

---

## Zhodnocení vlivu práce výzkumných agentur na konstruktovou validitu škál<sup>1</sup>

MARTIN KREIDL\*

Sociologický ústav AV ČR, Praha  
a Fakulta filozofická, Západočeská univerzita, Plzeň

### Evaluating the Influence of Survey Agencies on the Construct Validity of Scales

**Abstract:** This article uses a multiple-indicators multiple-causes model to assess the validity of an attitude scale in different survey contexts. Data used in the article come from a methodological experiment that was carried out in the Czech Republic in May 2002. A set of standardised questions was added to the omnibus surveys of three major Czech agencies that routinely conduct public opinion surveys and publicize voting preferences. As all survey instruments in the experiment were standardised, the author argues that the only possible source of variable validity across individual agencies is their fieldwork procedures. Therefore, the model compares not only the construct validity of scales, but also the quality of the fieldwork in individual agencies. It turns out that the scales are valid in all survey contexts, and the assessment fails to identify any major differences between agencies in the quality of measurement.

*Sociologický časopis/Czech Sociological Review, 2005, Vol. 41, No. 1: 103–123*

### 1. Kvalita měření

Kvalita kvantitativního sociálního výzkumu závisí kritickým způsobem na kvalitě dat. Kvalita dat má dvě klíčové složky: reprezentativitu a kvalitu měření. Reprezentativita vypovídá o vztahu výběrového souboru a základní populace výzkumu, zatímco kvalita měření je klasicky posuzována pomocí konceptů reliability a validity. Rozumějme reliabilitou přesnost, konzistentnost měření, tj. schopnost dosáhnout

---

<sup>1</sup> Práce na tomto textu byly podpořeny Grantovou agenturou Akademie věd ČR (grant číslo S7028204). Rád bych poděkoval zejména Jindřichovi Krejčímu za pomoc při psaní a úpravách dřívějších verzí tohoto textu a anonymním recenzentům Sociologického časopisu za cenné rady a připomínky k předchozím verzím textu, zejména pak k aplikaci myšlenky bootstrapu v analýze. Odpovědnost za případné zbylé chyby a nepřesnosti samozřejmě náleží pouze autorovi. Pracovní verze textu byla prezentována na semináři Sociologického ústavu AV ČR v Praze 6. března 2003. Části tohoto článku byly již dříve využity v pracovním textu, který jsem připravil společně s Tomášem Lebedou [Kreidl, Lebeda 2003]. Připomínky recenzentů této dřívější verze, Hynka Jeřábka a Ladislava Cabady, byly využity rovněž při práci na tomto článku.

\* Veškerou korespondenci pošlete na adresu: PhDr. Martin Kreidl, M.A., Ph.D., Sociologický ústav AV ČR, Jilská 1, 110 00 Praha 1, e-mail: martin.kreidl@soc.cas.cz.

stejného výsledku měření v případech, že se stav pozorovaného předmětu nezměnil, a validitou schopnost měřit koncept, který skutečně zamýšlíme měřit [viz např. Řehák 1998a; Schutt 2004].

Důležitost kvalitního měření v poslední době přitáhla vlnu zájmu sociálních vědců. Přestože klasický model měření má své základy už v monografii Lorda a Novicka [1968], širšího metodologického uplatnění a popularizace se tématu kvality měření dostalo až v osmdesátých a především devadesátých letech minulého století [např. Alwin 1989; Lyberg et al. 1997; Saris, Meurs 1990; Scherpenzeel 1995],<sup>2</sup> nicméně zájem o kvalitu měření a dat neustoupil ani v novém tisíciletí [např. Biemer, Lyberg 2003]. Do českého prostředí uvádí teorii měření Jeřábek [1992], empirické nástroje pro zjištění kvality měření potom série článků Řeháka a jeho kolegů [Řehák 1998a, 1998b; Řehák, Bártová, Hamanová 1998].

Zájem o měření reliability, ověření validity měřicích instrumentů a jejich implikace pro praktický výzkum narůstá nejen v sociologii, ale i v příbuzných sociálních vědách. Například v sociální demografii Wu, Martin a Long [1999] zkoumali reliability retrospektivně zjišťovaných údajů o plodnosti a sexuálním chování; v sociální psychologii např. Lynch [2000], Rees, Hardy, Ingledew a Evans [2000], nebo Russell [1996] analyzovali validitu škál používaných k měření konceptů jako je pocit osamocení, úzkosti, a sociální opory; Smith a Patterson [1984] uvedli některé pokročilé nástroje k měření validity do kriminologie. Podobně ve výzkumu vzdělání např. Huang a Michael [2000] určují validitu měřicích nástrojů napříč socio-demografickými skupinami; Ward [1994] měří validitu psychologických konceptů, jako je potřeba kontroly a excelence; a Reed-Ashcraft, Kirk a Fraser [2001] zkoumají validitu škál měřících kvalitu rodinného prostředí – nástroje běžného v praxi sociální práce. Podobné příklady bychom ale našli i v medicíně a výzkumu veřejného zdraví [Bjorner, Kristensen 1999; Sukhwinder, Kuttalaliangam, Seneviratna, Orrell 1999].

Možné zdroje nekvalitního měření lze rozdělit do dvou zásadně odlišných skupin: *zdrojem chyb měření je buď samotný měřicí nástroj, nebo je to kontext měření* (srov. podrobný rozbor nabídnutý Biemerem a Lybergem [Biemer, Lyberg 2003: kap. 4–7]). V sociologii je nejběžnějším měřicím instrumentem dotazník, respektive jeho jednotlivé otázky. V případě jednotlivých otázek mají vliv na spolehlivost měření jednak znění otázky, nabídka odpovědí, jejich obsah a počet kategorií (sudý vs. lichý), stejně jako explicitně uvedená možnost „nevím“ a použití karet nebo jiných vizuálních pomůcek. Mezi vlivy kontextu patří design a délka dotazníku, pořadí otázek v dotazníku, způsob distribuce dotazníků (např. poštou, ve skupině, na internetu, emailem, nebo přímý rozhovor s tazatelem), trénink tazatelů, jejich vystupování, neutralita při kladení otázek, důvěryhodnost a zdrženlivost v reakcích na odpovědi. Předpoklady kvalitního měření také stojí na vzájemné neznalosti tazatele a respondenta, na pravidle uskutečnění pouze jediného rozhovoru v domácnosti a vylou-

<sup>2</sup> Přehled některých dalších klíčových prací z tohoto období tzv. „revoluce kvality“ uvádí např. Biemer a Lyberg [2003: 12–13].

čení opakovaných dotazování jedné osoby [např. Alwin 1989; Biemer, Lyberg 2003; Lyberg et al. 1997; Řehák 1998a; Tourangeau, Singer, Presser 2003].

V tomto textu nejdříve představím jeden ze základních instrumentů používaných k měření konstruktové validity škál, tzv. MIMIC model, a poté tento nástroj použiji k odhadu konstruktové validity škály pravo-levé politické orientace respondentů. Poté porovnáím kvalitu škály v datech sebraných třemi hlavními českými výzkumnými agenturami: CVVM, STEM a TNS-Factum. Vycházím z předpokladu, že v případě, že použijeme v různých měřicích kontextech stejný měřicí instrument, může být jediným zdrojem rozdílné kvality měření právě jen odlišný kontext. Na základě této metodologie lze odhalit rozdíly ve kvalitě měření mezi agenturami, ale nelze a priori určit, která agentura dosahuje lepších výsledků. Seřadit agentury podle výsledků lze lépe s použitím jiných nástrojů, například přístupů založených na zkoumání prediktivní validity [Lebeda 2003].

## 2. Metody měření reliability a validity

Existuje celá řada metod, jak měřit reliabilitu a validitu měření. Jak už jsem uvedl, reliabilita je schopnost měřicího nástroje dosáhnout stejného výsledku měření, pokud se nezměnil stav měřeného objektu. Reliabilita je podmínkou nutnou, nicméně nepostačující k dosažení validity. Nástroje k měření reliability jsou nicméně poněkud snáze dostupné než nástroje nutné pro odhad validity, a proto se často v empirických pracích testuje pouze reliabilita měřicích instrumentů.

Mezi základní čtveřici nástrojů ke zhodnocení reliability patří *metoda v čase opakovaného měření* (například tzv. „Quasi-simplex model“ [viz Řehák 1998b]), která je založená na předpokladu, že se stav objektu v určitém časovém intervalu nemění, a proto by opakovaná měření měla dospět ke stejnému výsledku. Druhou poměrně rozšířenou metodou měření reliability je *mezi-položková reliabilita*, neboli konzistence v odpovědích na baterii otázek určovaná například pomocí indexu Cronbachovo alfa. Další běžnou metodou je reliabilita zjišťovaná pomocí *alternativních forem* jedné otázky, která se liší například formulací úvodního stimulu, množstvím nebo řazením nabízených odpovědí, použitím karet a podobně. Alternativní verze téže otázky se buď mohou objevit v tomtéž dotazníku, kde na ně odpovídají titíž respondenti, nebo je každá verze nabídnuta jen části respondentů v tzv. *split-halves* designu [Schutt 2004]. Konečně poslední rozšířenou metodou zjišťování reliability je *porovnání více pozorovatelů*, kteří hodnotí stejnou situaci, nebo věc. Tato metoda je často používána např. ve školách, kdy je každá zkouška hodnocena nezávisle více než jedním hodnotitelem s použitím stejného nástroje, např. hodnotících instrukcí. Čím vyšší je mezi hodnotiteli shoda, tím lepší je reliabilita použitého nástroje.

