

UNIVERZITA KARLOVA V PRAZE
FAKULTA SOCIÁLNÍCH VĚD
Institut sociologických studií
Katedra sociologie

Výzkum volebních preferencí v ČR:

Návrh metodologické optimalizace

Diplomová práce

Mgr. Daniel Prokop

Autor práce: **Mgr. Daniel Prokop**

Vedoucí práce: **Mgr. Jindřich Krejčí, Ph.D.**

Oponent práce:

Rok obhajoby: **2012**

Hodnocení:

Bibliografický záznam

PROKOP, Daniel. (2012). *Výzkum volebních preferencí v ČR: Návrh metodologické optimalizace*. Praha: Univerzita Karlova, Fakulta sociálních věd, Institut sociologických studií. Vedoucí diplomové práce Mgr. Jindřich Krejčí, Ph.D.

Anotace

Diplomová práce se zabývá tématem volebních preferenčních výzkumů v České republice. Studie vychází z premisy, že díky charakteru mediální recepce výzkumů je nutným cílem volebních výzkumů v předvolebním období co možná nejpřesnější odhad výsledků voleb. S využitím dat společnosti MEDIAN s.r.o. pocházejících ze sběru volebních preferencí osobním a telefonním dotazováním ve volebním roce 2010 formuluje možnosti metodologické optimalizace výzkumu, které by mohly pomoci omezit systematické vychýlení předvolebních výzkumů od výsledků voleb. Zabývá například možnostmi – kombinované metodologie sběru dat; vážení dat podle specifických proměnných ovlivňujících volební preference; zařazováním preferencí dosud nerozhodnutých respondentů; odhadu pravděpodobnosti účasti respondenta u voleb; redukce oscilace modelu v čase zohledněním trendů ve vývoji preferencí. Na základě těchto analýz a srovnání s veřejnými údaji o preferenčních výzkumech v ČR se autor pokouší o zobecnění zjištění, které by mohly pomoci k rozšíření diskuse o kvalitě a metodologii preferenčních výzkumů v ČR obecně. V práci jsou použity základní statistické metody (kontingenční analýza, analýzy odchylek predikce, atd.) i pokročilé statistické metody (regresní stromy, exponenciální vyhlazování, atd.)

Klíčová slova

volební průzkumy, preference, volební model volební predikce, kombinovaná metodika sběru dat, reprezentativita, volební chování, metodologický optimalizace, vliv nerozhodnutých voličů, volební účast, vážení dat, výzkum veřejného mínění, MEDIAN

Annotation:

The thesis focuses on the election-polls and prediction of election results in the Czech Republic. Using data of research company MEDIAN s.r.o. from face-to-face (CAPI) and telephone interviewing (CATI) in election year 2010 it examines possibilities of methodological optimizations which could lead to reducing systematic bias and discrepancies of pre-election polls the election results. In particular, it discusses these methodological solutions: mix-mode data collection (combination of CATI and CAPI), data weighting focused on specific factors correlated with voting behavior, including preferences of undecided voters, prediction of the respondents' participation in elections, election-polls results time-series smoothing. Based on these analyses the thesis tries to articulate general findings which could be fruitful in discussion about Czech election-polls and their methodology in general. In the thesis, basic and advanced statistic methods (CART, exponential smoothing, etc.) are being used to achieve given research goals.

Keywords:

election polls, quantitative surveys, election behavior, election results prediction, mix-method design, representativeness, methodological optimization, undecided voters, election turn-out, data weighting, public opinion research, MEDIAN

Prohlášení

1. Prohlašuji, že jsem předkládanou diplomovou práci zpracoval samostatně s použitím uvedených a citovaných pramenů a literatury.
2. Vlastní text práce (bez anotací, obsahu, bibliografie, příloh a poznámek pod čarou) má celkem 212 836 znaků včetně mezer.
3. Souhlasím s tím, aby práce byla zpřístupněna veřejnosti pro nekomerční účely výzkumu a studia.

V Praze dne 7. ledna 2011

Mgr. Daniel Prokop

Poděkování

Děkuji Mgr. Jindřichu Krejčímu, Ph.D. za ochotu být vedoucím této práce a za studie týkající se volebního výzkumu v ČR, z kterých jsem při psaní této práce čerpal.

Obzvláště děkuji výzkumné společnosti MEDIAN s.r.o. za bezplatné poskytnutí dat z volebního výzkumu ve volebním roce 2010 k analýze, i za možnost využít některých analytických prostředků zmiňovaných v práci a konzultace ohledně používaných metodologií.

Děkuji také svým blízkým za podporu při realizaci diplomové práce a pedagogům ISS FSV UK za dřívější výuku a inspirativní prostředí panující na této škole.

Obsah:

1. Úvod	8
1.1 K ambici volebních výzkumů odhadovat výsledky voleb.....	8
1.2 K některým používaným pojmům	10
1.2.1 Dvě pojetí reprezentativity.....	10
1.2.2 Stranické preference – zdánlivé jasno v pojmech	12
1.2.3 Statistická významnost	15
1.3 Popis základních datových zdrojů	15
2. Vstupní analýzy.....	17
2.1 Srovnání predikční kvality volebních výzkumů v ČR.....	17
2.1.1 Predikční kvalita volebních modelů v roce 2006.....	17
2.1.2 Predikční kvalita volebních modelů v roce 2010.....	20
2.1.3 Srovnání kvality predikce v roce 2010 a 2006	24
2.3 Faktory ovlivňující kvalitu volební předpovědi.....	24
2.3 Faktory ovlivňující volební preference respondentů	27
2.3.1 Ideální typy volebního chování a jejich projevy v ČR	27
2.3.2 Faktory ovlivňující politické preference v ČR 2010	29
2.3.2.1 Souvislost volebních preferencí se základními sociodemografickými faktory	30
2.3.2.2 Souvislost voleb. preferencí se specifickými sociodemografickými a life-stylovými faktory	33
2.3.3 Predikce volební preference pomocí regresních stromů	39
2.3.4 Důsledky zjištění pro preferenční výzkumy.....	41
2.3.5 Srovnání preferencí nadhodnocených / podhodnocených stran podle faktorů, které nejsou předmětem sociodemografického vážení.....	42
3. Optimalizace volebního modelu na příkladu společnosti MEDIAN.....	45
3.1 Optimalizace volební predikce kombinací metod sběru dat (CAPI + CATI).....	45
3.1.1 Teoretický úvod k řešení	45
3.1.2 Vliv kombinace osobního a telefonního sběru na reprezentativitu.....	47
3.1.2.1 Reprezentativita podle základních sociodemografických ukazatelů – plné vzorky	47
3.1.2.2 Reprezentativita podle základních sociodemografických ukazatelů – vzorky voličů	49
3.1.2.3 Reprezentativita podle speciálních sociodemografických ukazatelů.....	50
3.1.2.4 Reprezentativita podle life-stylových faktorů a minulého volebního chování	51
3.1.3 Predikční kvalita volebních modelů vzniklých kombinovanou metodologií	57
3.2 Optimalizace volebního modelu alternativním vážením.....	59

3.2.1 Zkvalitnění predikčních vlastností volebního modelu vážením.....	60
3.2.2 Vliv vážení na oscilaci volebního modelu v čase	65
3.2.2 Aplikace vážení na data ze spojeného sběru (CATI + CAPI).....	69
3.3 Optimalizace volebního modelu zařazováním nerozhodnutých voličů	71
3.3.1 Aplikace zařazování nerozhodnutých na data z kombinovaného sběru (CAPI + CATI)	73
3.3.1 Aplikace zařazování nerozhodnutých na data z osobního (CAPI) sběr.....	77
3.3.2 Výběr nejlepšího modelu zařazování a alternativa „donucení“ k single-response	79
3.4 Optimalizace prognózou odhadu účasti respondenta u voleb	81
3.5 Optimalizace modelu potlačením oscilace v čase	85
4. Závěr	89
4.1 Shrnutí	89
4.1 Diskuse.....	91
Summary	92
Bibliografie:	94
Příloha 1: Metadata - metodologie CAPI a CATI výzkumu, dotazový aparát	100
Příloha 2. Tabulky kategorizace chybovosti a odchýlení + rekód aktivit.....	102
Příloha 3: Srovnání reprezentativity jednotlivých (CAPI,CATI) a spojených sběrů (CAPI+CAPI).....	104
Příloha 4: Kvalita predikce při zařazení nerozhodnutých voličů	107
Příloha 5: Popis všech testovaných vážení	113
Příloha 6. Řešení pomocí klasifikačních a regresních stromů	114

1. Úvod

Nekorespondence výsledků volebních průzkumů s reálnými výsledky voleb má obecně tři důvody – limity statistické indukce (usuzování z omezeného vzorku na cílovou populaci), občasné nedostatečné pochopení výzkumných produktů ze strany jeho uživatelů (např. média) a konečně samotné systematické nedostatky výběrových šetření – volebních průzkumů. (Krejčí 2003, s. 36)

Cílem této práce je přispět do diskuse o kvalitě volebních průzkumů v České republice ve třetí ze zmiňovaných oblastí. Na základě dat společnosti MEDIAN z volebního roku 2010 se pokusíme identifikovat možnosti systematických optimalizací vedoucích ke zkvalitnění predikce volebních výsledků a zvýšení důvěryhodnosti volebního výzkumu.

Konkrétně se zaměříme na pět oblastí: a) otázku reprezentativity¹ a hledisek reprezentativity, která jsou důležitá pro volební výzkum; b) jak lze reprezentativitu a predikční kvalitu volebního výzkumu zvyšovat použitím kombinované techniky sběru dat (CAPI + CATI) a specifickým vážením, které se zaměřuje na faktory související s volebními preferencemi; c) nakládání s nerozhodnutými voliči, kteří před volbami většinou nejsou „nepopsanými listy“, ale váhající mezi několika stranami; d) možností odhadování skutečné volební účasti respondentů jinak než na základě důvěry deklarativní odpovědi; e) možností zohledněním trendů ve vývoji preferencí redukovat oscilaci volebního modelu v čase a předejít tím desinterpretacím náhodných výkyvů způsobených náhodnou složkou výběrové chyby jako reálných procesů na politické scéně.

Díky obecným charakterům těchto otázek (musí se s nimi potýkat každé volební šetření) a shodným postupům řady agentur v některých z nich, nemá tato práce ambici být jen „případovou studií“, ale naopak pokusit se popsat některé (ačkoli jistě ne všechny) obecněji aplikovatelné možnosti optimalizace volebních výzkumů.

1.1 K ambici volebních výzkumů odhadovat výsledky voleb

Autor této studie se v říjnu 2010, půl roku po parlamentních volbách, zúčastnil konference *Parlamentní volby 2010: otázky a analýzy* organizované Sociologického ústavem AV ČR (SÚ 2010). Pátý blok konference byl ve znamení diskuse mezi zástupci

¹ K chápání pojmu reprezentativita v této studii viz kapitola 1.2.

hlavních výzkumných agentur publikujících předvolební výsledky a politologem Tomášem Lebedou. (Lebeda & Hartl & Herzmann & Vinopal & Čech 2010)

Zástupci agentur v ní formulovali ve výzkumné praxi obecně sdílený postoj, že pokud agentura nepublikuje přímo volební prognózy lze i finální výstupy volebních preferenčních výzkumů považovat toliko za indikátory vývoje politických nálad ve společnosti, které se pouze podílejí na finálním výsledku voleb. Někteří zástupci agentur přitom v nepochopení tohoto faktu médií vidí příčinu částečně neoprávněného odsouzení volebních průzkumů před volbami v roce 2010 jako neúspěšných.²

Tomáš Lebeda namítal, že výzkumné agentury by buď měly mít ambici výsledky voleb odhadovat, nebo volební průzkumy vůbec nepublikovat.³ Díky charakteru mediální produkce totiž podle něj mohou očekávat, že finální výstupy výzkumů budou před volbami v médiích jako predikce výsledku prezentovány.

Výchozím bodem této studie je souhlas s postojem T. Lebedy. Většina mediálních analýz týkajících se tématu zpravodajských hodnot (Lipmann 1922) (Galtung & Rugeová 1968) (Westerstahl & Johansson 1994) (Shulz 2006) (Prokop 2007) se shoduje, že mezi dominantní zpravodajské hodnoty patří *jednoznačnost*, *kulturní blízkost* a *konfliktnost* zpravodajského materiálu⁴ a předpoklad, že tyto obsažené hodnoty identifikuje a ocení publikum. Lze tedy oprávněně předpokládat, že zejména masové média určená nejširšímu publiku budou majoritně rámcovat téma volebních průzkumů formou, která je *jednoznačná*, *konfliktní* a *kulturně-blízká* jejich konzumentům. Tedy jako odhad výsledku voleb.⁵

Ke správnému rozlišování produktů volebních výzkumů přitom nepomáhá médiím ani fakt, že kupříkladu agentury STEM a Factum vydávaly před volbami 2010 odhady zisku mandátů ve Sněmovně, přičemž společnost Factum (2010a) připouštěla, že tento finální výstup je prognózou výsledku voleb, zatímco STEM (2010a) se tomuto slovnímu spojení

² Postojkritickým k médií prezentuje ze zástupců hlavních agentur zřejmě nejčastěji Jan Hartl. (Čro 2010).

