybě měření.

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

V souladu s klasickou testovou teorií (viz rovnice 1 a 2) rozděluje QSM rozptyl měřených proměnných y_i do dvou částí: jedna část y_i odpovídá latentní proměnné pravého skóru (η_i), druhá náhodné chybě měření (ϵ_i). Latentní proměnné pravého skóru η_2 , η_3 a η_4 reflektují vliv, který reprezentuje možnou změnu v hodnotě pravého skóru, ke které mohlo dojít v čase, který uplynul mezi vlnami dotazování. Na Obrázku 1 je tento vliv 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. 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 3 vyjadřující část měření kvazisimplexového modelu (podíly rozptylu měřených proměnných) a rovnice 4 vyjadřující strukturální část kvazisimplexového modelu (podíly rozptylu latentních proměnných) jsou:

$$Y_i = \eta_i + \varepsilon_i \quad (3)$$

$$\eta_i = \beta_{i,i-1} + \eta_{i-1} + \zeta_i \quad (4)$$

Aby bylo možné s kvazisimplexovým modelem analyticky pracovat v softwaru pro SEM, je nutné na některé parametry uplatnit tzv. 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. 2004]. 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. Analyticky a analytici s ní pracují buď jako s nástrojem pro zjištění reliability používaných dotazníkových položek [Achen 1975; Alwin 1989; Alwin 2007; Alwin, Krosnick 1991; Biemer, Christ, Wiesen 2009; Blok, Saris 1983; Erikson 1979; Ehrhardt, Saris, Veenhoven 2000; Jagodzinski, Kuhnel 1987; 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 [Judd, Milburn 1980; Kenny, Campbell 1989; Green, Palmquist 1990; Schilling 2006].

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 jiné známé metody odhadu reliability test-retest byly již v předcházející části textu uvedeny, kritiku dalších metod odhadu reliability obsahují např. práce Bollena [1989] nebo Revilly a Sarise [2013a]. V následující části textu shrnu kritické připomínky, které vznesli někteří autoři a autorky k použití kvazisimplexového modelu. Saris a Gallhoferová [2014] identifikují dva zásadní problémy QSM přístupu. Zprvu 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 reliability 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 již se nemusí objevit v následující vlně. Kvazisimplexový model však tento rozptyl nerozpozná jako substanciálně 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 reliability. Z výše uvedených důvodů autoři zavrhnou QSM jako nevhodnou metodu měření reliability a razantně upřednostňují vlastní metodu 2 Split ballot MTMM.

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 kvazisimplexového modelu vůbec statisticky testovat prostřednictvím 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 na Obrázku 1 a použitým pro analýzu v tomto článku. Tento model 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 reliability 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. 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 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 reliability 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/ka 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. Podle Wileyeého a Wileyeého [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 vícepoložkových škál. Pokud je testována reliabilita jedné položky, jako je tomu např. v tomto článku, nelze Biemerem a jeho kolegy navržený přístup uplatnit. Autoři práce dále upozorňují na nebezpečí toho, že náhodné chyby mohou v da-

tech 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 zapomenout. Za krátkou dobu považuje odborná literatura období kratší než měsíc [van der Veld, Saris 2005: 6].

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 2.0.² Design 2 SB MTMM je variací na klasické MTMM s tím rozdílem, že je výběrový soubor rozdělen do dvou podsouborů, z nichž v každém odpovídají respondenti na 3×2 verze dotazníkové položky, nikoliv na 3×3 verze, jak je tomu v klasickém MTMM [Saris, Satorra, Coenders 2004]. Realizace šetření s využitím 2 SB MTMM je organizačně obtížná a stojí více peněz než běžné výběrové šetření. Navíc stejně jako jiné metody vycházející z klasické testové teorie 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 stejně jako QSM metodou, která má problémy, jež nelze eliminovat.

Závěrem lze tedy na obhajobu QSM 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 se držím v následující analýze, kde přijímám pouze výsledky z modelů, které měly vynikající nebo dobré ukazatele shody modelu s daty.

Zdroje dat k analýze kvazisimplexovým modelem a šetření European Union Statistics on Income and Living Conditions (EU SILC)

V České republice existuje pouze 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

² Databáze je uložena na adrese <http://www.sqp.nl>.