Nástroje a postupy k měření validity jsou obecně vzácnější. Asi nejběžnější, nicméně také nejsnazší a nejméně uspokojivá je metoda zjevné validizace. *Zjevná validita* je založena na intuitivním předpokladu, že výsledek měření skutečně odráží výzkumníkův koncept. Přístupy spoléhající na zjevnou validizaci jsou zpravidla

neuspokojivé, Disman [1993] je dokonce přesvědčen, že zjevná validizace je jen eufemismem pro situace, kdy kontrola validity vůbec nebyla provedena. Pojem *obsahové validity* se zaměřuje na úplnost významové domény jevu, který zkoumáme. Jak zjevná, tak obsahová validita nemohou být zkoumány jinak než odkazem na existující literaturu, teorii daného jevu, empirický výzkum a mínění expertů.

Třetí ze skupiny nástrojů k určení validity, *kritériální validita*, je naproti tomu objektivněji měřitelná, a to porovnáním s nějakým objektivním, již validizovaným kritériem buď souběžně (tzv. *souběžná validita*), nebo prospektivně (*prediktivní validita*). Například volební prognóza může být validizována skutečnými volebními výsledky. Slabinou přístupů založených na kritériální validitě je absence spolehlivě validizovaných kritérií.

Posledním, čtvrtým přístupem k určení validity je *konstruktová validita*. Konstruktová validita je doložena, pokud je v datech zjištěn takový vztah mezi daným indikátorem a dalšími proměnnými, jaký bychom a priori očekávali na základě teorie. Konvergentní validity je dosaženo, pokud je jeden indikátor zkoumaného konceptu asociován s dalšími, nicméně odlišnými indikátory téhož konceptu. Diskriminační validity je naproti tomu dosaženo v případě, že hodnoty proměnné měřící daný koncept nejsou příliš silně asociovány s měřením odlišných, nicméně příbuzných konceptů. Základním nástrojem k měření konvergentní a diskriminační validity jsou explorační a konfirmační faktorová analýza a jejich různé aspekty (např. Bell, Lee 2002; Campbell, Arthur 1997; Jones et al. 2002; Lynch 2000; Russell 1996; Zhu 2000).

Posledním, a teoreticky i metodologicky nejnáročnějším přístupem k měření konstruktové validity je potom tzv. MIMIC model, který umísťuje zkoumaný měřicí nástroj, zpravidla vícepoložkovou škálu, do kontextu jak konceptuálně příbuzných tak odlišných proměnných, přesně specifikuje možné kauzální vazby mezi proměnnými a kontrastuje v datech zjištěný vzorec vztahů s teorií. Pokud jsou v datech zjištěny vztahy odpovídající teorii, je toto zjištění bráno jako potvrzení konstruktové validity zkoumané proměnné [např. Bagozzi 1978; Edmundson, Koch 1993; Jöreskog, Sörbom 1975; Ruble, Stout 1990; Williams 1994].

### 3. Kvalita měření volebních preferencí: základní výzkumná otázka

Zatímco otázky kvality měření jsou neustále v centru pozornosti empirických sociálních věd, stojí většinou celkem pochopitelně mimo zájem veřejnosti. Nicméně stejně jako v určitých, zejména předvolebních, obdobích periodicky roste zájem o mapování postojů a chování lidí, vyvstává se stejnou periodicitou i zájem o kvalitu měření. Například před volbami do Poslanecké sněmovny v roce 2002 se někteří novináři snažili vysvětlit systematicky rozdílné výsledky výzkumů veřejného mínění, které nabízely jednotlivé agentury, zejména nejčastěji citované CVVM, STEM a TNS Factum. Tím se do středu zájmu dostaly i odlišné metody měření volebních preferencí. Například Hospodářské noviny 19. dubna 2002 přebírají zprávu

ČTK, která konstatuje rozdíly změřených volebních preferencí mezi agenturami [čtk 2002]. Lipold [2002] poukazuje na inherentně danou nepřesnost volebních odhadů a Šídlo [2002] v podobně laděném článku objasňuje poměrně podrobně i vznik a metodologii odhadů volebních preferencí a volebních prognóz.

Autor skrytý pod redakční značkou *luk* [2002] se v Právu dostává nejbližšímu tématu. Autor spekuluje o tom, že odlišné volební preference jsou důsledkem odlišných dotazovacích metod, které jednotlivé agentury používají. Konečně Černý [2002], v té době mimo jiné i zaměstnanec TNS-Factum, odhaduje, že výsledek volební prognózy závisí i na kontextu konkrétního dotazníku a tématech dalších otázek, které tazatel respondentovi ve výzkumu klade. Do debaty o kvalitě volebních preferencí se „vysvětlujícími“ texty zapojila i Seidlová [2001, 2002], v té době jinak ředitelka CVVM.

Tento text má dvojí cíl. Jednak představuje české sociologii jeden z nejrozšířenějších metodologických nástrojů k měření konstruktové validity, tzv. MIMIC model (viz níže), a zároveň jej aplikuje na problém kvality měření českých výzkumných agentur. Tento druhý cíl tak v odborné rovině doplňuje mediální zájem o kvalitu sociologického měření. Česká sociologie má, bohužel až na velmi řídké výjimky, v empirické analýze běžně používaných výzkumných metod zatím [ale viz např. Řehák 1998a, 1998b; Řehák, Bártová, Hamanová 1998] nesplacený dluh k odborné i laické veřejnosti.

#### 4. Použitá data

Používám data z metodologického šetření, které bylo provedeno v předvolebních měsících roku 2002. Do omnibusových šetření, která byla v té době v STEM, TNS-Factum a CVVM využívána i pro měření volebních preferencí, byl přidán soubor několika standardních otázek, z nichž v tomto textu používám pětipoložkovou baterii k měření postavení respondenta na tzv. „objektivní“ škále pravolevé politické orientace (viz níže), subjektivní politickou orientaci měřenou na sedmibodové pravolevé škále jdoucí od 1 (levice) až po 7 (právice), desetibodovou škálu subjektivního sociálního statusu (1 nejvyšší status, 10 nejnižší status)<sup>3</sup> a vzdělání měřené čtyřstupňovou škálou (ZŠ, vyučení, maturita, vysoká škola).

Protože znění otázek, použité měřicí škály, další instrumenty i termíny položení otázek (květen 2002) byly ve všech kontextech (tj. agenturách) v maximální dosažitelné míře shodné, lze jediné zjištěné rozdíly v kvalitě měření připsat rozdílné kvalitě práce jednotlivých agentur. Zdrojem nereliabilního měření může být, jak už bylo řečeno, kvalita tazatelské sítě, tj. kvalita tréninku a školení tazatelů, sociální skladba sítě, neutralita a důvěryhodnost tazatelů, přesné dodržení pokynů k výzkumu, neopakování rozhovorů ve stále stejných domácnostech, provádění rozhovorů pouze respondenty, které tazatel před zahájením výzkumu neznal apod. Zdrojem nekvalitní

---

<sup>3</sup> Přesné znění těchto otázek je uvedeno v Příloze.

ho měření může ovšem být i nekvalitní *zpracování dat v samotné agentuře*, kódování a kontroly dat [srov. Biemer, Lyberg 2003]. Kvalitu měření ovlivňuje i celková délka rozhovoru (tj. počet dalších otázek v omnibusovém šetření), nebo kontext dalších otázek ve výzkumu, neboť ty mohou vést ke krátkodobé změně postoje, mohou navádět k jisté odpovědi, mohou iniciovat proces tvorby nového postoje nebo mohou zvyšovat nebo snižovat ochotu respondentů k účasti ve výzkumu.<sup>4</sup>

Na první pohled nabízejí ve svých měsíčních omnibusech všechny tři agentury velmi podobné podmínky. Základní metodou je standardizovaný rozhovor vedený v osobním *face-to-face* kontaktu s respondentem, někdy s pomocí karet. Také v otázkách našeho metodologického experimentu byly použity karty. Respondentovy odpovědi jsou tazatelem zaznamenávány do předtištěných políček dotazníku.