³ Jedná se o otázku, která je diskutována v české odborné obci diskutována už od voleb 2002 (Lebeda & Krejčí & Leontiyeva 2004, s. 10) (Krejčí 200?) (Vinopal 2006, s.3). Do roku 2006 byla ale diskutována zejména v kontextu častých chybných mediálních interpretací stranických preferencí, které obsahují i nevoliče a nerozhodnuté a suma stran tedy nedává 100 %, jako ukazatele vhodného pro odhad výsledku voleb. Říjnová debata na půdě v SÚ AV ČR se zaměřovala na fakt, zda-li je zodpovědné publikovat volební model, který neobsahuje nevoliče a nerozhodnuté, či dokonce odhady zisku mandátů ve Sněmovně a stále tyto výstupy prezentovat jako něco jiného než odhad volebního výsledku.

⁴ Chceme-li být důslední je třeba dodat, že toerie zpravodajských hodnot je uplatnitelná zejména na mediální krajiny, které se pohybují v tzv. liberálním modelu - hlavním determinanem mediální produkce není státně-propagandistická úloha médií, ani společenská zodpovědnost (Siebert & Peterson & Schramm 1956).

⁵ Instituce voleb je divákovy/čtenáři dlouhodobě známá, přináší jednoznačené a definitivní výsledky a je institucionalizací boje (konfliktu) politických stran o přízeň voličů. Rámcování volebních průzkumů jako odhadu výsledku voleb tedy v tomto tématu posiluje zpravodajské hodnoty přítomné v samotných volbách.

vyhýbalo. Z metodologického i interpretačního hlediska je přitom těžko obhajitelné, že přepočtení na mandáty je něčím jiným než odhadem výsledku voleb.⁶

Pokud pomineme možnost, že by agentura publikovala výstupy typu stranických sympatií (Lebeda & Krejčí & Leontiyeva 2004) a důsledně upozorňovala média na jejich nepredikční charakter, nabízí se podle názoru autora této statě skutečně dvě čistá řešení nabízená Tomášem Lebedou – buď volební průzkumy nepublikovat vůbec, nebo jejich finální výstupy dostatečně optimalizovat a v období před volbami připouštět, že se jedná o přibližné⁷ predikce volebního výsledku.

Cílem této práce je na příkladu voleb 2010 a dat agentury MEDIAN ukázat, že existuje prostor pro poměrně jednoduché optimalizace volebních výzkumů a konstrukce jejich výstupů, které umožňují omezení systematických vychýlení volebních výzkumů. Práce tedy ve výsledku oponuje i názoru některých expertů, že nesoulad mezi výsledky volebních výzkumů a výsledkem 2010 byl způsoben výhradně reálným vývojem na politické scéně a v elektorátu stran v posledních týdnech a dnech před volbami, který nešel predikovat. (Válková 2010)

1.2 K některým používaným pojmům

1.2.1 Dvě pojetí reprezentativity

V určitých částech studie operujeme pojmem „reprezentativita“. Je přitom nutné zmínit, že tento pojem je v literatuře české sociálně-vědní výzkumné teorie a praxe používán ve dvou zásadně odlišných významech.

První pojetí říká, že reprezentativita: „(...) *představuje takovou strukturu charakteristik, kterou můžeme považovat za shodnou s populací, z níž vybraný vzorek pochází.*“ (Krejčí 2008, s. 9) (viz také Jeřábek 1992, kapitola 5.2)

⁶ Výstup přímo odpovídá reálnému výstupu voleb a na rozdíl stranických preferencí se zde tedy nedá uplatnit argument, že popisuje jen procesy v politickém naladění celé populace oprávněných voličů. Přepočtení na mandáty navíc, má-li být regulerní, nutně obsahuje i expertní prvek typický pro volební prognózy: Díky podobě d' Hondtova volebního přepočtu a struktury regionů, na které je v ČR aplikován, je pro regulerní přepočtení nutno alespoň rámcově odhadovat volební účast a volební zisky stran v krajích jednotlivě. (Lebeda 2001) To lze těžko udělat pouze z aktuálního výzkumu, na který je přepočtení aplikováno, protože kraje jsou při celkovém N=1000 zastoupeny jen cca 50 až 120 responenty. Agentura tedy musí do přepočtu zanášet expertní úsudek o struktuře volební účasti a podpoře stran v krajích, který je založen na externích informacích (krajské výzkumy, výsledky minulých voleb, spojená data agentury z celorepublikového výzkumu za další období, apod.)

⁷ Maximou žádného výzkumu a ani této práce samozřejmě není popřít limity statistické indukce a vytvářet 100% přesnou predikci. Maximou by však mělo být eliminovat systematické složky vychýlení a poučit publikující média dostatečně o náhodných složkách vychýlení.

Reprezentativita v tomto pojetí je primárně vlastností vzorku vzhledem k populaci, která může mít úrovně (Krejčí 2008, s. 9) - vzorek může populaci z daných kritérií reprezentovat hůře či lépe. Chápeme-li reprezentativitu takto, je také nutno připustit, že vzorek nemůže být reprezentativní podle všech myslitelných ukazatelů a jejich kombinací zároveň.⁸ Proto je kromě faktu, pro jakou populaci je vzorek reprezentativní, důležité uvádět, z kterých hledisek je reprezentativní. V běžném úzu to ovšem vždy neplatí a reprezentativitu je např. v kvótních výzkumech nevyřčeně myšlena reprezentativita podle základních sociodemografických znaků (Médeia Research 2010a) (STEM 2010a).

Před nebezpečím spokojit se s takto generalizovaným a nespecifickým konstatováním reprezentativity u volebních výzkumů varují například metodologické instrukce asociací ESOMAR/WAPOR. Autoři upozorňují, že například v Británii volební preference více než s kterýmkoli základním sociodemografickým znakem (věk, pohlaví, region, atd.) souvisí s členstvím osoby v odborech, v jiných zemích například s úrovní religiozity a etnicitou (USA), apod. „*Je možné sebrat vzorek reprezentativní z hlediska věku, pohlaví a sociální třídy a stále mít politicky nereprezentativní vzorek.*“ (Esomar/Wapor 2003, s. 18)

Tématu, nakolik volební preference korelují s jednotlivými charakteristikami voličů a podle kterých ukazatelů bychom tedy měly požadovat reprezentativitu se budeme zabývat v kapitole 2.3 a návrzích optimalizací v kapitolách 3.1 a 3.2. Kdykoli mluvíme v těchto kapitolách o reprezentativitě (a není uvedeno jinak), máme na mysli reprezentativitu v právě popsaném prvním pojetí, tedy vlastnost vzorku.

Druhé pojetí (Soukup & Rabušic 2007) vykládá reprezentativitu jinak⁹ – nikoli jako vlastnost vzorku a jeho struktury vzhledem ke struktuře populace, ale jako vlastnost samotné výběrové metody a říká, že: „*Zřejmě jedinou cestou, jak dosáhnout reprezentativity, je provést náhodný výběr ze všech jednotek základního souboru za situace, kdy všechny jednotky mají stejnou pravděpodobnost vybraní.*“ (Soukup & Rabušic 2007, s. 382)

V tomto pojetí není reprezentativita vlastností vzorku, která by mohla mít úrovně, ale dichotomickou proměnnou (výběr buď je reprezentativní, či není). Rovněž není nutno ani

⁸ Přinejmenší proto, že existují znaky unikátní (např. rodné číslo) a v konečné populaci také řada unikátních kombinací neunikátních znaků.

⁹ Soukup s Rabušicem (2007) jsou radikálnější a říkají, že první pojetí reprezentativity je mylné. Pro tutu studii však vycházíme z liberálnějšího nominalistické pojetí definice, jak ji razil např. Karl R. Popper (1985), podle něž definice termínu nemůže být správná či špatná, protože nenese informaci – jedná se pouze o zkrácení dlouhého příběhu na jedné straně závorky termínem na straně druhé.

nelze definovat hlediska, podle nichž je výběr reprezentativní. Praktickým problémem ovšem je, že pravděpodobnost zařazení jednotky základního souboru do vzorku závisí nejen na pravděpodobnosti, s níž bude tato jednotka oslovena, ale také na její ochotě se výzkumu zúčastnit.¹⁰ Návratnost pravděpodobnostních výběrových metod používaných v českém sociálním výzkumu dlouhodobě klesá (Krejčí 2007) a navíc není nezávislá na charakteristikách, které mohou mít souvislost s předmětem výzkumu, o čemž se přesvědčila agentura SC&C v exit-pollu roce 2006, kdy poměrně vysoká míra odmítnutí (40 %) projevila vychýlením struktury vzorku a odhadů exit-pollu (SC&C 2010b).

1.2.2 Stranické preference – zdánlivé jasno v pojmech

Tématu terminologie preferenčního volebního výzkumu se věnuje řada studií (Vinopal 2006) (Lebeda & Krejčí & Leontiyeva 2004) (Lebeda 2006) (Krejčí 2003) i prohlášení zástupců agentur (Čro 2010). Z většiny z nich lze vyrozumět, že terminologie výzkumných zpráv (stranické preference, voličské preference, volební model, volební prognóza) je od vstupu SIMAR v letech 2001 dostatečně standardizovaná a kodifikovaná a hlavní zodpovědnost za např. desinterpretaci stranických preferencí jako odhadu voleb je v nepochopení podstaty produktů preferenčních výzkumů v mediálním prostředí či nejednotná praxe výzkumných agentur (Lebeda & Krejčí & Leontiyeva 2004, s. 6) (Lebeda 2006, s. 208) (Zahradníček 2010).

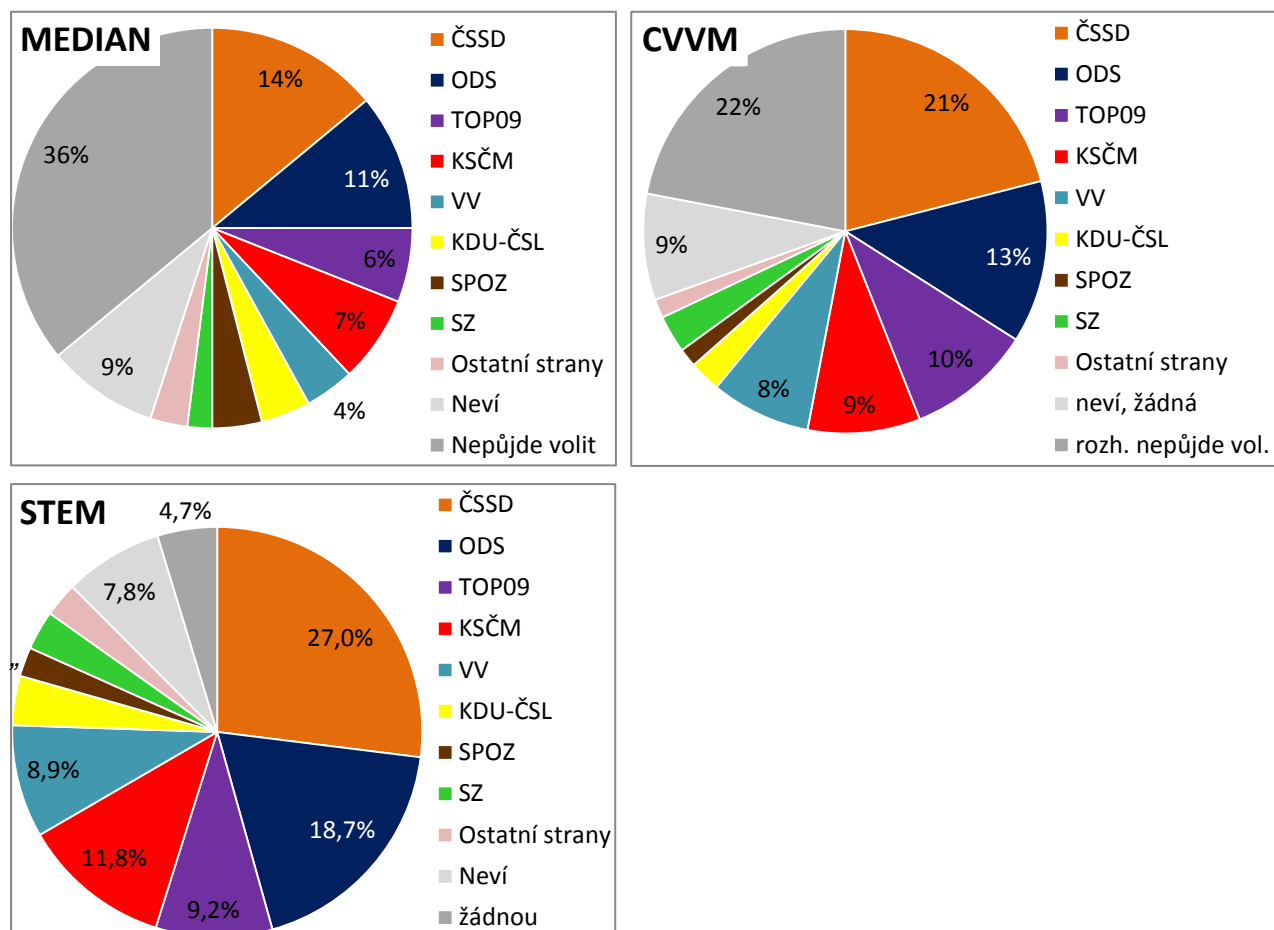
Ať už je vina kladena majoritně na stranu médií či agentur, opomíjí se hodnocení základní vstupní informace – tedy kvality definičního aparátu. Jeho problematičnost se dá dobře ilustrovat např. na *stranických preferencích*.