Tabulka 1. Složení rotačního panelu EU SILC a průřezových souborů, data ČR

Rok	2005	2006	2007	2008	2009	2010	2011	2012
Účast v panelu v daném roce	x	x	x	x				
		x	x	x	x			
			x	x	x	x		
				x	x	x	x	
					x	x	x	x

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. Přislíbem je např. projekt SHARE, který v současné době nabízí srovnatelná data pro tři vlny.³ 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 let.⁴ V tomto roce bylo rovněž zahájeno velké panelové šetření nazvané Proměny české společnosti, v jehož rámci bude proveden sběr dat ve čtyřech vlnách v letech 2015–2018.⁵

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. Pod hlavičkou EU SILC jsou shromažďovány informace o příjmech domácností, chudobě, sociální exkluzi a životních podmínkách obyvatel. Respondenti odpovídají na otázky zjišťující 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 poskytuje článek Martiny Mysíkové [2011].

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 1, údaje ve sloupcích; v tabulce je vidět, že první průřezový soubor obsahující čtyři panelové podsoubory vznikl až v roce 2008).

³ Data ze třetí vlny šetření nazvaného SHARE LIFE nejsou srovnatelná s daty z ostatních vln šetření.

⁴ <http://share.cerge-ei.cz>

⁵ <http://www.promenyceskespolecnosti.cz>

Vzhledem k tomu, že pro analýzu v tomto článku používám panelová data, nikoliv průřezová, podrobně 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 1, rotační panely jsou vyznačeny v řádcích.

V současné době je pro Českou republiku k dispozici již pět řad čtyřletých panelových dat (v době psaní tohoto 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

Existence většího množství čtyřletých panelů v šetřeních EU SILC je 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.

Položky z dotazníků EU SILC vhodné k 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).⁶ 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 dat,⁷ 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

⁶ Položky v dotazníku A byly pro analýzu nevhodné.

⁷ 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).

Tabulka 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? (navazuje na položku výše, pozn. aut.)	
PH010	Jak celkově hodnotíte svůj zdravotní stav?	1. velmi dobrý 2. dobrý 3. přijatelný 4. špatný 5. velmi špatný

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 2.

Z pěti položek uvedených v Tabulce 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/ka respondenta upozorní na nesrovnalost a vyplní údaj správně. Děje se tak v případě CAPI (Computer Assisted Personal Interview) i PAPI (Paper and Pencil Interview) módu sběru dat [Mysíková 2011].⁸ Data pro položky PL190 a PL200

⁸ Tuto skutečnost mi Martina Mysíková potvrdila i osobně na základě informací, které jí dodaly pracovnice ČSÚ.

tedy nereflektovala skutečnou odpověď respondenta, a proto nebyla pro analýzu reliability vhodná. Abych čtenářkám a čtenářům usnadnila orientaci v textu, budu v následujícím textu nazývat položku PH010 *PH010 Zdraví*, položku HS120 *HS120 Příjem* a položku HS130 *HS130 Odhad příjmu*

Výběr případů ke QSM analýze

Základním požadavkem QSM analýzy reliability je, aby se respondent/ka zúčastnil/a všech šetřených vln a aby byl/a ve všech šetřených vlnách totožný/á. 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/ka neúčastnil/a 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 [Saris, Gallhofer 2014], ale obecně v analýze reliability tato zásada platí.⁹

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, PAPI a velmi okrajově SAQ (Self-administered Questionnaire).¹⁰ V prvních třech rotačních panelech významně převládal mód sběru dat PAPI (cca 90 % 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 % CAPI, řada 2009–2012: 27 % 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 módů sběru dat jsem z analýzy vypustila. Proporci použitých módů sběru dat analyzovaných položek

⁹ 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.

¹⁰ 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.

Tabulka 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	2005–2008	2006–2009	2007–2010	2008–2011	2009–2012
Mód sběru	PAPI	5353	4796	3572	89	1678
	CAPI	Nepoužito	Nepoužito	Nepoužito	1283	661
	Ostatní	59	32	9	373	404
	Položka	Časová řada				
	HS120	2005–2008	2006–2009	2007–2010	2008–2011	2009–2012
Mód sběru	PAPI	2326	2028	1560	65	795
	CAPI	Nepoužito	Nepoužito	Nepoužito	1222	461
	Ostatní	15	10	6	537	274
	Položka	Časová řada				
	HS130	2005–2008	2006–2009	2007–2010	2008–2011	2009–2012
Mód sběru	PAPI	2312	2028	1560	90	795
	CAPI	Nepoužito	Nepoužito	Nepoužito	857	461
	Ostatní	14	10	6	182	274

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

PH010 Zdraví, HS120 Příjem a HS130 Odhad příjmu ukazuje Tabulka 3 (zvýrazněné buňky značí podsoubory EU SILC dat použitých pro analýzu QSM).