Všechny tři firmy standardně v roce 2002, kdy byla data sebrána, používaly metodu kvótního výběru. Agentura TNS Factum použila 5 kvótních znaků: regiony-NUTS3, velikost místa bydliště, věk (5 skupin), pohlaví a vzdělání (4 kategorie); agentura nám s datovým souborem neposkytla technickou zprávu, proto nám nejsou známy podrobnosti o členění věkových skupin a velikostní kategorie místa bydliště v kvótním předpisu. I STEM používá kvótní předpis s pěti znaky: regiony (definované jako 8 regionů NUTS2), velikost místa bydliště (6 kategorií), pohlaví, věk (4 kategorie) a vzdělání (4 kategorie). CVVM používá 5 kvótních znaků: region (8 starých krajů), velikost místa bydliště (6 kategorií), pohlaví, věk (4 kategorie) a vzdělání (rovněž 4 skupiny).

STEM ani CVVM klientům ve svých výzkumech nenabízejí váhy, zatímco TNS Factum nabízí hned dva alternativní způsoby vážení. Prvním je sociodemografická váha, která převažuje výsledky znovu podle základních kvótních znaků. Druhá váha upravuje výsledky podle regionu, velikosti místa bydliště a volebního chování v roce 1998. Podrobnosti o tvorbě vah nám s datovým souborem nebyly poskytnuty, což je poněkud překvapivý postup. Váhy nejsou v analýze dat TNS Factum použity, neboť software použitý pro analýzu jejich použití neumožňuje (viz níže). Popisné statistiky dat ze souboru TNS Factum nabízím jak nevážené, tak s použitím obou dodaných vah. Vliv vah na distribuci hodnot jednotlivých proměnných je ovšem minimální, zpravidla v řádu jednoho procenta (viz níže).

Výzkum STEM proběhl v prvním květnovém týdnu roku 2002 na celkem 2080 respondentech. Výzkum CVVM proběhl ve stejném termínu na celkovém sou-

---

<sup>4</sup> Vliv těchto kontextuálních efektů nelze nikdy vyloučit, ani standardizovat, byť by to např. v metodologickém šetření tohoto druhu bylo vhodné. Protože kontexty otázky na volební preference se v omnibusových šetřeních mění od měsíce k měsíci, bude se pravděpodobně s nimi měnit i samotná kvalita dat. Proto by bylo vhodné experiment podobný našemu opakovat častěji a získat tak stabilnější a robustnější výsledky. Zodpovědnost za tyto kontextuální efekty v případě omnibusových šetření náleží ovšem agentuře, která má složit dotazník tak, aby byly reaktivní efekty mezi otázkami minimalizovány, popřípadě odmítnout do omnibusu zakázky, díky nimž by došlo ke zkrácení celého šetření nebo k jeho prodloužení nad únosnou délku. Je ovšem vhodné, aby se zainteresovaní uživatelé výsledků sami o obsah omnibusu zajímali.

boru 1083 respondentů a TNS Factum se v týdnu od 13. 5. 2002 dotázal 1014 osob. Protože jednotlivé agentury používají odlišné věkové vymezení základní populace (15 let a více vs. 18 let a více<sup>5</sup>), rozhodl jsem se věkové rozložení sjednotit a do analýzy zahrnout pouze respondenty ve věku 18 let a více (972 respondentů TNS, 2077 respondentů STEM a 991 respondentů CVVM). Po vyloučení respondentů s chybějícími daty zbylo pro samotnou analýzu 754 respondentů TNS Factum, 1817 respondentů STEM a 710 respondentů CVVM.

Protože při porovnávání modelů (viz níže) používám testové statistiky na bázi chí-kvadrátu<sup>6</sup>, jejichž velikost závisí na velikosti vzorku, nabízím výsledky analýzy v několika podobách: výsledky založené na kompletním souboru všech dostupných respondentů všech agentur; výsledky založené na všech dostupných respondentech každé agentury zvlášť; a nakonec výsledky založené na re-samplingu respondentů každé agentury (podrobněji o myšlence re-samplingu a bootstrapu, viz [Efron, Tibshirani 1993]. Mezi respondenty každé agentury provádím 100 opakovaných náhodných výběrů s opakováním o standardizované velikosti. Jako standardní velikost výběrových souborů pro re-sampling jsem zvolil velikost nejmenšího datového souboru (CVVM, 710 respondentů). Test validity opakuji na každém takto získaném datovém souboru a uvádím souhrnné statistiky těchto opakování (medián, směrodatnou odchylku testového kritéria, 95% interval spolehlivosti) a ukáží distribuci testového kritéria pomocí krabicového grafu. Aplikací myšlenky bootstrapu v tomto kontextu jednak optimálně využiji informace o všech respondentech, tak standardizují velikost jednotlivých datových souborů a přitom není výsledek chí-kvadrátového testu ovlivněn odlišnou původní velikostí jednotlivých datových souborů, což by například výběr pomocí náhodné procedury bez opakování neumožnil [srov. Efron, Tibshirani 1993].

## 5. Základní charakteristiky MIMIC modelu

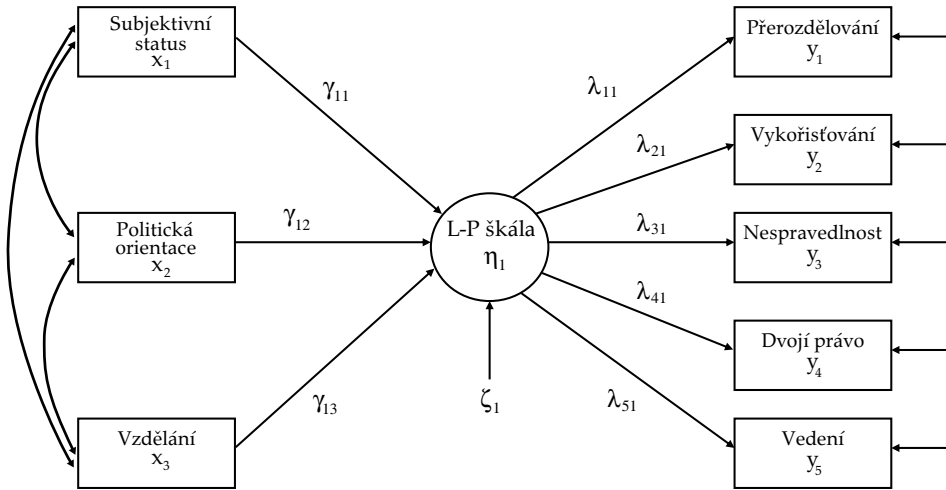
MIMIC model (MIMIC je zkratka za **M**ultiple **I**ndicators – **M**ulti**P**le **C**auses) je jedním z nejběžnějších existujících nástrojů pro měření konstruktové validity měřících instrumentů. Podstatou MIMIC modelu je testování validity vícepoložkové škály, která má měřit jeden latentní konstrukt, v nějakém teoreticky relevantním kontextu dalších proměnných. Nejednoduší verze MIMIC modelu je situace s jednou latentní škálou, která je měřena pomocí několika indikátorů (označovaných v tradici strukturního modelování zpravidla  $y$ ) a je zároveň ovlivňována několika přímo pozorovanými proměnnými [ $x$ , viz Jöreskog, Sörbom 1989]. Takový nejjednodušší mo-

---

<sup>5</sup> Přestože STEM udává základní populaci výzkumu v rozmezí 18 a více let, našli jsme v souboru 3 respondenty, kteří udávali nižší věk.

<sup>6</sup> Příkaz *testnl* v programu STATA testuje nelineární omezení v rámci odhadnutého modelu pomocí Waldova testu (je to obdoba příkazu *test*, který nicméně testuje pouze lineární omezení); v tomto konkrétním případě pomocí klasického chí-kvadrátu [podrobnosti viz STATA Corp. 2003b].

Obrázek 1. MIMIC model k ověření konstruktové validity škály pravolevé politické orientace



del používají například Hodge a Treiman [1968], kteří testují závislost latentní škály sociální participace (měřené pomocí tří indikátorů) na třech různých žebříčcích sociálního statusu (příjem, zaměstnání, vzdělání).<sup>7</sup>

V našem konkrétním případě je výchozí MIMIC model prezentován na Obrázku 1. Model obsahuje 8 přímo měřených proměnných a jednu latentní proměnnou ( $\eta$ ).<sup>8</sup> Z osmi přímo měřených proměnných jsou 3 nezávislé ( $x$ ) a 5 je závislých ( $y$ ). Závislé proměnné jsou položky používané standardně pro měření tzv. objektivní pravolevé orientace [viz např. Evans, Heath, Lalljee 1996; Matějů, Vlachová 2000]. Jde o souhlas (vyjádřený na čtyřbodové škále, kde 1 je „rozhodně souhlasím“, 2 znamená „spíše souhlasím“, 3 „spíše nesouhlasím“ a 4 „rozhodně nesouhlasím“) s následující pětici tvrzení (v závorkách jsou uvedeny zkrácené výrazy použité v obrázcích):

1. Vláda by měla přerozdělovat příjem od těch, kteří jsou na tom lépe, těm, kteří jsou na tom hůře. (přerozdělování)

<sup>7</sup> MIMIC model byl v posledních desetiletích použit v řadě různých disciplín, mimo jiné i v sociologii [Blake, Richardson, Bhattacharya 1991; Bohrnstedt 1977; Inverarity 1976; Warren, Hauser 1997]; kriminologii [např. Piquero, Macintosh, Hickman 2002; Smith, Patterson 1984]; ekonomii [Giles 1999]; politické vědě [Stapleton 1978]; výzkumu veřejného zdraví a medicíny [např. Bjorner, Kristensen 1999; Gallo, Anthony, Muthen 1994; Mast, MacNeill, Lichtenberg 2002]; sociální psychologii [např. Lynch 2000] i sociální práci [Reed-Ashcraft, Kirk, Fraser 2001].