V tom, jak je výzkumné agentury pojmají, je zásadní odlišnost. STEM se ptá preference všech respondentů 18+ včetně těch vylučující účast. Ve výstupu stranických preferencí tak má například v květnu 2010 jen 4,7 % odpovědi „žádná“ (STEM 2010a) – což jsou respondenti, kteří byli konzistentní v první a druhé otázce (nechtějí se účastnit, nevolili by žádnou stranu). CVVM se ptá na preferenci stran jen všech respondentů, kteří v otázce na plánovanou účast neodpověděli „určitě ne“. Graf stranických preferencí obsahuje tento nedotázaný sektor a preference zbytku. MEDIAN skládá v grafu stranických preferencí sektor lidí, kteří volit nechtějí (odpověděli určitě ne + spíše ne), jenž tak odpovídá agenturou očekávané volební neúčasti 36 %, a preference zbytku vzorku.

¹⁰ Mimo to i úroveň „contact rate“ má za předpokladu omezených časových a finančních možností výzkumu limity - je nutno stanovit smysluplný počet pokusů o zastižení (v pravděpodobnostních výběrech provozovaných společnostmi MEDIAN se většinou jedná o 4 pokusy).

Praxe je tedy zcela odlišná a vede k naprosto nesrovnatelným výsledkům – např. v posledním předvolebním výzkumu byly stranické preference ČSSD 27 % u STEM, 21 % u CVVM a 14 % u MEDIANu. (viz GRAF 1.2.1)

Graf 1.2.1 Stranické preference v posledním publikovaném předvolebním výzkumu



Zdroje: (Median 2010a) (CVVM 2010a) (STEM 2010a)

Problém je přitom už v původní definici SIMAR, podle níž jsou stranické preference:

„Podíl osob, které preferují určitou politickou stranu, ze souboru všech oprávněných voličů. Součet 100 % budou u tohoto údaje tvořit příznivci jednotlivých politických stran, kteří hodlají volit, ale v době realizace nevěděli, na kterou stranu se přiklonit, a lidé, kteří volit nechtějí.“ (Lebeda & Krejčí & Leontiyeva 2004, s. 9)

Podle této definice jsou totiž obhajitelné přístupy všech tří zmíněných agentur. STEM se zaměřuje zejména na první větu definice a předpokládá, že výrazem „preferují“ se míní i

preferenci lidí, kteří se voleb nezúčastní.¹¹ Přístup CVVM a MEDIAN lze naopak pochopit, čteme-li druhou větu z definice, podle níž sumu 100 % tvoří „(...) příznivci jednotlivých stran, lidé, kteří hodlají volit, ale v době realizace výzkumu nevěděli, na kterou stranu se přiklonit, a lidé, kteří volit nechtějí.“ (Lebeda & Krejčí & Leontiyeva 2004, s. 9) CVVM za „lidé, kteří volit nechtějí“ považuje ty, kteří účast s rozhodně vyloučili („určitě ne“). MEDIAN všechny ty, kteří uvedli, že by k volbám nešli (určitě + spíše). Výhodou přístupu CVVM je, že stranické preference popisují preference všech lidí, u nichž je jistá šance, že se voleb zúčastní. Pokud ovšem chápeme negaci výroku „lidé, kteří volit nechtějí“ výrok „lidé, kteří volit chtějí“, je regulérnější přístup MEDIANu.¹²

Fakt, že všechny tři přístupy pojetí stranických preferencí odpovídají definici SIMAR a přitom jsou jejich výstupy zásadním způsobem nesrovnatelné, vybízí k otázce, zda hlavní chyba není v (ne)kvalitě této kodifikace, která rezultovala nikoli ve standardizaci výstupů z preferenčních výzkumů, ale používání standardizované terminologie pro nestandardizované výstupy. Ostatně i samotný pojem „stranické preference“ je diskutabilní, protože vyvolává dojem, že se jedná o něco jiného než neočištěný popis předpokládaného volebního chování.¹³

Díky nesrovnatelnosti stranických preferencí (a dalších výstupů)¹⁴ a naší tezi z kapitoly 1.1 budeme v následující práci porovnávat volební modely agentur, přičemž volebním modelem myslíme finální výstup agentury ve formě procentuálního zisku stran, v nichž suma zisku stran se rovná 100 %, a pozorovatel (např. novinář) jej tak může považovat za nejrelevantnější odhad výsledku voleb publikovaný danou agenturou.

¹¹ Což zní sice logicky vzhledem k českému jazykovému ekvivalentu slova „preferovat“ = upřednostňovat. Musíme si ale uvědomit, že volební otázka u STEM ani dalších agentur nezní „Jakou stranu preferujete?“, ale ptá se přímo na (hypotetické) volební chování.

¹² Méně informovaný čtenář zde také nehledá rozpor mezi odhadovanou volební (ne)účastí a údajem o nevolících v grafu stranických preferencí, protože jsou konzistentní. Stranické preference díky své nízké hodnotě rovněž nenavádějí méně informovaného čtenáře interpretovat je jako výsledek voleb - u stranických preferencí STEM či CVVM si např. u ČSSD (27 % respektive 21 %) může novinář díky relativní blízkosti volebním výsledkům této strany splést s odhadem voleb.

¹³ Volební chování voličů nekopíruje 100% politické preference (viz kapitola 2.3). Například příznivec Strany zelených, který se ve volbách 2010 rozhodl hlasovat pro TOP 09, což byl podle exit-pollu SC&C poměrně zastoupený jev (SC&C 2010) - například díky reálnému riziku nepřekročení 5% klauzule u SZ, nemusí projevovat tímto rozhodnutím změnu politických preferencí, ale jen volební kalkul. Otázky, z nichž vycházejí stranické preference jsou formulovány jako odhad vlastního volebního chování (STEM 2010a) (CVVM 2010a) (MEDIAN 2010a) – zahrnují tak i podobné typy předpokládaných volebních kalkulů, které nemusí odpovídat tomu, kterou stranu politického spektra respondent skutečně politicky preferuje (česky upřednostňuje).

¹⁴ Např. voličských preferencí, které ze stranických preferencí vycházejí a závisí na tom, co agentura považuje za odmítnutí účasti u voleb.

1.2.3 Statistická významnost

V práci se jen omezeně vyskytuje argumentace statistickou významností. Jak upozorňují například Soukup a Rabušic (2007), statistická významnost je někdy v českém sociálně-vědním diskursu používána takřka obsesivně a i v případech, kde se její použití nehodí.

Většina našich analýz (například srovnání vedoucí k určení predikční kvality volebních modelů – viz kapitola 2.1 a 3) patří právě do této kategorie. Říci například, že se výsledky výzkumu statisticky významně neliší od výsledků voleb (na základě testů dobré shody či t-testů u zisků jednotlivých stran),¹⁵ neříká takřka nic. Pokud bychom volební odhad založili na vzorku 100 respondentů vybraných pravděpodobnostním výběrem, dospěli bychom k závěru, že se statisticky významně nemýlí, ačkoli kvalita predikce mohla být zcela nedostačující. Naopak výzkum na 5000 respondentech by se mohl statisticky významně odchylovat, ačkoli jeho predikční kvalita bude velmi dobrá. (Soukup & Rabušic 2007) Ve většině studie tedy používáme jiné ukazatele predikční kvality. (viz kapitola 2.1)

1.3 Popis základních datových zdrojů

Hlavní analytická část této práce vychází z dat agentury MEDIAN. MEDIAN v období 2006-2010 kontinuálně sbíral výzkum volebních preferencí jakou součást svého CAPI omnibusu. Publikace probíhala v měsíčních fázích, přičemž jedna fáze sběru obsahovala 850 – 1050 respondentů ve věku 18+. MEDIAN poskytl pro účely této práce data z volebního modulu CAPI omnibusu a požadované identifikační proměnné za období duben 2010 až prosinec 2010.

Mimo CAPI omnibusu zařadila společnost MEDIAN volební modul v měsících duben, květen, červenec 2010 i do telefonicky dotazovaného (CATI) výzkumu. Jeho výsledky před volbami 2010 nebyly publikovány, aby agentura dodržela požadavek konzistentní metodologie v celém volebním období 2006-2010 (Esomar 2003, s. 18). Umožňují ale detailnější analýzy v otázce optimalizace volebních výzkumů použitím kombinované metodiky sběru dat.

CAPI omnibus společnost MEDIAN sbírala dotazováním školenými tazateli v domácnostech. Výběr respondentů probíhal *vícetupňovým stratifikovaným náhodným adresním výběrem*, což je metodologie používaná řadou mezinárodních výzkumných

¹⁵ Je to příklad jinek pěkné a informačně obsáhlé bakalářské práce Kristýny Chábové z roku 2011, která obsahovala rozhovory se zástupci výzkumných agentur a je tak v jistých případech cenným informačním zdrojem (Chábová 2011)

šetření (např. ISSP, ESS). CATI volební data společnost MEDIAN sbírala telefonickým dotazováním tazatelů z centra v Praze. Výběr telefonních čísel z kombinované opory pevných telefonních linek a mobilních kontaktů byl náhodný (stratifikovaný náhodný výběr u pevných linek, prostý náhodný výběr mobilních kontaktů). V CATI výzkumu byl na rozdíl od osobního dotazování v měsíční fázi obsažen prvek kvótního výzkumu – v podobě eliminování respondentů vybraných náhodným výběrem, kteří by narušovali očekávanou strukturu vzorku z hlediska základních sociodemografických ukazatelů (pohlaví, věk, vzdělání, kraj, velikost sídla). Počet respondentů 18+ v měsíční fázi je ve využívaných fázích mírně vyšší, než u CAPI omnibusu (N=1000 až 1350). Podrobněji k metodologii a znění dotazníku volebního průzkumu viz Příloha 1.

Tabulka 1.3.1 popisuje datové zdroje, které budeme v naší analýze nejčastěji využívat,¹⁶ a zároveň ukazuje, jak budeme jednotlivé fáze značit (např. odlišné značení plovoucích a spojených dat z období duben a květen 2010).

TAB 1.3.1 Počet respondentů v datových zdrojích a jejich značení

		Měsíční fáze			Plovoucí data ¹⁷
		Duben 2010	Květen 2010	Červenec 2010	Duben-Květen 2010
CAPI	N (12-79)	1016	980	1077	962
	N (18-79)	914	883	964	874
CATI	N (12-79)	1358	1343	1140	1420
	N (18-79)	1312	1261	1059	1349

Zdroj: Společnost MEDIAN, poskytnutá data

POZN (). Datové úseky obsahující měsíc květen budeme v některých případech omezovat do začátku voleb (28.10.2010 14:00). Tyto případy jsou označeny hvězdičkou a komentovány.*

¹⁶ V analýze vlivu alternativních vážení na oscilaci volebního modelu využijeme i data z osobního sběru v červnu 2010 a srpnu-prosinci 2010 (viz kapitola 3.2).

¹⁷ Plovoucími daty rozumíme data, která z druhé poloviny dubna a první poloviny května (tj. jsou sebrána v období 17.4. – 16.5. 2010).

2. Vstupní analýzy

2.1 Srovnání predikční kvality volebních výzkumů v ČR

Úroveň predikce výsledků voleb do PSP ČR, které proběhly v květnu 2010, českými výzkumnými agenturami je žurnalistickou a akademickou obcí obecně považována za slabou. Analýzy novinářů (Válková 2010) i některých vysokoškolských prací (Chábová 2011) však často nabízejí minimum hodnotících kritérií pro skutečné posouzení, jakou kvalitu měla skutečně predikce jednotlivých agentur před volbami 2010, a to například ve srovnání s volbami minulými.

O tuto klasifikaci se v úvodě naší analýzy pokusíme. Za rok 2010 jsme do srovnání volebních modelů zařadili všechny celostátní veřejně dostupné preferenční výzkum publikované v květnu 2010, či poslední předvolební výzkum dané agentury (např. MEDEA – 28. 4. 2010) daného typu (důvod zařazení i výzkumu STEM pro ČSSD). Za rok 2006 jsme do srovnání zařadili poslední zveřejněné volební výzkumy před volbami v červnu 2006 od tehdy hlavních agentur – STEM, CVVM, Factum a SC&C.

2.1.1 Predikční kvalita volebních modelů v roce 2006

Na volbách 2006 jsme stanovili kritéria hodnocení kvality. Tabulka 2.1.1 ukazuje srovnání 5 modelů 4 hlavních agentur a exit-pollů SC&C pro Českou televizi.