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 10–14 %, ¹¹ CAPI soubor z panelu 2008–2011 se zmenšil o 30–42 % a v případě panelu 2009–2012 se CAPI soubor zmenšil o 37–45 % a PAPI soubor o 19–31 %. Soubory PAPI dat odpovídající položce HS120 Příjem pro první tři panely byly v mé analýze o 31–33 % menší než původní soubory. V panelu 2008–2011 se soubor CAPI dat zmenšil o 12–20 %, v následujícím panelu byl soubor PAPI dat menší o 38–44 %, soubor CAPI dat se zmenšil o 37–48 %. Pro proměnnou HS130 Odhad příjmu platilo to samé nebo velmi podobné, co pro položku HS120 Příjem.

¹¹ Porovnávám změnu velikosti souboru pro každý rok v panelu, proto uvádím interval procentuální změny ve všech čtyřech letech.

QSM analýza 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 1 tohoto textu. Analýza druhého jmenovaného modelu skončila s velmi špatnými ukazateli shody modelu s daty,¹² proto odhady reliability 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 1 v první části textu). Rozptyly latentních proměnných (parametry Ψ) byly ponechány volné, stejně jako regresní koeficienty mezi latentními proměnnými (parametry β). Faktorové zátěže mezi měřenou proměnnou a jí příslušející latentní proměnnou (parametry λ) byly fixovány na hodnotu 1, což je standardní a nutný postup, jehož prostřednictvím získává latentní proměnná měřítko [Bollen 1989]. Rozptyl náhodné chyby první měřené proměnné byl ponechán volný (parametr θ), 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 QS modelu s daty pro položku *HS130 Odhad příjmu* buď nekonvergovala, nebo hlásila jiné problémy, v jejichž důsledku software nebyl schopen najít řešení. Na radu recenzenta, kterému bych chtěla tímto poděkovat, se mi problém s nekonvergencí modelu podařilo vyřešit 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, 5 a 6. Velikosti tzv. stabilit (β) , které nejsou pro odečtení reliability podstatné, jsou uvedeny ve zvláštní tabulce v Příloze 2 tohoto článku.

Naprostá většina jednotlivých QSM analýz dospěla k výsledku, který ukazoval na vysokou shodu modelu s daty. 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

¹² 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 tzv. Heywood cases, tedy odhadnuté rozptyly, faktorové zátěže a regresní koeficienty větší než ± 1 a odhadnuté rozptyly záporné hodnoty [Revilla, Saris 2013a].

Tabulka 4. Výsledky QSM analýzy pro položku PH010 Zdraví

Max. Likelihood											
položka PH010	řada 2005–2008										
Mod 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	0,000	0,0; 0,016	1,00	0,000	1,00
položka PH010	řada 2006–2009										
Mod 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	0,000	1,0
položka PH010	řada 2007–2010										
Mod 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	1,00
položka PH010	řada 2008–2011										
Mod 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	0,040	0,0052; 0,078	0,60	0,007	1,00
položka PH010	řada 2009–2012										
Mod 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	0,000	0,0; 0,041	0,98	0,002	1,00
CAPI	661	0,792	0,774	0,792	0,792	0,788	0,031	0,0; 0,090	0,61	0,006	1,00

Tabulka 5. Výsledky QSM analýzy pro položku HS120 Příjem

Max. Likelihood											
položka HS120	řada 2005–2008										
Mod 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	0,810	0,0	0,0; 0,031	1,00	0,002	1,00
položka HS120	řada 2006–2009										
Mod 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	1,00
položka HS120	řada 2007–2010										
Mod 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	0,000	0,0; 0,048	0,96	0,003	1,00
položka HS120	řada 2008–2011										
Mod sběru dat	počet případů	rel 2008	rel 2009	rel 2010	rel 2011	průměr rel	RMSEA	RMSEA 90% int	P value RMSEA	SRMR	CFI
CAPI	1222	0,757	0,774	0,74	0,74	0,753	0,022	0,0; 0,065	0,83	0,007	1,00
položka HS120	řada 2009–2012										
Mod 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	1,00
CAPI	461	0,723	0,689	0,706	0,656	0,695	0,000	0,0; 0,086	0,75	0,007	1,00