<sup>8</sup> Při popisu MIMIC modelu používám notaci, která je běžná ve strukturním modelování, například v programu LISREL [Jöreskog, Sörbom 1989].



2. Vlastníci velkých podniků bohatnou na úkor dělníků. (*vykořisťování*)
3. Obvyčejní pracující lidé nedostávají spravedlivý podíl z národního bohatství. (*nespravedlnost*)
4. Existuje jedno právo pro bohaté a druhé pro chudé. (*dvojí právo*)
5. Vedení podniku se bude vždy snažit vyzrát na zaměstnance, pokud dostane šanci. (*vedení*)

Těchto pět postojů je v souladu s teorií bráno jako pět indikátorů latentního konstruktů ( $\eta_1$  v Obrázku 1). Zároveň z teorie determinant politických postojů a její aplikace na českou společnost odvozují, že měřená pozice na „objektivní“ pravo-levé škále závisí na třech exogenních, nezávislých proměnných: objektivním sociálním statusu, který je zde přiblížen vzděláním respondenta, na subjektivním sociálním postavení respondenta a na subjektivní pravo-levé politické identifikaci [viz např. Kreidl 1998; Matějů, Vlachová 2000].

Model dále obsahuje dva druhy chyb, chybu v rovnicích ( $\zeta_i$ ) a chyby měření [ $\varepsilon_i$ , viz Jöreskog, Sörbom 1989]. Vztahy mezi proměnnými v MIMIC modelu mohou být vyjádřeny dvěma rovnicemi, rovnicí strukturální („*structural equation*“) a rovnicí pro model měření („*measurement equation*“).

Strukturální rovnici můžeme zapsat jako:

$$\eta = \Gamma X + \zeta, \quad (1)$$

kde  $\eta$  je sloupcový vektor (obecně)  $m$  latentních proměnných,  $X$  je sloupcový vektor  $n$  pozorovaných nezávislých proměnných,  $\Gamma$  je maticí ( $m \times n$ ) parametrů, která vyjadřuje efekty nezávislých proměnných ( $X$ ) na vektor  $m$  latentních závislých proměnných ( $\eta$ ) a  $\zeta$  je sloupcový vektor (obecně)  $m$  disturbancí v měření  $m$  latentních konstruktů.

Model měření je potom:

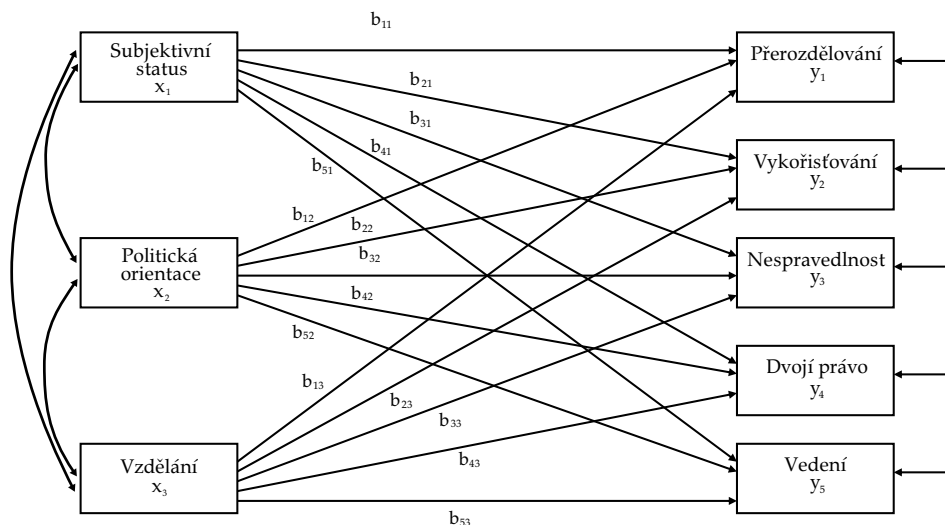
$$Y = \Lambda \eta + \varepsilon, \quad (2)$$

kde  $Y$  je sloupcový vektor  $p$  pozorovaných indikátorů,  $\eta$  je sloupcový vektor (obecně)  $m$  latentních proměnných,  $\Lambda$  je ( $p \times m$ ) parametrická matice obsahující faktorové zátěže a  $\varepsilon$  je sloupcový vektor  $p$  chyb měření u indikátorových proměnných.

## 6. Měření validity škál pomocí MIMIC modelu

Metodologicky o modelech typu MIMIC poprvé pojednávají pravděpodobně Hauser a Goldberger [1971]. Jöreskog a Goldberger [1975] dále ukazují, že se lze na MIMIC model dívat jako na mnohorozměrný regresní model s určitou soustavou omezení, tzv. poměrových omezení („*proportionality restrictions*“). Tato poměrová omezení hraje klíčovou roli při použití MIMIC modelu ke zhodnocení konstruktové validity škál,

**Obrázek 2. Neomezený regresní model předpokládající přímý vliv nezávislých proměnných na indikátory pravolevé škály<sup>9</sup>**



a proto v této sekci textu jednak vysvětlím, co poměrová omezení znamenají, tak princip jejich testování.

Vezměme mnohorozměrný regresní model závislosti pěti proměnných (proměnné „Přerozdělování“, „Vykořisťování“, „Nespravedlnost“, „Dvojí právo“, „Vedení“) na třech nezávislých proměnných („Subjektivní status“, „Vzdělání“, a „Politická orientace“), jak je znázorněno na Obrázku 2. Tento model, na rozdíl o MIMIC modelu (Obrázek 1) nepředpokládá existenci latentní škály indikované pětici přímo měřených proměnných. Na rozdíl od MIMIC modelu nepředpokládá tento regresní model žádná škálovací omezení a nechává každou závislou proměnnou reagovat individuálně na změny v hodnotách nezávislých proměnných, tj. relativní síla efektů jednotlivých nezávislých proměnných na závislou proměnnou se může v každé z rovnic libovolně měnit. Každá z přímých šipek v obrázku 2 odpovídá jednomu z celkem 15 regresních koeficientů v modelu. Tento model můžeme zapsat jako:

$$\begin{aligned}
 y_1 &= b_{11} x_1 + b_{12} x_2 + b_{13} x_3 + \varepsilon_1 \\
 y_2 &= b_{21} x_1 + b_{22} x_2 + b_{23} x_3 + \varepsilon_2 \\
 y_3 &= b_{31} x_1 + b_{32} x_2 + b_{33} x_3 + \varepsilon_3 \\
 y_4 &= b_{41} x_1 + b_{42} x_2 + b_{43} x_3 + \varepsilon_4 \\
 y_5 &= b_{51} x_1 + b_{52} x_2 + b_{53} x_3 + \varepsilon_5
 \end{aligned} \tag{3}$$

<sup>9</sup> Z důvodu zachování přehlednosti nejsou v obrázku zakresleny korelace mezi chybami měření závislých proměnných (vektor Y).