TAB 2.1.1 Volební modely před volbami do PSP 2006 (%)

ZDROJ	Volby do PSP	Srovnání volebních modelů						
		SC&C exit poll 16:00 ^c	SC&C exit poll 14:00 ^c	STEM ^a	CVVM ^b	FACTUM ^a	SC&C pro MF DNES ^c	SC&C pro ČRo ^c
ukončení sběru dat	3.6.	3.6.	3.6.	9.5.	8.5.	24.5.	26.5.	22.5.
ODS	35,4	37,0	38,0	33,2	32,0	27,8	35,6	35,6
ČSSD	32,3	31,0	30,0	31,4	28,0	28,5	27,6	28,7
KSČM	12,8	12,0	12,0	16,2	15,5	17,3	13,8	14,9
KDU-ČSL	7,2	8,0	8,0	6,8	5,5	9,0	6,9	5,7
SZ	6,3	7,0	7,0	8,8	10,5	8,6	10,3	11,5
ostatní strany	6,0	5,0	5,0	3,6	8,5	8,8	5,7	3,4

ZDROJE: a - STEM 2006), b - (CVVM 2006), c – (SC&C 2006)

Barevně vyznačené buňky v tabulce označují hodnoty silně vychýlené oproti skutečným výsledkům jednotlivých stran ve volbách. Modré buňky označují silně nadhodnocené a červené silně podhodnocené strany. Silné vychýlení nelze definovat absolutní hodnotou odchylky (penalizuje odchylky u velkých stran) ani relativní hodnotou odchylky (penalizuje odchylky u malých stran). Pro definici silného vychýlení jsme tedy použili srovnání

s intervalem spolehlivosti okolo reálného volebního výsledku jednotlivých stran, který by vznikl, pokud by se jednalo o náhodný výběr vzorku $N=1000$, s reálnou volební účastí 62,6 % (tedy spolehlivostní interval na daném kvantilu pro náhodný výběr vzorku $N=626$).¹⁸

Pro regulérnější srovnání jednotlivých výzkumů byl navíc tento interval rozšířen o 10 % (nikoli procentních bodů) na každou stranu pro výzkumy, u nichž byl sběr dat ukončen více než 21 dní před začátkem voleb a je u nich tedy větší šance, že odchylka od voleb je dána reálným posunem na politické scéně. Výsledné konfidenční intervaly pro definici silně vychýlených odhadů ukazuje TAB 2 v Příloze 2.

Pro srovnání celkové predikční kvality jednotlivých volebních modelů nestačí identifikace silných odchýlení (viz výše). Predikční kvalita se dá posuzovat hned z několika hledisek. Například:

- Suma čtverců odchylek¹⁹ jednotlivých stran
- Průměrná absolutní odchylka u strany
- Průměrná absolutní odchylka u strany
- Maximální absolutní odchylka u strany
- Průměrná relativní odchylka u strany²⁰
- Maximální relativní odchylka u strany
- Absolutní odchylka v rozdílu mezi první a druhou stranou
- Počet stran, u nichž model chybně predikuje (ne)kvalifikaci do Sněmovny²¹

Každé z těchto kritérií má svou legitimitu. *Suma čtverců odchylek* je standardní statistickou metodou pro analýzu odchýlení predikovaných a pozorovaných hodnot (Lukeš & Machek 1988). *Maximální absolutní odchylka* patří mezi mediálně nejviditelnější chyby predikce. *Maximální relativní chyba* postihuje v hodnocení případy, kdy model celkově predikuje správně, ale silně nadhodnocuje nějakou z malých stran (případ průzkumu WWR Publishing z května 2010, podle nějž měla Suverenita získat 10,2 %, ale jinak predikoval nadprůměrně dobře – viz TAB 2.1.5). Počet stran, u nichž model chybně predikuje, zda se kvalifikují, či

¹⁸ Používat kritérium, zda jsou hodnoty statisticky významně odchýleny od voleb na základě parametrů vlastního výzkumu nelze, protože většina agentur používala kvótní výběr. Navíc není regulérní pro srovnání vycházející z parametrů samotných výzkumů. Např. u malých vzorků bychom dospěli k zjištění, že odchylka od voleb není statisticky významná – to však neznamená vyšší kvalitu predikce, velikost vzorku je jedním z parametrů kvality výzkumu a zodpovědností výzkumné agentury.

¹⁹ Odchylkou rozumíme odchylku predikované hodnoty pro danou stranu ve volebním modelu (%) od skutečného volebního výsledku.

²⁰ Relativní odchylkou u strany s volebním ziskem X a předpovídanou hodnotou Y rozumíme hodnotu $100 * |Y-X| / X$. Z těchto relativních odchylek je vytvořen aritmetický průměr.

²¹ Toto kritérium nebylo uplatněno jako srovnání absolutní predikované hodnoty a kvalifikačního kritéria 5 %, protože při správné interpretaci a zmínění statistické odchylky nelze např. hodnotu 5,5 % ve volebním modelu interpretovat jako jistou predikci kvalifikace dané strany do PSP. Pro definici chybných predikcí o kvalifikaci do Sněmovny využít referenční výzkum $N=1000$ s volební účastí 62,6 % (tj. $n=626$). Chybnou predikcí je případ, kdy se z preferencí strany dá se statistickou významností usuzovat na nepřekročí / překročí 5 %, tj. predikovaná hodnota leží mimo interval $<3,3 \% - 6,8 \%$.

nekvalifikují do PSP ČR, je důležitý z hlediska správnosti predikce složení Sněmovny i z hlediska reálného vlivu výzkumu na politické rozhodování voličů, které zvažují volbu menších stran. (Vinopal 2006) *Absolutní odchylka v určení rozdílu mezi první a druhou stranou* je rovněž důležitá z hlediska správnosti predikce složení budoucí Sněmovny a vlády²² a při mediální interpretaci voleb jako souboje dvou hlavních politických stran (od roku 1996 ODS a ČSSD) je také mediálně výrazně viditelnou chybou.

Za účelem srovnání celkové predikční kvality byly zkalkulovány všechny výše zmíněné ukazatele pro volební modely publikované před volbami do PSP ČR v roce 2006 a 2010. Dále byla pro každý ukazatel arbitrárně vytvořena evaluační škála, která kategorizuje predikční kvalitu volebního modelu podle daného hodnotícího kritéria mezi hodnotami 1 (vynikající) až 7 (zcela chybná).

Tato evaluační škála byla vytvořena nejprve na volebních modelech z roku 2006²³ s přihlédnutím k tomu, jakých hodnot nabývá u exit-pollu agentury SC&C publikovaného 29. 5. 2010 v 15:40, který lze považovat za referenční "*vynikající predikci*" (hodnoty 1 u všech predikčních kritérií). Viz Příloha 2, Tab 1.

Ze sedmi různých hodnot predikčních kvalit, které odpovídají sedmi použitým hodnotícím kritériím, byl poté aritmetickým průměrem spočítán *Index celkové predikční kvality modelu*. Stejně evaluační škály byly aplikovány na volební modely z roku 2006 (TAB 2.1.2) a 2010 (TAB 2.1.4). Jejich kalibrace může být předmětem diskuse. Účel tohoto nástroje je však zejména srovnávací a slouží ke:

- a) srovnání kvality predikce volebních modelů různých agentur
- b) obecné srovnání kvality predikce před volbami 2006 a před volbami 2010;
- c) ohodnocení kvality budoucích návrhů na optimalizaci designu volebního výzkumu a konstrukce volebního modelu na základě data společnosti MEDIAN

Při srovnání v roce 2006 se ukazuje celkově spíše dobrá kvalita predikce volebních modelů většiny agentur (STEM, CVVM, SC&C pro MF DNES) – viz TAB 2.1.2. V případě STEM se přitom kvalita predikce (index 2,4) výrazně neliší od prvního exit-pollu SC&C (2,1). Výrazně hůře dopadly z hlediska celkové predikční kvality jen Factum (index 3,9) a výzkum SC&C pro Český rozhlas (index 3,6).

²² A to i díky volebnímu spektru v ČR, v kterém od roku 1996 nikdy nevznikla situace, že by vítěz voleb sestavoval koalici s druhou nejúspěšnější stranou (nepočítáme-li případ opoziční smlouvy z roku 1998) a odchylka první dvou stran je tedy jedním z faktorů determinujících možnost sestavení vládní koalice a míru její většiny ve Sněmovně.

²³ Škála byla zkonstruována tak, aby při srovnání v roce 2006 zhruba rovnoměrně distribuovala hodnoty všech kritérií kvality predikce – tj. aby průměrná kvalita výzkumů z daného kritéria byla srovnatelná u všech kritérií. To neplatí pro počet chyb v kvalifikaci stran do PSP, kde se žádná z agentur nezmýlíla.

TAB 2.1.2 Volební modely před volbami do PSP 2006

HODNOTÍCÍ KRITÉRIUM	Hodnocení chybovosti						
	SC&C exit poll 2	SC&C exit poll 1	STEM	CVVM	FACTUM	SC&C pro MF DNES	SC&C pro ČRo
Součet čtverců odchylek	7,0	14,8	29,4	64,1	108,8	39,3	53,5
Průměrná abs. odchylka	1,0	1,4	2,0	3,1	3,8	1,8	2,5
Maximální abs. odchylka	1,6	2,6	3,4	4,3	7,6	4,7	5,2
Průměrná relativní odchylka	8,9%	9,9%	20,1%	29,3%	29,4%	15,9%	29,1%
Maximální relativní odchylka	16,7%	16,7%	40,0%	66,7%	46,7%	63,5%	82,5%
Odchylka v rozdílu mezi 2 lídry	2,9	4,9	1,3	0,9	3,8	4,9	3,8
Počet chyb v kvalifikaci do PSP	0	0	0	0	0	0	0
Kategorie kvality predikce							
Součet čtverců odchylek	1	2	2	3	5	2	3
Průměrná abs. odchylka	2	2	3	4	5	3	4
Maximální abs. odchylka	2	2	3	3	5	3	4
Průměrná relativní odchylka	2	2	3	4	4	3	4
Maximální relativní odchylka	1	1	3	4	3	4	5
Odchylka v rozdílu mezi 2 lídry	3	5	2	1	4	5	4
Počet chyb v kvalifikaci stran do PSP	1	1	1	1	1	1	1
Index celkové predikční kvality modelu							
INDEX (mean)	1,7	2,1	2,4	2,9	3,9	3,0	3,6

2.1.2 Predikční kvalita volebních modelů v roce 2010

V roce 2010 se většina volebních odhadů se silně mýlila u 3 až 4 stran. (viz TAB 2.1.5 níže). Pro definici silného vychýlení u hodnot jednotlivých stran jsme přitom aplikovali stejná kritéria jako v roce 2006 vycházející ze srovnání s konfidenčními intervaly kolem reálných výsledků v případě, že by vycházely z referenčního výzkumu s náhodným výběrem N=1000 a volební účastí 62,6 % (tj. n=626). Pro výzkumy s koncem sběru dat více než 21 dní před volbami bylo opět uplatněno rozšíření intervalu o 10 % na každou stranu. Viz intervaly slabého vychýlení v Příloze 2.

Srovnání 10 volebních modelů (TAB 2.1.3) ukazuje, že většina modelů se silně mýlila u 3 až 4 stran. U parlamentních stran bylo v roce 2010 typickým jevem nadhodnocení ČSSD a podhodnocení TOP09. U většiny agentur došlo navíc k podcenění malých neparlamentních stran, a to včetně Suverenity a v některých případech SPOZ. Ostatní silné výchylky jsou spíše typické pro modely jednotlivých či několika agentur (např. podcenění Věcí veřejných).

Zaměříme-li se na srovnání celkové predikční kvality modelů před volbami 2010 (viz TAB 2.1.4), lze konstatovat, že z agentur aplikující standardní kvantitativní metodologii předvolebních preferenčních průzkumů byly relativně nejúspěšnější agentura Factum a Médea Research (index 3 – 3,3).