Tabulka 6. Výsledky QSM analýzy pro položku HS130 Odhad příjmu

Max. Likelihood											
položka HS130	řada 2005–2008										
Mod sběru dat	počet případů	rel 2005	rel 2006	rel 2007	rel 2008	průměr rel	RMSEA	RMSEA 90% int	P value RMSEA	SRMR	CFI
PAPI	2312	0,903	0,903	0,903	0,903	0,903	0,0	0,0; 0,033	1,00	0,0013	1,00
položka HS130	řada 2006–2009										
Mod 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	0,903	0,903	0,903	0,903	0,903	0,068	0,044; 0,096	0,11	0,0052	1,00
položka HS130	řada 2007–2010										
Mod 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	0,922	0,903	0,903	0,903	0,907	0,021	0,0; 0,059	0,88	0,0025	1,00
položka HS130	řada 2008–2011										
Mod sběru dat	počet případů	rel 2008	rel 2009	rel 2010	rel 2011	průměr rel	RMSEA	RMSEA 90% int	P value RMSEA	SRMR	CFI
CAPÍ	857	0,922	0,903	0,903	0,88	0,903	0,069	0,031; 0,11	0,18	0,0074	1,00
položka HS130	řada 2009–2012										
Mod 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	0,922	0,922	0,922	0,922	0,922	0,0	0,0; 0,0	1,00	0,00021	1,00
CAPÍ	461	0,922	0,903	0,903	0,903	0,907	0,0	0,0; 0,091	0,70	0,00038	1,00

lepší shoda. Dále píše, že hodnoty tzv. P value RMSEA nesmí být nižší než 0,05 [ibid.], cituje však práci Jöreskoga a Sörboma [1996], podle kterých by P value RMSEA neměla být nižší než 0,5. Podstatná je rovněž šíře intervalů spolehlivosti hodnot RMSEA; horní mez intervalu by ideálně neměla být vyšší než 0,08 [Hooper, Coughlan, Mullen 2008]. Ukazatel CFI by se měl ideálně pohybovat kolem hodnoty 0,95; pro ukazatel SRMR platí, že hodnoty kolem 0,05 indikují vynikající shodu modelu s daty [Byrne 2009].

Hodnoty ukazatelů shody modelu s daty pro položku položky *PH010 Zdraví* a *HS120 Příjem* byly ve všech panelech vynikající až dobré (viz Tabulky 4 a 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 devadesátiprocentní interval spolehlivosti ukazatele RMSEA překračoval doporučené hodnoty a ukazatel P value RMSEA byl menší než 0,5 (výsledky pro tyto řady jsou v Tabulce 6 vyznačeny šedým stínováním). Vzhledem k problematickým aspektům metody QSM, popsáním v teoretické části tohoto textu, jsem v této analýze uplatňovala striktní požadavky na shodu modelu s daty. Proto výsledky z kvazisimplexových modelů ze zmíněných panelů nepovažuji za důvěryhodné. Hranice ideálních intervalů spolehlivosti RMSEA byla rovněž překročena v analýzách menších datových souborů, tj. *PH010* 2009–2012 CAPI (661 případů), *HS120* 2009–2012 PAPI (795 případů), *HS120* 2009–2012 CAPI (461 případů) a *HS130* 2009–2012 CAPI (461 případů). V těchto případech však bylo překročení limitu důsledkem relativně malé velikosti souboru a současně ostatní ukazatele shody modelu s daty byly v normě, proto výsledky z těchto analýz považuji za věrohodné.

Reliabilita položky *PH010 Zdraví* se ukázala jako 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 čtvrté shodně 0,846, průměr činil 0,842. Ve třetím panelu byla reliabilita 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 dat metodou CAPI. V panelu 2008–2011 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 reliabilita 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 reliabilita v souboru PAPI 0,837. Pro data sebraná metodou CAPI byla reliabilita v roce 2009 0,792, dále 0,774, o rok později 0,792 a pak rovněž 0,792. Průměrně byla reliabilita v souboru CAPI 0,788.

Výsledky analýzy položky *HS120 Příjem* byly srovnatelné s těmi získanými pro *PH010 Zdraví*. V rotační 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 odhady shodně 0,903, průměrná reliability v tomto panelu tedy činila 0,907.