V MIMIC modelu, kde je  $y_1$  referenčním indikátorem pro  $\eta$  a tudíž je  $\lambda_{11} = 1$ , platí:

$$\begin{aligned} y_1 &= \eta + \varepsilon_1 \\ y_2 &= \lambda_{21} \eta + \varepsilon_2 \\ y_3 &= \lambda_{31} \eta + \varepsilon_3 \\ y_4 &= \lambda_{41} \eta + \varepsilon_4 \\ y_5 &= \lambda_{51} \eta + \varepsilon_5 \end{aligned} \quad (4)$$

a zároveň:

$$\eta = \gamma_{11} x_1 + \gamma_{12} x_2 + \gamma_{13} x_3 + \zeta \quad (5)$$

MIMIC model zapsaný v rovnicích (4) a (5) je zúžením obecného regresního modelu (3) (jde o „nested models“). MIMIC model konkrétně předpokládá následující restrikce uvnitř regresního modelu:

$$\begin{array}{lll} b_{11} = \gamma_{11} & b_{12} = \gamma_{12} & b_{13} = \gamma_{13} \\ b_{21} = \gamma_{11} \lambda_{21} & b_{22} = \gamma_{12} \lambda_{21} & b_{23} = \gamma_{13} \lambda_{21} \\ b_{31} = \gamma_{11} \lambda_{31} & b_{32} = \gamma_{12} \lambda_{31} & b_{33} = \gamma_{13} \lambda_{31} \\ b_{41} = \gamma_{11} \lambda_{41} & b_{42} = \gamma_{12} \lambda_{41} & b_{43} = \gamma_{13} \lambda_{41} \\ b_{51} = \gamma_{11} \lambda_{51} & b_{52} = \gamma_{12} \lambda_{51} & b_{53} = \gamma_{13} \lambda_{51} \end{array} \quad (6)$$

Z toho vyplývají následující restrikce:

$$b_{21}/b_{11} = b_{22}/b_{12} = b_{23}/b_{13} \quad (7)$$

$$b_{31}/b_{11} = b_{32}/b_{12} = b_{33}/b_{13} \quad (8)$$

$$b_{41}/b_{11} = b_{42}/b_{12} = b_{43}/b_{13} \quad (9)$$

$$b_{51}/b_{11} = b_{52}/b_{12} = b_{53}/b_{13} \quad (10)$$

Namísto 15 regresních koeficientů (viz rovnice (3)) má MIMIC model jen 7 koeficientů ( $\gamma$  a  $\lambda$  v rovnicích (4) a (5)), což vyplývá z uvedených 8 dodatečných omezení (viz (7), (8), (9) a (10)). Tato omezení jsou známa právě jako „poměrová omezení“ („proportionality restrictions“). Z poměrových omezení vyplývá, že relativní efekty nezávislých proměnných jsou stejné pro všechny položky v baterii. Např.  $x_1$  může mít větší efekt než  $x_2$ , ale platí, že relativní velikost efektů  $x_1$  a  $x_2$  je stejná pro  $y_1$ ,  $y_2$ ,  $y_3$ ,  $y_4$  i  $y_5$ .

Principem MIMIC modelu coby nástroje k určení konstruktové validity je statistický test nulové hypotézy ( $H_0$ ) o platnosti poměrových omezení pomocí chí-kvadrátového testu. Pokud poměrová omezení platí, respektive nulová hypotéza není na základě statis-

tického testu zamítnuta, je škála považována za konstruktově validní. Je potvrzen předpoklad, že *vliv měřených nezávislých proměnných na měřené závislé indikátory není přímý, ale je zprostředkován jednou latentní proměnnou, tj. škálou*. V našem konkrétním případě (viz rovnice (7), (8), (9), (10)) se bude jednat o chí-kvadrátový test s osmi stupni volnosti. Vyšší hodnota chí-kvadrátu pak, při standardizované velikosti vzorku, znamená větší odchylku od poměrových omezení; pokud hodnota chí-kvadrátu překročí kritickou hladinu, musíme zamítnout platnost poměrových omezení a škálu v daném kontextu prohlásit za konstruktově nevalidní.

V principu lze odhadnout parametry MIMIC modelu s použitím běžné algebry. Testovat platnost poměrových omezení lze například i s použitím běžných statistických balíků jako je STATA [STATA Corp. 2003a], který byl použit i pro výpočet zde uváděných výsledků.<sup>10</sup> Odhad všech parametrů modelu je pak možné získat jen speciálním softwarem pro strukturní modelování, jako je např. LISREL, EQS, nebo Mplus.

## 7. Popis dat a výsledků

Distribuce odpovědí na otázky tvořící pravo-levou postojovou škálu jsou velmi podobné napříč agenturami (Tabulka 1). Jen v několika málo případech najdeme významnější odchylky. Nejsilnější rozdíly najdeme mezi CVVM a TNS Factum v odpovědích na 3 otázky použité škály. Zatímco 23 % respondentů TNS Factum rozhodně souhlasilo s výrokem „Vláda by měla přerozdělovat příjem od těch, kteří jsou na tom lépe, těm, kteří jsou na tom hůře“, pouze 15 % respondentů CVVM vyjádřilo podobně silný souhlas. Podobně 40 % respondentů TNS Factum a pouze 33 % respondentů CVVM rozhodně souhlasilo, že „Vlastníci velkých podniků bohatnou na úkor dělníků.“ Konečně 38 % respondentů TNS Factum a pouze 31 % respondentů CVVM rozhodně souhlasilo, že „Obyčejní pracující lidé nedostávají spravedlivý podíl z národního bohatství.“ Ostatní rozdíly mezi agenturami jsou již méně významné.

Výsledky testů platnosti proporčních omezení v testovaném MIMIC modelu jsou představeny v Tabulce 2. Pro ilustraci uvádím výsledky testu na souhrnném vzorku všech respondentů všech agentur, na kompletním vzorku každé agentury a výsledky založené na standardizovaném vzorku respondentů každé agentury. Velikost těchto standardizovaných vzorků (710 respondentů) odpovídá počtu respondentů v nejmenším vzorku, který patří CVVM. Při standardizaci velikosti vzorků ovšem vyvstává otázka, jak efektivně využít informaci o respondentech, kteří nebyli do vzorku vybráni. Abych předešel ztrátě informace, provedl jsem mezi respondenty všech agentur sto opakovaných náhodných výběrů s opakováním o velikosti

---

<sup>10</sup> Použité datové soubory a potřebné příkazové soubory („do files“) použité v programu STATA ke všem zde prezentovaným výpočtům jsou volně k dispozici ke stažení a k replikaci výsledků na následující internetové stránce: [www.soc.cas.cz/kreidl\\_data](http://www.soc.cas.cz/kreidl_data). Tato stránka také obsahuje písemný návod na spuštění programů pro čtenáře, kteří běžně program STATA nepoužívají.

**Tabulka 1. Procentní distribuce proměnných tvořících škálu pravo-levé politické orientace podle agentury, předvolební omnibusová šetření z roku 2002**

Agentura	Proměnná				Celkem (N)
	<i>Přerozdělování</i>				
	Rozhodně souhlasím	Spíše souhlasím	Spíše nesouhlasím	Rozhodně nesouhlasím	
TNS Factum 1	23 %	32 %	27 %	18 %	100 % (754)
TNS Factum 2	24 %	32 %	27 %	17 %	100 % (743)
TNS Factum 3	22 %	31 %	28 %	18 %	99 % (738)
STEM	19 %	35 %	30 %	16 %	100 % (1817)
CVVM	15 %	36 %	32 %	16 %	100 % (710)
Celkem	19 %	34 %	30 %	17 %	100 % (3281)
	<i>Vykořisťování</i>				
	Rozhodně souhlasím	Spíše souhlasím	Spíše nesouhlasím	Rozhodně nesouhlasím	Celkem
TNS Factum 1	40 %	34 %	19 %	7 %	100 % (754)
TNS Factum 2	41 %	33 %	19 %	6 %	99 % (743)
TNS Factum 3	41 %	33 %	20 %	7 %	101 % (738)
STEM	37 %	37 %	19 %	7 %	100 % (1817)
CVVM	33 %	42 %	18 %	7 %	100 % (710)
Celkem	37 %	37 %	19 %	7 %	100 % (3281)
	<i>Nespravedlnost</i>				
	Rozhodně souhlasím	Spíše souhlasím	Spíše nesouhlasím	Rozhodně nesouhlasím	Celkem
TNS Factum 1	38 %	38 %	19 %	5 %	100 % (754)
TNS Factum 2	39 %	38 %	18 %	5 %	100 % (743)
TNS Factum 3	38 %	38 %	19 %	5 %	100 % (738)
STEM	34 %	41 %	20 %	5 %	100 % (1817)
CVVM	31 %	46 %	17 %	5 %	99 % (710)
Celkem	34 %	42 %	19 %	5 %	100 % (3281)
	<i>Dvojí právo</i>				
	Rozhodně souhlasím	Spíše souhlasím	Spíše nesouhlasím	Rozhodně nesouhlasím	Celkem
TNS Factum 1	38 %	33 %	20 %	9 %	100 % (754)
TNS Factum 2	38 %	33 %	20 %	8 %	99 % (743)
TNS Factum 3	37 %	34 %	21 %	9 %	101 % (738)
STEM	35 %	37 %	19 %	8 %	99 % (1817)
CVVM	35 %	36 %	19 %	9 %	99 % (710)
Celkem	36 %	36 %	20 %	9 %	101 % (3281)
	<i>Vedení</i>				
	Rozhodně souhlasím	Spíše souhlasím	Spíše nesouhlasím	Rozhodně nesouhlasím	Celkem
TNS Factum 1	38 %	39 %	18 %	5 %	100 % (754)
TNS Factum 2	39 %	38 %	18 %	5 %	100 % (743)
TNS Factum 3	38 %	39 %	19 %	5 %	101 % (738)
STEM	38 %	41 %	17 %	4 %	100 % (1817)
CVVM	36 %	46 %	15 %	3 %	100 % (710)
Celkem	38 %	42 %	17 %	4 %	101 % (3281)

Pozn.: U TNS Factum uvádíme tři výsledky – první řádek platí pro nevážená data, druhý pro data vážená „demografickou“ vahou a třetí je výsledek po aplikaci převážení podle regionu, velikosti místa bydliště a volebního chování v roce 1998. Všechny hodnoty jsou zaokrouhleny na jedno desetinné místo, v důsledku chyby ze zaokrouhlení není součet v každém řádku roven přesně 100 %.