TAB 2.1.3 Volební modely agentur před volbami do PSP ČR v květnu 2010

	Volby do PSP	SC&C exit-poll 15:40 ^a	SC&C exit-poll 14:00 ^a	Factum 2 ^b	Factum 1 ^c	CVVM ^d	STEM ^{*e}	STEM pro ČSSD ^f	SANEP ^g	MEDEA ^h	WWR Publishing pro Suverenitu ^{*i}	100 chytrých ^j	MEDIAN ^k
datum publikace	29. 5. 2010	29. 5.	29. 5. 14:00	12.5.	28. 4.	10. 5.	8. 5.	13.4	4.5.	28.4.	24.5.	24. 5.	1.5.
ČSSD	22,1	22,5	20,0	26,3	27,5	30,5	30,8	33,2	29,9	27,5	21,3	28,2	26,2
ODS	20,2	20,5	20,0	22,9	21,7	19,0	21,3	23,0	22,3	23,0	18,9	22,2	19,0
TOP 09	16,7	16,5	17,0	10,9	11,1	14,0	10,5	10,0	10,1	15,8	19,3	12,2	10,7
KSČM	11,3	11,3	11,0	13,1	13,9	13,0	13,5	12,7	12,9	10,2	10,9	12,8	13,3
VV	10,9	10,7	11,0	12,6	11,0	11,5	10,2	8,5	9,8	12,1	6,4	7,6	7,6
KDU-ČSL	4,4	4,1	5,0	5,5	5,2	3,5	4,5	4,2	4,7	4,5	2,5	5,3	7,5
SPOZ	4,3	4,1	4,0	2,6	3,2	2,0	2,6	3,5	5,5	2,6	3,5**	3,2	6,8
Suverenita	3,7	3,4	3,0	1,7**	1,7**	1,0**	1,5**	1,5**	1,0**	0,6**	10,2	1,3	3,1
SZ	2,4	2,3	3,0	2,5	2,9	4,5	3,5	1,7	2,8	3,0	3,8	3,6	3,5
Ostatní	4,0	4,6	6,0	1,9	1,8	1,0	1,6	1,7	1,0	0,7	3,3	3,6	2,2

**Tyto agentury zveřejnily pouze stranické preference a odhad získaných mandátů ve Sněmovně. Volební model je pro účely analýzy vypočítán přepočtem na voličské preference rozhodnutých = stranické preference lidí, kteří nevyklučují účast u voleb a ví, jakou stranu by volili (tj. model přepočtu využívaný např. CVVM).*

*** V modelech těchto agentur byla Suverenita díky předpokladu malého volebního zisku zahrnuta pod „ostatní strany“. Pro účely analýzy a srovnání byla hodnota Suverenity dopočítána s předpokladem, že daná výzkumná agentura správně odhadla poměr zisku Suverenity ke zbytku ostatních stran (tj. 3,7:4). Tento postup byl uplatněn i u SPOZ ve volebním modelu WWR Publishing. Hodnota zbytku „ostatních stran“ byla přepočítána, aby došlo k uchování sumy volebního modelu 100 %.*

Zdroje: a – (SC&C 2010b); b – (Factum 2010b); c – (Factum 2010a); d – (CVVM 2010a); e – (STEM 2010a); f – (ČTK 2010a); g – (SANEP 2010a); h – (Médea Research 2010a); i – (Parlamentní listy 2010); j – (SCIO 2010); k – (MEDIAN 2010a)

TAB 2.1.4 Analýza predikční kvality volebních modelů před volbami do PSP ČR 2010

ZDROJ	Hodnocení chybovosti											
	SC&C exit poll 2	SC&C exit poll 1	Factum 2	Factum 1	CVVM	STEM	STEM pro ČSSD	SANEP	MEDEA	WWR Publishing pro Suverenitu	100 chytrých	MEDIAN
Součet čtverců odchylek	0,6	9,9	77,2	80,5	109,3	135,8	195,0	130,5	64,2	78,4	82,9	89,9
Průměrná abs. odchylka	0,3	0,7	2,3	2,2	2,6	2,6	3,1	2,7	2,0	2,1	2,3	2,6
Maximální abs. odchylka	0,7	2,1	5,8	5,6	8,4	8,7	11,1	7,8	5,4	6,5	6,1	6,0
Průměrná relativní odchylka	4,8%	13,0%	27,4%	26,3%	39,0%	31,6%	30,8%	30,9%	29,7%	38,5%	27,6%	33,4%
Maximální relativní odchylka	17,5%	50,0%	53,2%	54,5%	87,5%	60,0%	58,4%	74,0%	83,1%	175,2%	64,1%	70,4%
Chyba v rozdílu mezi 1. a 2. stranou	0,1	1,9	1,5	3,9	9,6	7,6	8,3	5,7	2,6	0,6	4,1	5,3
Počet chyb v kvalifikaci stran do PSP	0	0	0	0	0	0	0	0	0	0	0	1
ZDROJ	Kategorie kvality predikce											
	1	2	3	4	5	6	7	8	9	10	11	12
Součet čtverců odchylek	1	1	4	4	5	6	7	6	3	4	4	4
Průměrná abs. odchylka	1	2	3	3	4	4	4	4	3	3	3	4
Maximální abs. odchylka	1	2	4	4	5	5	7	5	4	4	4	4
Průměrná relativní odchylka	1	2	4	4	5	4	4	4	4	5	4	4
Maximální relativní odchylka	1	3	3	3	5	3	3	4	5	7	4	4
Chyba v rozdílu mezi 1. a 2. stranou	1	2	2	4	7	7	7	6	3	1	5	6
Počet chyb v kvalifikaci do PSP	1	1	1	1	1	1	1	1	1	3	1	3
Index celkové predikční kvality modelu												
INDEX (mean)	1,0	1,9	3,0	3,3	4,6	4,3	4,7	4,3	3,3	3,9	3,6	4,1

Poslední publikovaný volební model agentury MEDIAN se dá přes specifické predikční chyby (nadhodnocení KDU-ČSL a SPOZ, silné podcenění VV) označit ve srovnání s konkurencí v roce 2010 hodnotit jako průměrný (index 4,1). Oproti ostatním agenturám MEDIAN více chyboval v odhadu stran, které projdou do Sněmovny. Jeho nevýhodou oproti většině konkurenčních výzkumů však byl fakt, že díky měsíční frekvenci preferenčních volebních výzkumů MEDIANu sběr dat končil až 28 dní před začátkem voleb.

Horších, ačkoli nikoli nesrovnatelných výsledků, které výrazněji překračují hodnotu indexu 4, pak dosáhly volební modely STEM, CVVM, SANEP. U všech k tomu došlo zejména díky kombinaci silného nadhodnocení zisku ČSSD a dalších odchylek specifických pro jednotlivé agentury (viz TAB 2.1.3).

Nejlepším výzkumům (Factum, Médea Research) se vyrovnal odhad společnosti SCIO pomocí akce 100 chytrých, který nebyl koncipován jako kvantitativní běžný výzkum volebních preferencí na reprezentativní populaci 18+, ale metodou tzv. crowd-surfingu (6758 tipujících v anketě předvídalo výsledky voleb a volební model byl aritmetickým průměrem jejich odhadů).²⁴ Výzkum 100 chytrých se sice od reálných výsledků voleb odchýlil podobným směrem jako volební modely řady agentur (nadhodnocení ČSSD, podhodnocení TOP09, VV a Suverenity), ale ve většině případů v menší míře (viz TAB 2.1.3). Navíc se od většiny volebních modelů odlišoval správným odhadem, že odklon od velkých politických stran povede k tomu, že poměrně velké procento hlasů (odhad 3,6 %, volební realita 4 %) dostanou nejmenší strany se ziskem pod 2 % na jednu stranu. (viz TAB 2.1.3).

Je však otázkou, zda se však výsledek *100 chytrých* dá interpretovat tak, jak to činí jeho tvůrci – že crowd-surfing je plnohodnotnou alternativou preferenčního výzkumu. Z výsledků a toho, jakým způsobem replikují hlavní chyby preferenčních výzkumů, se totiž usuzovat, že většina hlasujících vycházela z volebních modelů či obecného povědomí o jejich výsledcích. Výsledek *100 chytrých* totiž zahladil individuální chyby jednotlivých agentur, ale uchoval si hlavní společné trendy vychýlení (nadhodnocení ČSSD a podhodnocení TOP09). Není tedy jasné, jak by účastníci hlasovali bez existence preferenčních výzkumů.

Výzkum *100 chytrých* správně předvídá mírně silnější postavení budoucí pravicové vládní koalice, než jaké odhadovala většina volebních modelů. Zdůvodnění lze hledat více: a) probíhal až do druhé poloviny května a mohl tak zachytit poslední procesy na politické scéně; b) za předpokladu, že respondenti v on-line anketě promítali kromě „expertního odhadu“ také vlastní přání výsledku, dal se posun k pravici předpokládat, protože struktura on-line populace je mladší a vzdělanější než struktura běžné populace 18+ (Eurostat 2010) a

²⁴ Viz (SCIO 2010)

má tedy větší sklon k volbě pravicových stran (viz kapitola 2.3). Tento vychylující vliv mohl sloužit jako vyvážení vlivů, které v preferenčních výzkumech roku 2010 vedly k nadhodnocení volebního výsledku levice (více Kapitola 3).

Specifickým případem je výzkum agentury WWR Publishing. Volební model zde silně nadhodnotil volební výsledek zadavatele výzkumu – Suverenity, a to o 175 % (predikce zisku 10,1 % oproti 3,7 % získaných ve volbách – viz TAB 2.1.3). Díky obecně správné predikci odlivu voličů od dvou nejsilnějších stran a zejména od ČSSD však model v ostatních kritériích predikuje poměrně dobře. (Viz TAB 2.1.4)²⁵

2.1.3 Srovnání kvality predikce v roce 2010 a 2006

Srovnání predikční kvality výzkumů v roce 2006 a 2010 ukazuje, že volební modely, které se v roce 2006 od výsledků voleb odchýlily nejvíce (Factum, SC&C pro ČRo), by v roce 2010 patřily k modelům nadprůměrným. Všechny ostatní modely z roku 2006 byly hlediska celkové predikční kvality lepší než jakýkoli model z roku 2010. Srovnej TAB 2.1.4 a TAB 2.1.7. Obecné vnímání, že předpovědi v roce 2010 byly neúspěchem, je tedy oprávněné.

Zajímavé je také srovnání predikční kvality exit-pollů, které v roce 2006 a 2010 realizovala pro Českou televizi agentura SC&C. Po roce 2006 byla tato agentura často terčem ostré kritiky za nepřesnost exit-pollu. (Kotrba 2006)

Naše analýza podle řady hodnotících kritérií potvrzuje, že celková predikce exit-pollů v roce 2010 byla kvalitnější než v roce 2006. Ukazuje se však, že rozdíl mezi kvalitou prvních odhadů (těsně po uzavření volebních místností v 14:00) nebyl v roce 2006 a 2010 příliš dramatický (index 1,9 versus 2,1). Zatímco v roce 2010 ale SC&C v průběhu odpoledne druhého volebního dne výsledky výrazně zpřesnila, v roce 2006 k výraznému zpřesnění nedošlo (index 1 versus 1,7). Srovnej TAB 2.1.4 a TAB 2.1.7.²⁶

2.3 Faktory ovlivňující kvalitu volební předpovědi

Hlavní otázku povolebních diskusí – proč se odhady výzkumů lišily od voleb – lze zkoumat srovnáním hlavních parametrů posledních předvolebních výzkumů jednotlivých agentur a celkové predikční kvality volebních modelů těchto výzkumů. Do tabulky 2.3.1 jsme shrnuli

²⁵ Obecně se dá konstatovat, že pro důvěryhodnost odhadu není důležité jen to, jaký je průměr jednotlivých kategorií kvality (tj. celkový index kvality), ale v případech hodnotících kritérií model získává velmi negativní známky 6 či 7. Protože právě na tyto aspekty se poutá pozornost novinářů a akademiků.

²⁶ Důvodem chybovosti v roce 2006 byla vysoká míra odmítnutí odpovídat v exit-pollu (40 %), díky níž byla výtěžnost (field response-rate) nižší než např. u CAPI volebního sběr MEDIAN. Analýza sociodemografie vzorku v exit-pollu SC&C navíc ukázala, že míra odmítnutí výrazně souvisela se sociodemografickými charakteristikami respondentů u volebních místností, což vedlo k vychýlení reprezentativity vzorku (viz SC&C 2010b).

základní veřejně přístupné údaje týkající se metodologie volebních průzkumů a konstrukce volebních modelů u agentur, které se volebnímu průzkumu věnovaly kontinuálně v letech 2006-2010.

TAB 2.3.1 Základní parametry výzkumu a úspěšnost predikce

Název společnosti	MEDIAN	STEM	Factum	CVVM
Počet respondentů (18+)	914	1257	1004	1016
Metoda sběru dat	Face-to-face	Face-to-face	Face-to-face	Face-to-face
Výběr	pravděpodobností	kvótní	kvótní	kvótní
Kritéria reprezentativity ²⁷	základní sociodem. + pravděpod. výběr	základní sociodemografie	základní sociodemografie	základní sociodemografie
Prodleva od voleb ²⁸	42 dní	25 dní	18 dní	21 dní
Typ hlavní volební otázky	otevřená	otevřená	uzavřená – karty	otevřená
Odhadovaná účast	64 %	62-86 % ²⁹	66,8 %	61 %
Zahrnutí avizujících neúčast ³⁰	Ne	Ano	Ano	Ano
Zahrnutí nerozhodnutých ³¹	Ne	Ne	Ne	Ne
Zohlednění minulých voleb ³²	Ne	Ano	Ano	Ne
Úspěšnost predikce ve volbách 2006	x	Ano	Ne	Ano
Kontinuální publikování výzkumů 2006-2010	Ano	Ano	Ano	Ano
Celkový index kvality predikce	4,1	4,3	3	4,6

Zdroje: (Chábová 2011, s. 32-44) (STEM 2010a) (CVVM 2010a) (MEDIAN 2010a) (Factum 2010a)

Pomineme-li vzhledem k charakteru vychýlení volebních odhadů možnost, že za větší úspěšností modelu společnosti Factum by stálo používání volebních karet,³³ nelze mezi hlavními parametry identifikovat žádný jednoznačný faktor, ani jejich kombinaci, která by vysvětlovala mírně rozdílnou úspěšnost jednotlivých agentur.³⁴ To se týká i údajů jako

²⁷ Kritérii myslíme sledované kvótní znaky, kterými v omnibusech agentur nesoucích volební výzkumy podle veřejných zdrojů jsou právě základní sociodemografické ukazatele (CVVM 2011f) (STEM 2011f) (Factum 2010d), či kategorie podle nichž probíhalo post-stratifikační vážení (Median – viz kapitola 3.2).

²⁸ Vzdálenost středního bodu sběru dat od prvního dne voleb (28.5.2010).

²⁹ STEM v posledním předvolebním výzkum uvádí, že 62 % respondentů avizuje účast a 24 % není rozhodnuto. Reálný odhad účasti se tak u STEM pohybuje mezi 62 % až 86 %.