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 reliability, např. Peter [1979] uvádí pravidlo podle Nunnallyho, který stanovuje, že dostatečná reliability 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 reliability 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 [Andrews 1984], obsahují hodnoce-

ní, týkají se zdraví, resp. příjmu a obě používají škálu odpovědi – položka *HS120 Příjem* šestibodovou škálu, položka *PH010 Zdraví* pětibodovou. V odborné literatuře lze najít informace o reliability (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 reliability 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-assessment) či vnímání sebe sama (self-perception), tedy položky odrážející koncepty odpovídající položkám *PH010 Zdraví* a *HS120 Příjem*, měly v Alwinově studii reliability průměrně 0,5 [ibid. 2007: 131]. Podíváme-li se na reliability položek v Alwinově studii podle tématu, zjistíme, že otázky ohledně finanční situace respondentů či jejich domácností měly průměrnou reliability 0,592 [Alwin 2007: 161]. Data týkající se zdraví v této analýze bohužel nebyla. Studie Kroha [2006] byla provedena na experimentálních SB MTMM datech německého Socioekonomického panelu (SOEP) a testovala tři položky zjišťující spokojenost se životem, zdravím a finančním příjmem. Reliability položky zjišťující 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 reliability 0,69, v případě jedenáctibodové škály to bylo 0,84. Z uvedených údajů lze usoudit, že reliability 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 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 příjmu* je ještě přesvědčivější než u ostatních položek.

Pohled do Tabulek 4 a 5¹³ napovídá, že mohl existovat rozdíl v reliability 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 reliability CAPI a PAPI dat nebo se jednalo pouze o minimální statisticky nevý-

¹³ 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.

znamný rozdíl?¹⁴ K zodpovězení této otázky se nabízelo použití statisticko-analytické metody Multi-group Structural Equation Modeling (MG SEM), využívající Multi-group 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 módu 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 Multi Group SEM modelech proti sobě stojí dva a více modelů, které se mezi sebou musí v něčem lišit. Základní 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 jsou stejné, neboť všechny faktorové zátěže odpovídající koeficientu reliability jsou fixovány na hodnotu 1. Z tohoto důvodu nelze testovat rozdíl reliabilit 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 QS modelu. Nabízí se sice možnost testovat shodu rozptylů náhodných chyb a rozptylů pravých skóre, jejichž souvislost s reliabilitou 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 reliabilita, faktem je ovšem to, že reliabilita může zůstat stejná i v případě, kdy se rozptyl pravého skóre či náhodné chyby mění. Naznačená analýza tak umí poznat pouze případy, kdy se reliabilita napříč skupinami nelišila a zároveň byly stejné rozptyly náhodné chyby a pravého skóre. Případy, kdy se reliabilita nelišila, ale rozptyly chyby či latentní proměnné se lišily, neumí identifikovat. Jinými slovy, neumí najít všechny případy, kdy mohla být reliabilita CAPI a PAPI dat stejná.

¹⁴ 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.

Důvěryhodnou analýzu ne/shody reliability v CAPI a PAPI datech tedy touto cestou získat nelze.¹⁵

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 reliabilitě mezi CAPI a PAPI módem 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, 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 [Schräpler et al. 2006; de Leeuw, Hox, Snijders 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 ojedinělý nález.

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 popsaného 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 měřicích 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, a to o 0,13.

Tento text 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 módu sběru dat bohužel nebylo možné prokázat vzhledem k omezením, která kladou principy modelování strukturálních rovnic na analýzu ekvivalence parametrů.

¹⁵ 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.

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

Literatura

- Achen, Ch. H. 1975. „Mass Political Attitudes and the Survey Response.“ *American Political Science Review* 69 (4): 1218–1231, <http://dx.doi.org/10.2307/1955282>.
- Alwin, D. F. 2007. *Margins of Error. A Study of Reliability in Survey Measurement*. New Jersey: Wiley.
- 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, <http://dx.doi.org/10.1177/0049124191020001005>.
- Andrews, F. 1984. „Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach.“ *Public Opinion Quarterly* 46: 409–442, <http://dx.doi.org/10.1086/268840>.
- 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, <http://dx.doi.org/10.1037/a0016618>.
- Blok, H., W. E. Saris. 1983. „Using Longitudinal Data to Estimate Reliability.“ *Applied Psychological Measurement* 7 (3): 295–301, <http://dx.doi.org/10.1177/014662168300700305>.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Byrne, B. 2009. *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. New Jersey.
- Cernat, A., P. Lugtig, S. C. N. Uhrig, N. Watson. 2014. „Assessing and Relaxing Assumptions in Quasi-simplex Models.“ ISER Working Paper Series.
- de Leeuw, E. D., J. Hox, G. Snijders. 1998. „The Effect of Computer-Assisted Interviewing on Data Quality. A Review.“ Pp. 173–198 in B. Blyth. *Market Research and Information Technology*. ESOMAR Monograph.
- Dillman, D. A., L. M. Christian. 2005. „Survey Mode as a Source of Instability in Responses across Surveys.“ *Field Methods* 17 (1): 30–52, <http://dx.doi.org/10.1177/1525822X04269550>.
- 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, <http://dx.doi.org/10.1023/A:1010084410679>.
- Erikson, R. S. 1979. „The SRC Panel Data and Mass Political Attitudes.“ *British Journal of Political Science* 9 (1): 89–114, <http://dx.doi.org/10.1017/S0007123400001630>.
- Green, D. P., B. Palmquist. 1990. „Of Artifacts and Partisan Instability.“ *American Journal of Political Science* 34 (3): 872–902, <http://dx.doi.org/10.2307/2111402>.
- Groves, R. M. 2004. *Survey Errors and Survey Costs*. Chichester: John Wiley and Sons.
- Heise, D. R. 1969. „Separating Reliability and Stability in Test-retest Correlation.“ *American Sociological Review* 34: 93–101, <http://dx.doi.org/10.2307/2092790>.
- 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.