**Tabulka 2. Test platnosti proporčních omezení v MIMIC modelu podle výzkumné agentury v roce 2002**

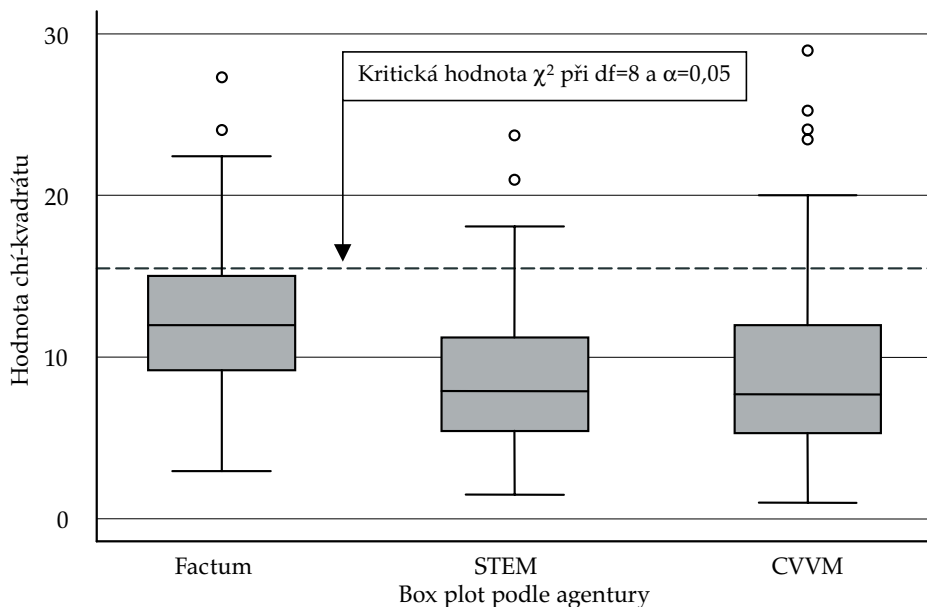
	$\chi^2$	s.d.	df	N	95 % C.I.
<b>A. bootstrapovaný vzorek*</b>					
CVVM	7,71	5,42	8	710	2,43; 24,64
STEM	7,89	4,41	8	710	1,73; 19,46
TNS	11,97	4,88	8	710	3,74; 23,20
<b>B. původní data</b>					
CVVM	4,59	--	8	710	--
STEM	12,52	--	8	1817	--
TNS	10,02	--	8	754	--
Všichni respondenti	9,05	--	8	3281	--

\* Bootstrapovaný vzorek znamená, že jsem v souboru respondentů každé agentury provedl 100 náhodných výběrů s opakováním ( $n = 710$ ) a v každém jsem vypočetl chí-kvadrátový test platnosti poměrových omezení. V tabulce prezentované hodnoty jsou střední hodnotou a směrodatnou odchylkou distribuce této stovky hodnot chí-kvadrátu. Konfidenční interval je založen na 2.5 a 97.5-tém percentilu v distribuci hodnot chí-kvadrátu, což je hrubé, nicméně přijatelné přiblížení skutečných hodnot [Efron, Tibshirani 1993: 5].

710 respondentů a na každém takto získaném souboru jsem otestoval platnost poměrových omezení pomocí chí-kvadrátového testu (idea *bootstrapu*, viz výše). Protože používám metodu náhodného výběru s vracením, můžu provádět výběr ze souboru každé agentury, a tedy i CVVM. V Tabulce 2 potom prezentuji střední hodnotu chí-kvadrátu v tomto souboru sta opakovaných testů v souboru dat každé agentury.

V panelu A Tabulky 2 jsou uvedeny výsledky na bootstrapovaných vzorcích a v panelu B na vzorcích nestandardizovaných. Podle výsledků v Tabulce 2 se zdá, že se data sebraná každou z agentur signifikantně neodlišují od předpokládaného MIMIC modelu. Chí-kvadrátový test proporčních omezení v nestandardizovaných datech CVVM dosahuje hodnoty 4.59 ( $df=8$ ), datech STEM 12.52 a u TNS Factum 10.02. Žádná z těchto hodnot nepřesahuje kritickou hodnotu chí-kvadrátu s 8 stupni volnosti (15.5) a proto u žádné agentury nemůžeme zamítnout hypotézu o platnosti proporčních omezení. *Použitá škála pravo-levé politické orientace se tedy jeví uspokojivě validní bez ohledu na kontext vzniku dat.* Podle velikosti dosaženého testového kritéria by se mohlo zdát, že je škála v datech CVVM validnější, než v datech TNS Faktum a že je u obou agentur dosaženo větší validity než u STEM. Tato interpretace by však byla chybná, jednak proto, že jde pouze o testové kritérium, a také proto, že je jeho hodnota ovlivněna i velikostí datového souboru. Proto je vhodné porovnat i hodnoty testového kritéria na bootstrapovaných vzorcích. Na standardizovaných vzorcích dosahuje střední hodnota chí-kvadrátu 7.71 v datech CVVM, 7.89 v datech STEM a 11.97 v datech TNS Factum. V datech každé agentury také nale-

**Graf 1. Krabicový graf hodnot chí-kvadrátu testu platnosti poměrových omezení ve stovce náhodných výběrů s opakováním z respondentů každé agentury**



Pozn.:

1. Každý box plot znázorňuje hodnoty chí-kvadrátu pro 100 nezávislých náhodných výběrů s opakováním ze souborů všech respondentů dané firmy. Velikost každého tohoto výběru je rovna 710 respondentům bez ohledu na původní velikost datového souboru dané agentury.
2. Každý box plot je vytvořen podle zásad navržených Clevelandem [1993: 25–27] a znázorňuje hodnotu mediánu a 25-ho a 75-ho percentilu („Interquartile range“ IQR), přilehlá pozorování v rozmezí  $x_{25}/x_{75} + 1.5 \cdot \text{IQR}$ . Vzdálená pozorování, tj. pozorování mimo oblast přilehlých pozorování, jsou v grafu znázorněna samostatnými body.
3. Čerchovaná horizontální čára ukazuje kritickou hodnotu  $\chi^2$  při osmi stupních volnosti a hladině významnosti 0.05.
4. Hodnota  $\chi^2$  na nestandardizovaném souboru 1817 respondentů STEM je 12.52, na nestandardizovaném souboru 754 respondentů TNS Factum 10.02 a na nestandardizovaném souboru 710 respondentů CVVM 4.59.

neme přibližně stejnou varianci (viz Tabulku 2).<sup>11</sup> Z toho vyplývá, že testovaná škála dosahuje ve všech kontextech v zásadě stejné úrovně validity.<sup>12</sup>

<sup>11</sup> Graf 1 ukazuje podrobněji distribuci hodnot chí-kvadrátu v opakovaných náhodných podvýběrech všech agentur. V grafu je pro orientaci uvedena i kritická hodnota chí-kvadrátu s osmi stupni volnosti.

<sup>12</sup> Bootstrapová metoda ovšem také ukazuje, že u každé agentury dochází v určitém procentu výběrů k překročení testového kritéria (viz Tabulku 2), což může zpochybnit dostatečnou

Logicky se nyní musíme zeptat, která z agentur dosáhla nejlepších výsledků. Tvrdit a priori, že agentura, která dosáhla nejmenší hodnoty testového kritéria v testu proporčních omezení, odvádí při sběru dat nejprofesionálnější práci, by bylo jen málo podložené. Je zřejmé, že určitý druh nedokonalé práce agentur (například podvádění tazatelů při vyplňování dotazníků) může teoreticky vést k takové struktuře dat, která se v MIMIC modelu projeví jako zvýšení validity. Jinými slovy, faktory specifické pro jednotlivé agentury mohou změřenou validitu jak zvyšovat, tak snižovat oproti skutečné validitě, protože nevíme, zda a v jaké míře byla v české populaci přítomna předpokládaná kovarianční struktura. Protože neznáme skutečnou validitu použité škály, je bez dodatečných znalostí nemožné klasifikovat agentury podle výkonu.