³⁰ Zde máme na mysli jakékoli zahrnutí respondentů avizujících neúčast u voleb do volebního modelu (finálního predikčního produktu), a to buď plné zahrnutí respondentů odpovídajících na otázku o volební účasti „spíše ne“ (CVVM 2010a), či způsoby částečného / váženého zahrnutí (STEM, Factum) – více viz kapitola 3.4.

³¹ Zde máme na mysli zahrnutí preferencí voličů, kteří jsou nerozhodnutí (zvažují více stran), do volebního modelu. To podle veřejných zdrojů neaplikovala žádná z agentur (MEDIAN 2010a) (CVVM 2010a) (STEM 2010a) (Factum 2010a) – více viz kapitola 3.3.

³² Zohlednění minulých voleb či minulého volební chování ve finálních odhadech. Připouští jej STEM (2010a) a Factum (2010a).

³³ Tato metodologie je směřována k potlačení odpovědí podle sociální desirability (např. potlačení neochoty přiznat KSČM), ale těžko se jí dá zdůvodnit obecně lepší odhad Factum v poměru zisku koaličních stran a ČSSD (viz kapitola 2.1.1).

³⁴ Zde se musíme spoléhat hlavně na vlastní úsudek z daných dat. Statistické metody typu lineární regrese nelze regulérně aplikovat, protože nejsou splněny doporučení o minimální poměru predikovaných případů. Lineární regrese sice lze spočítat index kvality predikce lineární kombinací ostatních faktorů, ale právě jen díky

velikost vzorku (Factum a MEDIAN měly s menší velikostí vzorku kvalitnější predikci než STEM a CVVM), jejichž role bývá někdy při snaze o identifikaci zdrojů chyb odhadů přeceňována. (Shamir 1986)

Protože konkrétní detaily nutné k detailnější analýze prací jednotlivých agentur³⁵ nejsou z veřejných zdrojů známy, je lepší se zaměřit na danou otázku opačně. Tedy, jaké mohou být zdroje faktu, že agentury se mýlili sice v rozdílné míře, ale podobným směrem (přecenění ČSSD, podcenění celkového zisku budoucí koalice – viz kapitola 2.1.2) - co dané výzkumy spojuje?

Jistě se jedná o fakt, že střední doba sběru ani jednoho z výzkumů nezasáhla do období 14 dnů před volbami, kdy se finálně rozhodovalo okolo třetiny budoucích voličů (SC&C 2010b, s. 14). Zaměříme-li se však na faktory, které agentury mohou reálně ovlivňovat,³⁶ existuje shoda ještě v dalších oblastech – všechny agentury se spoléhaly čistě na sběr dat osobním dotazováním, v kvótně sbíraných omnibusech sledovaly (jen) reprezentativitu podle základních sociodemografických ukazatelů (CVVM 2011f) (STEM 2011f) (Factum2010d) či podle těchto ukazatelů vážily data (MEDIAN – viz kapitola 3.2) a nepracovaly při tvorbě finálních výstupů volebních průzkumů s preferencemi nerozhodnutých voličů. (CVVM 2010a) (STEM 2010a) (Factum 2010b) (MEDIAN 2010a)

Právě na možnosti optimalizace z těchto hledisek se ve velké části zaměřuje naše analýza na příkladu dat výzkumné společnosti MEDIAN z volebního roku 2010.

omezenému počtu predikovaných případů (přínos žádného z faktorů pro zvýšení vysvětleného rozptylu není statisticky významný).

³⁵ Např. kontrolní procesy v tazatelské síti, počet tazatelů a oblastí sběru, konkrétní postupy nakládání s daty při tvorbě modelů u STEM a Factum, atd.

³⁶ Tj. nejsou determinovány dobou nutnou pro kvalitní sběr a analýzu dat a existencí týdenního embarga na publikaci volebních výzkumů.

2.3 Faktory ovlivňující volební preference respondentů

2.3.1 Ideální typy volebního chování a jejich projevy v ČR

Pro základní systematizaci faktorů ovlivňujících volební chování občanů můžeme vyjít z klasifikace tří modelů volebního chování, tak jak je uvádí Vlachová (2000). Tyto modely lze považovat za weberovské ideální typy volebního chování:

- a) **Sociologický model** - základním determinantem volebního chování jedince příslušnost ke skupině (sociální třídy, etnické skupiny, náboženské a věkové skupiny, atd.). Předpokladem chování podle tohoto modelu je, že existuje silný vztah mezi zaměřením jednotlivých politických stran a zájmy jednotlivých skupin.
- b) **Model stranické identifikace (sociálně-psychologický model)** - základním determinantem volebního chování je stranická identifikace a emotivní loajalita voliče k dané straně. Chování podle tohoto modelu sice může vyrůstat z modelu a). „Na pozadí této identifikace stojí příslušnost k určité skupině a vliv rodiny.“ (Vlachová 2000, s. 9) Samotná stranická identifikace však nemá sociální příčiny, ale jedná se spíše o sociálně-psychologický konstrukt s emotivním charakterem. „Je založena na emotivně vnímané víře a hlubinných mentálních stereotypech.“ (Kunštát 2007, str. 4) Zjednodušeně řečeno: Dělník volí sociální demokracii čistě podle tohoto modelu chování, ji nevolí protože sociální demokracie se programově zaměřuje na sociální třídu dělníků, ale protože existuje stereotyp, že sociální demokracie je stranou zaměřující se na dělníky.
- c) **Model racionálního volby (ekonomický)** - volba strany je ad-hoc rozhodnutí učiněné na základě všech dostupných informací o programu stran, tedy strany, jejíž vítězství by maximalizovalo užitek voliče. Takový typ volby je racionálním výběrem strany, kde racionalitu chápeme ve smyslu klasické ekonomie.

Chápeme-li tyto modely jako ideální typy volebního chování, nelze konkrétní volební chování jednotlivce popsat jako čistou aplikaci jednoho z nich, ale naopak kombinaci těchto ideálních typů.

Pro účely naší studie je však podstatná otázka, jaké typy volebního chování v současné České republice dominují a nakolik v těchto typech hrají roli statisticky sledovatelné a měřitelné charakteristiky jednotlivců. Tedy například sociální, demografické a jiné charakteristiky voličů / respondentů.

Daniel Kunštát (2007) popisuje, že v poválečné historii západního světa sledujeme klesající vliv determinace volebního chování příslušností k sociální třídě a skupině (typ volebního

chování A) a narůstající vliv konstruované stranické identifikace (typ B). Důvodem tohoto procesu je podle Kunštáta a politických teorií, o které se opírá, zejména rozpad a oslabení tradičních masových stran, se silnou vazbou na sociální skupiny obyvatelstva - dělnické strany, agrární strany, atd. Tento rozpad byl způsoben částečně oslabováním poptávky po nich (oslabování tradičních skupin dělníků, zemědělců apod. v post-industriální společnosti), ale také změnami na straně nabídky - orientace poválečné politiky na sociální konsensus a oslabování ideologických konfliktů (Kunštát 2007, s.3) Komplementárním jevem je posilování tzv. catch-all stran – částečně odideologizovaných „všelidových a univerzálních stran“ (Kunštát 2007, str.3), které se neprofilují jako propagátor zájmů sociálních tříd a skupin, ale naopak rozšiřují paletu svých volebních témat a vytváří si širší volební potenciál napříč sociálními a demografickými skupinami. Volba takových stran je pak často závislá na právě na stranické identifikaci v sociálně-psychologickém smyslu (typ volebního chování B).

Ideální typ C, kdy se jednotlivec chová racionálně (ve smyslu klasické ekonomie) a ad-hoc analyzuje přínosy nabízených volebních programů, se dá chápat jako alternativa volby podle emotivní stranické identifikace v prostředí, kde se již rozpadly jednoznačné vazby mezi sociální skupinami/třídami a politickými stranami.

Řada analytiků upozorňuje, že vliv racionální volby na volební chování nelze přeceňovat. Například Caplan ve své známé studii *Mýtus racionálního voliče* dochází k závěru, že většina voličů racionální volbu neaplikuje, protože očekávaný vliv vlastního hlasu na volební výsledek (tj. zisk z racionální volby) je u voličů menší než námaha a čas vynaložené na pochopení a racionální analýzu volebních program (tj. náklady racionální volby).³⁷ Chovají se tedy „racionálně“ již v předcházejícím kroku – zvažování racionální volby.

Pokud se ptáme, jak se tedy čeští voliči ve volbách poslední doby rozhodovali, narazíme však na problém. Výzkumy totiž popisují, že od 90. let klesá intenzita stranické identifikace (CVVM 2011c). Zároveň klesá identifikace voličů s tradičními sociálně-ideologickými proudy (CVVM 2001d), což naznačuje platnost Kunštátovy analýzy poklesu tradiční sociální determinace volebního chování v ČR (viz výše). Pokud zároveň v ČR platí Caplanovy

³⁷ Tento Caplanův závěr se dá vnímat jako přínosný, i bez ohledu na jeho další konkretizaci v níž se projevuje Caplanův libertariánský ideologický bias - tj. například chápání nedůvěry v tržní princip nutně jako voličského iracionálního předsudku (Caplan 2007, str. 31-33), bez ohledu na zkušenost voličů s výsledky politiky zaštiťované neoliberálními tržními principy (Kleinová 2010). Tento problém se však zdá u *modelu racionální volby* klíčový obecně – ten totiž předpokládá racionální volbu jako opakované ad-hoc rozhodnutí jednotlivce (Vlachová 2000, s. 9), čímž eliminuje možnost zahrnout do racionální úvahy zkušenost s výsledky reálné aplikace politických programů, zkušenost s možností jejich aplikace (to může být faktorem například u protikorupčních programů, nad jejichž potřebností jinak panuje konsensus) a pravděpodobnost, že daná strana bude skutečně vyvíjet snahu o naplnění programu. Zahrnutí těchto faktorů do racionální volby by zřejmě snižovalo teoretickou eleganci modelu racionální volby voličů, ale zvyšovalo jeho komplexitu a explanační schopnost.

obecné závěry o subjektivní nevýhodnosti a tedy časté absenci racionální volby, zůstává otázkou, podle čeho se voliči reálně rozhodují.

Oslabením všech tří základních ideálních typů volebního chování lze považovat za jeden z důvodů: a) klesajícího zájmu občanů ČR o politiku obecně; b) poměrně silného zastoupení voličů ve volbách 2010, kteří se rozhodovali tzv. negativní či taktickou volbou³⁸ (STEM 2010b) či podle mediální kampaně a vystupování kandidátů v TV debatách před volbami (Factum 2010c); c) nárůstu počtu voličů, kteří se pro jednu z kandidujících stran rozhodují v posledních dnech před volbami (SC&C 2010b).³⁹

Pokud by byly druhé dva důsledky dlouhodobé, výrazně se tím omezuje obecná šance předvolebních výzkumů v době 1 měsíc – 14 dní před volbami seriózně predikovat jejich výsledek. Nárůst rozhodování na poslední chvíli a uplatnění ad-hoc rozhodovacích mechanismů se však dá také zdůvodnit tím, že v roce 2010 bylo oproti roku 2006 větší množství stran, které měly podle předvolebních průzkumů šanci proniknout do Sněmovny,⁴⁰ z nichž některé vznikly odštěpením ze stran tradičních (TOP09) a jiné byly nově zformovány (VV), čímž se zejména středo-pravicově orientovaným voličům zkomplikovali volební rozhodování.

V kapitole (3.3) se pokusím ukázat, že vhodnou konstrukcí volebního modelu se zařazováním respondentů dosud nerozhodnutých pro jednu stranu, se dá chybě predikce v takto komplikované volební situaci částečně předejít.

2.3.2 Faktory ovlivňující politické preference v ČR 2010

Hlavní otázkou této kapitoly je, jaké sledovatelné a měřitelné faktory (sociální, demografické a jiné) mají souvislost s výběrem strany ve volbách do Sněmovny. Nelze totiž automaticky přepokládat, že oslabováním ideálního typu volby strany podle příslušnosti k sociální třídě / skupině (viz výše) se oslabí vztah sociálních/demografických faktorů a volebního chování.

³⁸ Negativní volbou se rozumí volba např. proti vládě pod vedením ČSSD. Taktickou volbou taková, která zvyšuje nikoli pravděpodobnost na vítězství volené strany, ale pravděpodobnost ustavení preferované koalice (např. volba VV místo ODS, při které se průnikem této strany do parlamentu zvýší šance na pravicovou koalici, apod.)

³⁹ V roce 2006 bylo více než 2 měsíce před volbami rozhodnuto 63 % voličů, v roce 2010 jen 51 %. Nárůst nastal i v počtu voličů, kteří se rozhodovali těsně před volbami. (SC&C 2010b).

⁴⁰ V roce 2010 mělo celkem 8 stran v předvolebních průzkumech nad 3 % - kromě současných parlamentních stran také KDU-ČSL, SPOZ a Stranu zelených. V roce 2006 mělo pouze 5 stran (ODS, ČSSD, KSČM, KDU-ČSL, SZ) v průzkumech nad 3 %. Viz Kapitola 2.1.

Tyto faktory se totiž projevují i při konstruování stranické identifikace⁴¹ (typ B) i při racionální ad-hoc volbě volené strany⁴².