- 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, <http://dx.doi.org/10.1177/0049124187015003003>.
- 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., D. Sörbom. 2004. *LISREL 8.7 for Windows* [počítačový software]. Skokie, IL: Scientific Software International, Inc.
- 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, <http://dx.doi.org/10.2307/2095012>.
- Kenny, D. A., D. T. Campbell. 1989. „On the Measurement of Stability in Over-time Data.“ *Journal of Personality* 57 (2), <http://dx.doi.org/10.1111/j.1467-6494.1989.tb00489.x>.
- Kroh, M. 2006. *An Experimental Evaluation of Popular Well-Being Measures*. DIW-Diskussionspapiere, No. 546.
- Lord, F. M., M. R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading MA: Addison-Wesley 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, <http://dx.doi.org/10.1037/a0024298>.
- Mysíková, M. 2011. „EU-SILC a jeho metodologická úskalí: mezinárodní srovnatelnost a příjmové proměnné.“ *Data a výzkum – SDA Info* 5 (2): 147–170.
- Peter, J. P. 1979. „Reliability: A Review of Psychometric Basics and Recent Marketing Practices.“ *Journal of Marketing Research* 16: 6–17, <http://dx.doi.org/10.2307/3150868>.
- Revilla, M., W. E. Saris. 2013a. „The Split-Ballot Multitrait-Multimethod Approach: Implementation and Problems.“ *Structural Equation Modeling* 20: 27–46, <http://dx.doi.org/10.1080/10705511.2013.742379>.
- 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, <http://dx.doi.org/10.1093/ijpor/eds007>.
- Řehák, J. 1998. „Kvalita dat II. Přístupy k ohodnocování výzkumných instrumentů založené na modelování kovariančních struktur.“ *Sociologický časopis* 34 (2): 195–204.
- Saris, W. E., I. Gallhofer. 2003. *Report on the MTMM Experiments in the Pilot Studies and Proposals for Round 1 of the ESS*. University of Amsterdam.
- Saris, W. E., I. N. Gallhofer. 2007. „Estimation of the Effects of Measurement Characteristics on the Quality of Survey Questions.“ *Survey Research Methods* 1 (1): 29–43, <http://dx.doi.org/10.1002/9780470165195.ch12>.
- Saris, W. E., I. N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New Jersey: Wiley, <http://dx.doi.org/10.1002/9781118634646>.
- 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, <http://dx.doi.org/10.1111/j.0081-1750.2004.00155.x>.
- Schilling, O. 2006. „Development of Life Satisfaction in Old Age: Another View on the Paradox.“ *Social Indicators Research* 75: 241–271, <http://dx.doi.org/10.1007/s11205-004-5297-2>.
- 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, <http://dx.doi.org/10.1002/acp.2350050304>.

- 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, <http://dx.doi.org/10.1080/13645579.2013.823003>.
- 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.
- Wiley, D., J. Wiley. 1970. „The Estimation of Measurement Error in Panel Data.“ *American Sociological Review* 35 (1): 112–117, <http://dx.doi.org/10.2307/2093858>.

Použitá data

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

Přílohy

Příloha 1. Syntaxe programu LISREL použité k analýze QS modelů podle (1) Heiseho [1969] a (2) Wileyeého a Wileyeého [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 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	0,90	0,91
Řada 2006–2009	PAPI	0,90	0,90	0,90
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	0,90	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	0,90	0,95	0,95
Řada 2009–2012	CAPI	0,93	0,94	0,99