## 8. Závěrečné shrnutí

Jednou ze dvou motivačních otázek tohoto textu bylo ukázat, zda lze doložit rozdíly v kvalitě měření mezi agenturami na základě zhodnocení validity standardizované škály. Vycházel jsem při tom z předpokladu, že existují v zásadě pouze dva možné zdroje chybného měření: měřicí nástroj a kontext měření. V případě že každá z agentur ve svém výzkumu použila stejný měřicí nástroj, mohou být jediným zdrojem chybového měření jevy specifické pro každou agenturu: kvalita tazatelské sítě a její proškolení a profesionalita, dodržování výběrových postupů, schopnost sestavit otázky omnibusového dotazníku způsobem, který nepovede ke zkreslení odpovědí, kvalitní práce s daty a podobně. Ukázalo se nicméně, že *podle zvoleného kritéria dosahují všechny agentury velmi podobných výsledků a nelze ukázat na data žádná z nich jako na data výrazně odlišná* [srov. Lebeda 2003], a tudíž potenciálně méně kvalitní.

I kdybychom však zjistili nějaké významné zdroje variance v měření volebních preferencí mezi agenturami, mohli bychom o jejich přesných zdrojích spíše jen spekulovat, protože agentury řadu důležitých údajů nezveřejňují. Zpravidla se nedozvídáme, kolik tazatelů se daného výzkumu zúčastnilo, ani kolik dalších výzkumů v dané době prováděli. Nedozvídáme se ani údaje o sociální skladbě tazatelské sítě, která je u kvótních výběrů obzvláště důležitá, ani o obměně tazatelů v čase (což souvisí s jejich zkušeností a rutinou jejich práce). Paradoxně tak musíme nahlížet na agenturu jako na černou skříňku, která na základě požadavků zadavatele produkuje nějaké výsledky, nenechává však nikoho nahlédnout do svého nitra, což ztěžuje práci s výsledky i jejich kritické zhodnocení, a přispívá tak k jejich zkreslenému mediálnímu zpracování. Tato skutečnost sama o sobě je dosti znepokojivá a to i přesto, že agentury dosáhly v našem testu velmi podobných výsledků.

---

validitu celé škály jako takové. V tomto textu jsem ovšem bootstrapovou metodu používal pro srovnání relativní validity škály v různých kontextech a celková validita škály se mi jeví jako uspokojivě doložená analýzou nestandardizovaných vzorků.



Je třeba mít na paměti, že zde uvedený test není a ani nemůže být komplexním zhodnocením kvality dat, která zmíněné agentury rutinně produkují [viz např. Krejčí 2004 pro přehled a aplikace dalších přístupů]. Jak již bylo řečeno v úvodu, závisí kvalita dat na mnoha okolnostech. V případě měření volebních preferencí a produkce volebních prognóz mimo jiné i na kvalitě výběru a použitých otázkách. Kvalita inference z výběru na základní populaci je dále závislá i na použité výběrové proceduře. Zatímco STEM i TNS Factum byly v roce 2002 schopny provádět i náročnější a nákladnější pravděpodobnostní výběry (byť je pravděpodobně naprostá většina jejich klientů nevyžaduje), CVVM v té době kapacitu k provádění pravděpodobnostních výběrů ještě nevybudovalo.

Pro koncové uživatele je rovněž důležitá velikost výběrového souboru, neboť, při jinak stejných okolnostech, znamená větší soubor větší přesnost odhadů. Z hlediska velikosti vzorku vedla v roce 2002 agentura STEM se svými měsíčně přibližně 2000 sebranými rozhovory jednoznačně nad konkurencí jak z CVVM, tak z TNS Factum. Posledním otázkem ve srovnání agentur a jejich volebních preferencí je potom samotný použitý instrument, tj. otázka, kterou agentura ke zjištění stranických preferencí používá [např. Luk 2002; Seidlová 2001, 2002; Šídlo 2002; Vlachová 2004]. Zde ovšem neexistuje v odborné veřejnosti jednoznačný konsenzus ohledně toho, který způsob dotazování je lepší a přesnější v předpovídání konečných volebních výsledků.

Z hlediska sociologické metodologie je vhodné připomenout, že tento text seznamuje českou sociologickou komunitu s jedním specifickým, nicméně užitečným metodologickým nástrojem, který nachází v soudobé sociální vědě široké uplatnění [např. Bagozzi 1978; Edmundson, Koch 1993; Jöreskog, Sörbom 1975; Ruble, Stout 1990; Warren, Hauser 1997; Williams 1994]. MIMIC model jde při hodnocení kvality škál nad některé jiné instrumenty, jako je Cronbachovo alfa. Logika MIMIC modelu rovněž poukazuje na důležitý fakt, že měřicí kvalita škály může být a velmi často je kontextuálně podmíněna. Vedlejším produktem tohoto textu je i ověření validity jedné často používané škály měřicí politickou ideologií.

MARTIN KREIDL je vědeckým pracovníkem Sociologického ústavu AV ČR se zaměřením na otázky vývoje sociálních nerovností, sociální demografii a jejich souvislosti. Vystudoval sociologii na Fakultě sociálních věd Univerzity Karlovy v Praze, kde v roce 2003 získal titul Ph.D., a na Kalifornské univerzitě v Los Angeles, kde v současné době pokračuje v postgraduálním studiu.

## Literatura

- Alwin, Duane F. 1989. „Problems in the Estimation and Interpretation of the Reliability of Survey Data“. *Quality and Quantity* 23: 277–331.
- Bagozzi, Richard P. 1978. „The Construct Validity of the Affective, Behavioral, and Cognitive Components of Attitude by Analysis of Covariance Structure“. *Multivariate Behavioral Research* 13: 9–31.

- Bell, S., C. Lee 2002. „Development of the Perceived Stress Questionnaire for Young Women“. *Psychology, Health & Medicine* 7 (2): 189–201.
- Biemer, Paul P., Lars E. Lyberg 2003. *Introduction to Survey Quality*. Hoboken: John Wiley & Sons, Inc.
- Bjorner, Jakob Bue, Tage Sondergaard Kristensen 1999. „Multi-Item Scales for Measuring Global Self-Rated Health. Investigation of Construct Validity Using Structural Equations Models“. *Research on Aging* 21 (3): 417–439.
- Blake, Judith, Barbara Richardson, Jenifer Bhattacharya 1991. „Number of siblings and sociability“. *Journal of Marriage and the Family* 53: 271–283.
- Bohrnstedt, Georg W. 1977. „Use of Multiple-Indicators Multiple-Causes (MIMIC) Model“. *American Sociological Review* 42: 656–663.
- Campbell, Todd, James Arthur 1997. „The Factor Structure of the Bem Sex-Role Inventory (BSRI): Confirmatory Factor Analysis of Long and Short Forms“. *Educational & Psychological Measurement* 57 (1): 118–124.
- Cleveland, William S. 1993. *Visualizing Data*. Summit: Hobart Press.
- Černý, Jiří 2002. „Proč průzkumy selhaly“. *Lidové noviny, příloha Orientace*, 20. 6. 2002.
- čtk. 2002. „Agentury se rozcházejí v odhadech volebních výsledků“. *Hospodářské Noviny*, 19. 4. 2002
- Disman, Miroslav 1993. *Jak se vyrábí sociologická znalost. Příručka pro uživatele*. Praha: Karolinum.
- Edmundson, Elizabeth, William W. Koch 1993. „A Facet Analysis Approach to Content and Construct Validity“. *Educational & Psychological Measurement* 53: 351–368.
- Efron, Bradley, Robert J. Tibshirani 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hill.
- Evans, Geoff, Anthony Heath, Mansur Lalljee 1996. „Measuring left-right and libertarian-authoritarian values in the British electorate“. *British Journal of Sociology* 47 (1): 93–112.
- Gallo, Joseph J., James C. Anthony, B. O. Muthen 1994. „Age Differences in the symptoms of Depression: a Latent Trait Analysis“. *Journal of Gerontology* 49 (6): 251–264.
- Gilles, David E. A. 1999. „Modeling the Hidden Economy and the Tax-Gap in New Zealand“. *Empirical Economics* 24: 621–640.
- Hauser Robert M., Arthur S. Goldberger 1971. „The Treatment of Unobservable Variables in Path Analysis“. Pp. 81–177 in: *Sociological Methodology*. Edited by H. L. Costner. San Francisco: Jossey-Bass.
- Hodge, Robert W., Donald J. Treiman 1968. „Social Participation and Social Status“. *American Sociological Review* 33: 723–740.
- Huang, Chiungjung, William B. Michael 2000. „A Confirmatory Factor Analysis of Scores on a Chinese Version of an Academic Self-concept Scale and its Invariance Across Groups“. *Educational and Psychological Measurement* 60: 722–786.
- Inverarity, James M. 1976. „Populism and Lynching in Louisiana, 1889–1896: a Test of Erikson's Theory of the Relationship Between Boundary Crises and Representative Justice“. *American Sociological Review* 41: 262–280.
- Jeřábek, Hynek 1992. „Koncepte měření v sociálních vědách“. *Sociologický časopis* 28: 103–118.
- Jones, Nicole Tuomi, Peter Ji, Mary Beck, Niels Beck 2002. „The Reliability and Validity of the Revised Conflict Tactics Scale (CTS2) in a Female Incarcerated Population“. *Journal of Family Issues* 23: 441–457.
- Jöreskog, Karl G., Arthur S. Goldberger 1975. „Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable“. *Journal of American Statistical Association* 70: 631–639.