2.3.2.1 Souvislost volebních preferencí se základními sociodemografickými faktory

Při analýze dat z osobního sběru volebních preference agentury MEDIAN jsme se zaměřili na období červenec – prosinec 2010. Tedy na období relativně aktuální, které ale zároveň popisuje vztah sociální, demografických a jiných faktorů k politickým stranám na politické scéně, jak ji definovaly volby do PSP ČR 2010. Vycházíme z nevážených dat. Zaměřujeme se na respondenty ve věku 18+, kteří vstupují do volebního modelu (avizují účast u voleb a jsou rozhodnutí, koho volit). Základní sociodemografické faktory křížíme přes otázku na aktuální volební preference (tedy ne na stranu volenou v minulých volbách).⁴³

Základními sociodemografickými faktory přitom rozumíme veličiny, které jsou v preferenčních výzkumech agentur používajících kvótní výběr využívány jako kvótními znaky (CVVM 2011f) (STEM 2011f) (Factum2010d). U společnosti MEDIAN, která aplikuje víceetapový stratifikovaný náhodný adresní výběr, tyto veličiny vstupovaly v preferenčních výzkumech v roce 2010 do post-stratifikačního vážení.⁴⁴ Jedná se o proměnné: pohlaví, věk, kraj / region, velikost obce, nejvyšší dosažené vzdělání.⁴⁵

TAB 2.3.1 Síla vztahu základních sociodem. ukazatelů a volebních preferencí

	Contingency Coefficient	Phi	Chi-Square Sig
Věk	,426	,471	,000
Kraj	,372	,401	,000
Vzdělání	,327	,346	,000
Velikost obce	,178	,181	,000
Pohlaví	,080	,080	,002

Zdroj: Data volebního sběru MEDIAN červenec – prosinec 2010. Nevážená. Respondenti avizující účast u voleb, pokud by se konaly nyní, preferující jednu stranu (n=3794)

Analýza na datech MEDIAN (TAB 2.3.1) ukazuje, že volební preference nejsilněji závisí na věku respondenta, kraji bydliště (NUTS3) a nejvyšším dosaženém vzdělání. Závislost na

⁴¹ Jakkoli může být tato konstrukce částečně emotivní a stereotypní a nebrat ohled reálný program stran, lze předpokládat, že při ní volič vychází i z reflexe vlastního postavení ve společnosti.

⁴² Volič racionálně hodnotí přínos programu strany pro sebe a vlastní rodinu, což lze těžko dělat bez ohledu na vlastní ekonomické a sociální postavení ve společnosti.

⁴³ Srovnání s minulým volebním chováním by nebylo regulérní, protože u řady sociodemografických faktorů mohlo dojít v době mezi volbami 2010 a dotázaním respondenta ke změně (věk, kraj, VMB i vzdělání). Do přiznávání volené strany se navíc v době krátce po volbách mohou projevit vlivy band-wagon efektu, neochoty přiznat neúčast u voleb, atd. (viz Kapitola 3.1)

⁴⁴ MEDIAN v preferenčních výzkumech před volbami v roce 2010 vážil data podle pohlaví, věku (7 kategorií), vzdělání (4 kategorie), kraje NUTS3, velikosti obce (5 kategorií) a podle kombinací faktorů pohlaví x věk, pohlaví x vzdělání a velikost obce x region NUTS2.

⁴⁵ Agentury se drobně odlišují v úrovni s jakou podrobností kvótují – například Factum využívá proměnnou kraje (NUTS3) a CVVM proměnnou regionů NUTS2, apod.

velikosti obce/města, v němž respondent žije, je slabší a závislost na pohlaví sice stále signifikantní, ale oproti ostatním faktorům poměrně zanedbatelná.

Konkrétní podoba zjištěných závislostí základních sociodemografických faktorů a volebních preference (TAB 2.3.2) do velké míry kopíruje zjištění z exit-pollu SC&C v roce 2010 a analýzu Daniela Kunštáta (2007), pokud ji vztáhneme na aktualizované spektrum hlavních stran pro rok 2010. Podpora pravice (ODS+TOP09) pochází zejména od lidí do 44 let, přičemž v nejmladších ročnících (do 24 let) je preferována TOP 09. Podpora pravice je silnější ve velkých městech (nad 100 obyvatel) a mezi vysokoškoláky. Z regionálního hlediska má pravice nadstandardně silné postavení zejména v oblasti Praha a střední Čechy.

Podpora levicových stran (ČSSD + KSČM) pochází častěji od starších lidí (nad 55 let), přičemž se stoupajícím věkem roste podpora zejména u KSČM. Podpora levice je také vyšší u skupin s nižším vzděláním. Z regionálního hlediska má v nevážených datech KSČM nejsilnější postavení ve zbytku Čech (Čechy s vyjmutím Prahy a Středočeského kraje). Podpora ČSSD je regionálně poměrně homogenní.

Z ostatních silných stran se Věci veřejné a Strana zelených strukturou voličů podobají pravici v tom, že je preferují mladší lidé, a Věci veřejné také silnou podporou v Praze a středních Čechách. KDU-ČSL se demografickou strukturou voličů podobá levici v tom, že ji preferují nejstarší věkové skupiny (55+) a navíc má silné postavení na Moravě a v malých obcích – tedy v místech s větší mírou religiozity. Podobné jevy, i když ve slabší míře, platí u SPOZ.

Statistickou souvislost volebních preferencí a jednotlivých sociodemografických faktorů samozřejmě nelze vykládat jako kauzální vztah - např. v preferenci levice méně vzdělanými lidmi hraje výraznou roli významnou roli vztah vzdělání a finančních příjmů respondenta a jeho domácnosti a také vztah vzdělání a stáří (větší zastoupení méně vzdělaných lidí mezi nejstarší generací), nelze tedy vykládat, jak se někdy stává při mediálních zkratkách, jako kauzální vztah vzdělání (rozhledu, intelektu) a vzdělanosti s politickými preferencemi.

Více o konkrétní struktuře vztahů vypoví kapitola 2.3.3 – predikce volební preference pomocí klasifikačních a regresních stromů. V otázce, zda respondent preferuje strany pravicové koalice (ODS+TOP09+VV), levicové opozice (ČSSD+KSČM) či strany jiné, se ukazuje, že vzdělání hraje spíše druhotnou roli a pro zkvalitnění predikce je vhodné jen v určitých sub segmentech obyvatel a ani zde není finálním predikujícím faktorem. (více viz kapitola 2.3.3)

TAB 2.3.2 Podoba vztahu základních sociodemografických ukazatelů a volebních preferencí

STRANA	TOTAL	pohlaví		věk						vzdělání			
		muži	ženy	18 - 24 let	25 - 34 let	35 - 44 let	45 - 54 let	55 - 64 let	65 - 79 let	základní	vyučen(a) / SŠ bez maturity	SŠ s maturitou	vysokoškolské
ČSSD	31,0%	30,1%	31,8%	16,0%	18,4%	26,6%	34,0%	41,2%	33,9%	28,4%	39,5%	28,3%	15,1%
ODS	20,2%	22,1%	18,6%	24,7%	32,1%	26,2%	24,9%	14,4%	8,6%	10,9%	11,7%	24,6%	36,1%
TOP 09	12,9%	12,0%	13,6%	31,5%	22,8%	16,8%	11,9%	6,2%	6,2%	8,3%	7,8%	15,8%	21,2%
KSČM	13,1%	14,2%	12,1%	,9%	2,2%	3,6%	8,3%	16,5%	30,6%	29,7%	17,8%	8,1%	6,7%
VV	7,5%	6,2%	8,6%	11,9%	11,4%	10,0%	7,5%	5,4%	4,2%	4,8%	5,6%	9,6%	8,3%
KDU-ČSL	5,7%	5,0%	6,2%	1,8%	2,6%	4,5%	3,5%	7,7%	9,3%	8,7%	6,6%	4,8%	3,9%
SPOZ	3,3%	3,8%	2,8%	,9%	2,4%	4,0%	2,5%	3,5%	4,2%	1,7%	3,9%	3,0%	2,8%
SZ	3,4%	3,3%	3,4%	6,4%	4,2%	4,2%	4,1%	2,6%	1,6%	3,9%	3,4%	3,2%	3,5%
Ostatní	3,1%	3,3%	2,9%	5,9%	3,8%	4,0%	3,3%	2,5%	1,5%	3,5%	3,7%	2,6%	2,4%
Total	100%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

STRANA	TOTAL	oblast			velikost místa bydliště				
		Praha a Střední Čechy	zbytek Čech	Morava	do 999 obyvatel	1 000 - 4 999 obyvatel	5 000 - 19 999 obyvatel	20 000 - 99 999 obyvatel	100 000 obyvatel a více
ČSSD	31,0%	30,1%	31,2%	31,3%	27,7%	32,2%	31,4%	34,6%	28,6%
ODS	20,2%	24,9%	16,4%	21,4%	18,7%	20,6%	18,4%	15,9%	26,9%
TOP 09	12,9%	18,1%	12,6%	10,2%	11,4%	11,7%	11,0%	11,9%	17,5%
KSČM	13,1%	7,4%	16,5%	12,7%	14,6%	14,3%	15,0%	14,8%	7,5%
VV	7,5%	9,5%	7,4%	6,5%	8,7%	6,5%	7,4%	8,1%	7,0%
KDU-ČSL	5,7%	3,7%	3,7%	8,8%	9,0%	6,1%	5,1%	4,1%	4,9%
SPOZ	3,3%	2,9%	3,9%	2,9%	4,9%	2,7%	4,2%	2,8%	2,3%
SZ	3,4%	1,9%	3,6%	3,9%	2,6%	3,5%	3,3%	4,1%	3,2%
Ostatní	3,1%	1,5%	4,7%	2,2%	2,6%	2,6%	4,3%	3,8%	2,0%
Total		100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Zdroj: Data volebního sběru MEDIAN červenec – prosinec 2010. Nevážená. Respondenti avizující účast u voleb, pokud by se konaly nyní, preferující jednu stranu (n=3794)

Pozn.: Červeně vybarvená pole o více než 25 % (nikoli p.b.) překračují zisk strany v celkovém volebním modelu. Modře vybarvená o více než 25 % nedosahují hodnoty v celkovém modelu. Červený / modrý font značí statisticky významnou odlišnost ($p=0,05$) od hodnoty v celkovém modelu.

V rámci základních sociodemografických je vhodné se zaměřit i na vliv jejich kombinací⁴⁶, které mohou být mimo jiné předmětem optimalizace post-stratifikačního vážení (viz Kapitola 3.2).

TAB 2.3.3 Síla vztahu kombinovaných zákl. sociodemografických ukazatelů a volebních preference

Kombinace...	Contingency Coefficient	Phi	Chi-Square Sig
vek a region NUTS 2	,497	,573	,000
vek a VMB	,472	,536	,000
vek a vzdělání	,463	,522	,000
vek a pohlaví	,432	,478	,000
vzdělání a VMB	,387	,419	,000
vzdělání a region ⁴⁷	,378	,409	,000
region NUTS2 a VMB	,346	,368	,000
pohlaví a vzdělání	,340	,361	,000
pohlaví a regionu NUTS2	,293	,306	,000
pohlaví a VMB	,208	,212	,000

Zdroj: Data volebního sběru MEDIAN červenec – prosinec 2010. Nevážená. Respondenti avizující účast u voleb, pokud by se konaly nyní, preferující jednu stranu (n=3794)

Všechny kombinované sociodemografické faktory s volební preferencí statisticky významně souvisí (sig.<0,01 se splněním podmínek chi-square testu). Považujeme-li za kritérium síly vztahu to, že vztah by měl být silnější než u silně působících nekombinovaných základních sociodemografických faktorů (viz výše), dají se za kategorie silně svázané s volebními preferencemi respondentů považovat všechny kombinované faktory, na nichž se podílí věk (tedy věk v kombinaci se *vzděláním*, *regionem*, *velikostí obce* a *pohlavím*). Tento jev se dá bez pomoci rozhodovacích stromů vysvětlit například tak, že uvnitř jednotlivých segmentů populace určených *regionem* / *vzděláním* / *pohlavím* / *velikostí obce* dále politické preference výrazně závisí na věku respondentů.

2.3.2.2 Souvislost voleb. preferencí se specifickými sociodemografickými a life-stylovými faktory

Silněji než s některými ze základních sociodemografických veličin (viz TAB 2.3.1), které jsou standardně předmětem kontrol reprezentativity vzorků a následného vážení či kvótních předpisů, a jejich kombinacích, jsou volební preference respondentů svázané se specifickými sociodemografickými proměnnými. (viz TAB 2.3.4) Za poměrně kvalitní prediktor volební

⁴⁶ V programu SPSS byly při analýze dat vytvořeny proměnné kombinující všechny hodnoty dvou daných proměnných (např. muž 18-24 let, žena 18-24 let, muž 25-34 let, žena 25-34 let, atd.). Tyto proměnné byly zkříženy s údajem o volebních preferencích. Pro účely této analýzy byly sloučeny neparlamentní strany do jedné hodnoty „ostatní strany“, v opačném případě totiž většina kontingenčních tabulek obsahuje více než 20 % polí s hodnotou menší než 5.