- Jöreskog, Karl G., Dag Sörbom 1989. *LISREL 7. A Guide to the Program and Applications*. 2<sup>nd</sup> Edition. Chicago: SPSS Inc.
- Kreidl, Martin 1998. „Changes in the Perception of Poverty and Wealth: Czech Republic Between 1991–1995“. *Czech Sociological Review* 6: 73–97.
- Kreidl, Martin, Tomáš Lebeda 2003. *Pre-election polls, election results, and validity of measurement before the 2002 elections*. Sociologické texty/Sociological Papers 03:8. Praha: Sociologický ústav AV ČR.
- Krejčí, Jindřich (ed.) 2004. *Kvalita výzkumů volebních preferencí*. Praha: SOÚ AV ČR.
- Lebeda, Tomáš 2003. „Průzkumy volebních preferencí před volbami do Sněmovny 2002. Srovnání výzkumů agentur CVVM, STEM a TNSF“. *Politologický časopis* 10 (1): 22–37.
- Lipold, Jan 2002. „Proč se prognózy pokaždé aspoň trochu mýlí“. *Hospodářské noviny*, 10. 6. 2002.
- Lord, Frederic M., Melvin R. Novick 1968. *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley Publishing Company Inc.
- luk. 2002. „Volební průzkumy bývají divoké“. *Právo*, 1. 6. 2002.
- Lyberg, Lars, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo Norbert Schwarz, Dennis Trewin. 1997. *Survey Measurement and Process Quality*. New York: John Wiley & Sons, Inc.
- Lynch, Scott M. 2000. „Measurement and Prediction of Aging Anxiety“. *Research on Aging* 22 (5): 533–558.
- Mast Benjamin T., Susan E. MacNeill, Peter A. Lichtenberg 2002. „A MIMIC Model Approach to Research in Geriatric Neuropsychology: The Case of Vascular Dementia“. *Aging Neuropsychology and Cognition* 9 (1): 21–37.
- Matějů, Petr, Klára Vlachová a kol. 2000. *Nerovnost, spravedlnost a politika. Česká republika 1991–1997*. Praha: SLON.
- Piquero, Alex R., Randall Macintosh, Matthew Hickman 2002. „The Validity of a Self-Reported Delinquency Scale. Comparisons Across Gender, Age, Race, and Place of Residence“. *Sociological Methods and Research* 30 (4): 492–529.
- Reed-Ashcraft, Kellie, Raymond S. Kirk, Mark W. Fraser 2001. „The Reliability and Validity of the North Carolina Family Assessment Scale“. *Research on Social Work Practice* 11 (4): 503–520.
- Rees, Tim, Lew Hardy, David K. Ingledew, Lynne Evans 2000. „Examination of the Validity of the Social Support Survey using Confirmatory Factor Analysis“. *Research Quarterly for Exercise and Sport* 71 (4): 322–330.
- Ruble, T. L., D. E. Stout 1990. „Reliability, Construct Validity, and Response – Set Bias of the Revised Learning-Style Inventory (LSI-1985)“. *Educational & Psychological Measurement* (3) 50: 619–629.
- Russell, Daniel W. 1996. „UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure“. *Journal of Personality Assessment* 66: 20–40.
- Řehák, Jan 1998a. „Kvalita Dat I. Klasický model měření reliability a jeho praktický aplikační výzkum“. *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.
- Řehák, Jan, Irena Bártová, Jana Hamanová 1998. „Kvalita dat III. Empirické výsledky měření reliability pro vybrané míry a stupnice“. *Sociologický časopis* 34 (3): 363–372.
- Saris, W. E., A. van Meurs (eds.) 1990. *Evaluation of Measurement Instruments by Meta-Analysis of Multi-Trait Multi-Method Studies*. Amsterdam: North-Holland.
- Seidlová, Adéla. 2001. „Volební preference politických stran CVVM“. *Neviditelný pes* 21. 11. 2001, [http://pes.internet.cz/clanky/2001/11/17258\\_0\\_0\\_0.html](http://pes.internet.cz/clanky/2001/11/17258_0_0_0.html)

- Seidlová, Adéla 2002. „Výzkumy stranických preferencí“. *Neviditelný pes* 20. 3. 2002, [http://pes.eunet.cz/clanky/2002/03/20314\\_12\\_0\\_0.html](http://pes.eunet.cz/clanky/2002/03/20314_12_0_0.html)
- Scherpenzeel, Annette. 1995. *A Question of Quality. Evaluating Survey Questions by Multitrait-Multimethod Studies*. Leidschendam: KPN Research
- Schutt, Russell K. 2004. *Investigating the Social World. The Process and Practice of Research*. 4<sup>th</sup> edition. Thousand Oaks: Pine Forge Press.
- Smith, Douglas A., E. Britt Patterson 1984. „Applications and a Generalization of MIMIC Models to Criminological Research“. *Journal of Research in Crime and Delinquency* 21: 333–352.
- Stapleton, David C. 1978. „Analyzing Political Participation Data with a MIMIC Model“. *Sociological Methodology* 9: 52–74.
- STATA Corp. 2003a. *STATA Statistical Software: Release 8.0*. College Station: Stata Corporation.
- STATA Corp. 2003b. *STATA 8. Reference Su–Z*. College Station: Stata Corporation.
- Sukhwinder S. Shergill, Kuttalaliangam K. Shankar, Knightley Seneviratna, Martin W. Orrell. 1999. „The Validity and Reliability of the Health of the Nation Outcome Scales (HoNOS) in the Elderly“. *Journal of Mental Health* 8 (5): 511–521.
- Šídlo, Jindřich 2002. „Volební preference: hra bez záruky“. *MF Dnes*, 29. 4. 2002.
- Tourangeau, Roger, Eleanor Singer, Stanley Presser 2003. „Context Effects in Attitude Surveys. Effects on Remote Items and Impact on Predictive Validity“. *Sociological Methods and Research* 31: 486–513.
- Vlachová, Klára 2004. „Výzkumy volebních preferencí, žurnalistika a problém formulace otázek“. Pp. 131–138 in Krejčí, Jindřich (ed.): *Kvalita výzkumů volebních preferencí*. Editoval Jindřich Krejčí. Praha: SOÚ AV ČR.
- Ward, Edward A. 1994. „Construct Validity of Need for Achievement and Locus of Control Scales“. *Educational and Psychological Measurement* 54: 983–992.
- Warren, John Robert, Robert M. Hauser 1997. „Social Stratification Across Three Generations: New Evidence from the Wisconsin Longitudinal Study“. *American Sociological Review* 62: 561–572.
- Williams, Jannice E. 1994. „Anxiety Measurement: Construct Validity and Test Performance“. *Measurement & Evaluation in Counseling & Development* 27: 302–307.
- Wu, Lawrence L., Steven P. Martin, Daniel A. Long 1999. *Comparing Data Quality of Fertility and First Sexual Intercourse Histories*. Center for Demography and Ecology WP 99/08. Madison: University of Wisconsin.
- Zhu, Weimo 2000. „Which Should It Be Called: Convergent Validity or Discriminant Validity?“ *Research Quarterly for Exercise and Sport* 71: 190–194.

## **Příloha: znění otázek v metodologickém výzkumu**

### **Subjektivní politická orientace:**

*„V souvislosti s politikou se často hovoří o levici a pravici. Kam byste se zařadil Vy osobně na této stupnici, kde 1 znamená krajní levici a 7 krajní pravici, 4 je střed?“ (tazatel předkládá kartu zobrazenou se škálou).*

### **Subjektivní sociální status:**

*„V naší společnosti jsou skupiny, kterým je přisuzováno vysoké sociální postavení a naopak skupiny s nízkým sociálním postavením. Kam byste se Vy osobně zařadil(a) na následující škále?“ (tazatel předkládá kartu zobrazenou se škálou).*

### **Baterie otázek měřící pravolevou ideologii:**

*„Nyní Vám budu číst výroky a Vy mi prosím řekněte, nakolik s nimi souhlasíte, nebo ne-souhlasíte.“ (tazatel předkládá kartu se čtyřbodovou odpovědní škálou).*

- a) *Vláda by měla přerozdělovat příjem od těch, kteří jsou na tom lépe, těm, kteří jsou na tom hůře.*
- b) *Vlastníci velkých podniků bohatnou na úkor dělníků.*
- c) *Obyčejní pracující lidé nedostávají spravedlivý podíl z národního bohatství.*
- d) *Existuje jedno právo pro bohaté a druhé pro chudé.*
- e) *Vedení podniku se bude vždy snažit vyžrát na zaměstnance, pokud dostane šanci.“*