⁴⁷ Pro analýzy vlivu kombinované sociodemografické kategorie region x vzdělání byla použit hrubší geografická klasifikace než NUTS2, a to Praha+Stř.Čechy, zbytek Čech, Morava. Při rozdělení na 32 kategorií (8x4) a 9 možností volby, nejsou splněny podmínky testu dobré shody o minimálních očekávaných četnostech.

preferenze se dá považovat zejména *pracovní zařazení*⁴⁸ respondenta a jeho *osobní výše příjmů a výše příjmů jeho domácnosti*. Je přitom vhodné v úvahu brát obě tyto příjmové charakteristiky, které jinak popisují ekonomické zázemí respondenta, a v jistých případech se informačně doplňují.⁴⁹

TAB 2.3.4 Síla vztahu volebních preferencí a specifických sociodem. a life-stylových faktorů

	Contingency Coefficient	Phi	Chi-Square Sig
Pracovní zařazení	,437	,486	,000
Síla užívání internetu	,389	,422	,000
Čistý měsíční příjem respondenta	,330	,350	,000
Čistý měsíční příjem domácnosti	,328	,347	,000
Čistý měs. příjem domácnosti na člena ⁵⁰	,275	,286	,000
Preference ve volnočasových aktivitách	,245	,252	,000
Počet členů domácností	,209	,214	,000

Vliv life-stylových faktorů – intenzita užívání internetu a preference trávení volného času

Neplatí přitom námitka, že se jedná jen o „falešnou korelaci“ přes základní sociodemografické proměnné (např. přes věk a vzdělání). *Pracovní zařazení* statisticky významně ($\text{sig chi-square} < 0,05$) souvisí s volebními preferencemi uvnitř všech skupin vzdělání, uvnitř ve 4 ze 6 věkových skupin (konkrétně u lidí mezi 25-64 let)⁵¹ i uvnitř dalších základních sociodemografických kategorií.

To samé platí *čistém příjmu respondentovy domácnosti*, která s preferencemi souvisí uvnitř všech vzdělanostních a věkových skupin. A rovněž o *čistém příjmu respondenta*, který

⁴⁸ Ve srovnání v TAB 2.3.4 používáme 9hodnotovou proměnnou pracovního statusu (1 - samostatně výdělečně činný(á), podnikatel bez zaměstnanců, 2 - podnikatel se zaměstnanci, 3 - zaměstnanec - řadový pracovník, 4 - zaměstnanec - vedoucí pracovník, 5 - nezaměstnaný, 6 - nepracující důchodce, 7 - v domácnosti, rodičovská dovolená, 8- student, žák, učeň, 9 – jiné).

⁴⁹ Čistě z příjmu respondenta můžeme za „bohatého“ mylně označit samoživitele velké rodiny, která má v přepočtu na člena ve skutečnosti. Tj. reálná hodnota této proměnné pro usuzování na ekonomické zázemí respondenta závisí na tom, kolik osob je na jeho příjmu ekonomicky závislých. Hodnota příjmu domácnosti může být ovšem také zkreslující bez informace, kolik členů domácnost má (tj. ekonomické zázemí 4členné domácnosti s příjmem 25 tisíc je jiné než ekonomické zázemí jednočlenné domácnosti s příjmem 25 tisíc). Čistý měsíční příjem domácnosti tyto negativa eliminuje.

⁵⁰ Pro popis závislosti preferencí na ekonomickém zázemí respondenta jsme se pokusili použít rovněž proměnnou příjmu domácnosti přepočteného na jednoho člena. Tato proměnná byla spočtena tak, že kategorie příjmu domácnosti byly nahrazeny průměrem (např. kategorie 8 001 – 10 000 Kč hodnotou 9 000 Kč), který byl vydělen počtem členů v domácnosti. Výsledná pseudokardinální proměnná byla kategorizována (viz kategorie v TAB 2.3.5). Důvodem, proč má tato proměnná menší vztah s preferencemi může být fakt, že spojuje do jedné kategorie sociodemograficky velmi odlišné části populace – např. samostatně žijícího důchodce s důchodem 10 000 a člena 4členné domácnosti se poměrně vysokým čistým měsíčním příjmem 40 000.

⁵¹ Mezi mladými (velké množství studentů) a starými (velké množství důchodců) se schopnost diverzifikace ztrácí.

s volebními preferencemi významně souvisí velké většiny kategorií věku a vzdělání.⁵² Kategorie *příjmu a pracovního zařazení* navíc nejsou ani plně vzájemně nahraditelné.⁵³

Faktorem silněji vázaným s politickými preferencemi silněji než některé základní sociodemografie (pohlaví a velikosti místa bydliště) a jejich kombinace, je i ukazatel, v jak velké rodině respondent žije. Statistický vztah je zde sice částečně „falešnou korelací“ přes proměnné *věku* (domácnosti starších respondentů mají častěji méně členů) a *příjem domácnosti na člena* (větší domácnosti mají ceteris paribus menší příjem na člena), ale to opět pouze jen do jisté míry.⁵⁴

Do analýzy jsme zařadili také dvě proměnné, které mají spíše life-stylový, než sociodemografický charakter. Zkoumali jsme, nakolik volební preference souvisí s *intenzitou užívání internetu* a s faktem, zda respondent preferuje *volnočasové aktivity vykonávané doma či venku*.

Zařazení prvního faktoru je relevantní zejména díky předpokladu, že lidé častěji využívající internet budou mít jinou strukturu informačních zdrojů o politice, což může ovlivnit jejich preference. Zároveň je také intenzita využívání internetu jedním z life-stylových veličin, které by mohly vstupovat do ex-post kontroly reprezentativity vzorku a potenciálně i do post-stratifikačního vážení. Existuje totiž důvěryhodná externí⁵⁵ datová opora o intenzitě užívání internetu českou populací v podobě dat Eurostatu.

Intenzita užívání internetu souvisí s volebními preferencemi velmi silně ($\Phi = 0,422$ – což je větší hodnota než v případě všech základních sociodemografických ukazatelů kromě věku – viz TAB 2.3.1). Opět tomu přitom není jen díky „falešné korelaci“ přes věk, vzdělání či příjem respondenta – tj. přes proměnné, které souvisí s intenzitou využívání internetu a zároveň s politickými preferencemi (viz výše). Intenzita užívání internetu významně souvisí s volebními preferencemi uvnitř všech 6 věkových kategorií, uvnitř všech 4 kategorií vzdělání i uvnitř všech kategorií proměnné příjmu domácnosti na člena. Intenzita užívání internetu (a s ní spojená struktura informačních zdrojů, životní styl či účast na kampaních vedených po

⁵² Výjimkou jsou základní školáci a lidé do 24 let – zřejmě díky poměrně nízké diverzifikaci příjmů v těchto skupinách.

⁵³ Například i mezi důchodci výše měsíčního příjmu domácnosti statisticky významně souvisí s preferencemi.

⁵⁴ Signifikantní vztah mezi počtem členů domácnosti a preferencemi existuje uvnitř 3 z 5 kategorií proměnné čistý měsíční příjem domácnosti na člena domácnosti a uvnitř 3 ze 6 věkových kategorií. Přímý vliv počtu členů domácnosti na preference respondenta lze vysvětlit například odlišnou orientací respondentů z různých početných domácností na prarodinnou politiku či rozdílnou afinitou ke konzervativním politickým hodnotám, které jsou spojeny se vztahem k rodině.

⁵⁵ Problém u jiných typů media-exposure je, že jediná důvěryhodná datová opora pro vážení a kontrolu reprezentativity pochází rovněž od agentury MEDIAN, a to z výzkumů Media projekt a Radio projekt, které jsou oficiální měnou pro určování poslechovosti a čtenosti médií v ČR. Díky tomu, že tato datová opora není z hlediska svého původce a aplikované metodologie, zcela nezávislá na vážených datech, nebylo by její využití pro vážení a kontrolu reprezentativity zcela regulérní.

sociálních sítích) tedy pravděpodobně volební preference sama ovlivňuje. Je s ní tedy nutno uvažovat při analýze reprezentativity vzorku a potenciálně i při vážení dat.

Toto zjištění je mimo jiné i významným zpochybněním snahy některých výzkumných společností o sběr preferenčních výzkumů čistou CAWI metodou. On-line metoda dotazování totiž a-priory omezuje zkoumanou populaci na část, která má významně jiné politické preference, než obecná populace voličů (viz TAB 2.3.6). K relevantním výsledkům tedy CAWI výzkum může docházet jen za předpokladu, že volební účast v internetové populaci je reálně významně vyšší než v neinternetové, čímž se zvyšuje její vliv na výsledek voleb, či použitím korekcí pro redukci systematického vychýlení výzkumu.⁵⁶

Druhý z faktorů – *preference domácích / venkovních volnočasových aktivit* – je pro analýzu relevantní, protože se může jednat o zdroj vychýlení volebních výzkumů prováděných osobním dotazováním v domácnostech. Lze předpokládat, že pokud respondent vybraný Kishovou tabulkou v domácnosti vybrané vícestupňovým stratifikovaným výběrem (více viz Příloha 1), často absentuje v domácnosti, sníží to (při nemožnosti substituce na úrovni respondentů) i při povinnosti opakované návštěvy šanci zahrnutí takového respondenta do výzkumu.⁵⁷

Omnibusové CAPI šetření, kterým společnost MEDIAN dotazuje i volební preference, obsahuje mimo jiné baterii 27 otázek na četnost vykonávání různých druhů volnočasových aktivit se škálou odpovědí 1 až 6.⁵⁸ U celkem 7 z těchto činností se dá předpokládat, že je respondent drtivou většinou vykonává doma. U 13 se dá předpokládat, že je respondent drtivou většinou vykonává mimo domov. Zbývajících 7 nelze rozřadit. Z odpovědí respondenta na otázky o intenzitě vykonávání volnočasových jsme v analýze vytvořili jeden index, který určuje celkovou náklonost respondenta k trávení volného času doma a venku. Tato pseudo-kardinální proměnná nabývající teoreticky hodnot v intervalu (-5;5) byla kategorizována na pěti-stupňovou typologii respondenta s krajními kategoriemi 1 „*velmi domácí*“ a 5 „*velmi venkovní*“.⁵⁹

⁵⁶ Nabízí se například nadhodnocování respondentů, kteří sice internet používají (a mohou odpovídat CAWI), ale to jen zřídka, vážením - což by mohlo suplovat vliv neinternetové populace na výsledky preferenčního výzkumu. Popřípadě přepočítání volebního modelu z CAWI podle předem provedeného face-to-face výzkumu, na jehož základě výzkumníci kvantifikují nakolik a jakými směry se CAWI výzkum od reprezentativních výzkumů vychyluje. Z veřejně dostupných informací však není zřejmé, že by společnost SANEP takové korekce prováděla.

⁵⁷ A to díky faktické absenci v domácnosti i kvůli tomu, že tento typ respondentů může projevoval menší ochotu trávit čas v domácnosti odpovídáním na otázky tazatele.

⁵⁸ 1=Denně, téměř denně; 2=Alespoň 1x týdně; 3=Alespoň 1x měsíčně; 4=Alespoň 1x ročně; 5=Méně často; 6=Nikdy.

⁵⁹ Index byl vytvořen odečtením průměrné hodnoty respondenta u venkovních volnočasových aktivit od průměrné hodnoty u domácích aktivit. Výsledná pseudo-kardinální proměnná byla kategorizována, tak aby dělila vzorek na kvintily.

Potvrzuje se přitom, že *preference ve volnočasových aktivitách* skutečně s volebními preferencemi respondenta významně souvisí (viz TAB 2.3.4) a to opět silněji než některé ze základních sociodemografických ukazatelů a jejich kombinací (TAB 2.3.1, TAB 2.3.2). Opět přitom neplatí námitka, že by *preference ve volnočasových aktivitách* souvisela s volební preferencí jen jako „falešná korelace“ přes proměnnou věku, příjmu či popřípadě pracovního zařazení respondenta.⁶⁰ Proměnná preference ve volnočasových aktivitách totiž s volebními preferencemi významně souvisí uvnitř všech kategorií čistého měsíčního příjmu respondenta, uvnitř segmentu ekonomicky aktivních i ekonomicky neaktivních respondentů⁶¹ a také uvnitř poloviny věkových kategorií.⁶² Lze tedy předpokládat, že i ona přináší u jistých respondentů obtížně nahraditelnou informaci o životním stylu respondenta (a s tím souvisejícím hodnotové orientaci; struktuře osob, s nimiž respondent komunikuje; atd.), která se projevuje do volebních preferencí.

Tabulka 2.3.5 ukazuje konkrétní podobu souvislostí specifických sociodemografických a life-styleových faktorů s volebními preferencemi respondentů. V otázce příjmů a pracovního zařazení respondentů se potvrzují dřívější zjištění jiných analýz (Kunštát 2007).

Elektorát pravice (ODS + TOP 09) je situovaný do vyšších příjmových skupin. Přičemž tento vztah je silnější u ODS, než u TOP 09.⁶³ Voliči ODS jsou nejsilněji zastoupeni zejména v nejvyšších příjmových skupinách (příjem respondenta nad 25 000 čistého měsíčně, příjem domácnosti nad 40 000 čistého, příjem domácnosti na člena 20 000). Ve vyšších středních příjmových skupinách pak občanským demokratům silněji konkuruje TOP09 i další středopravé politické strany (VV). Z hlediska pracovního zařazení, je elektorát pravice je silný mezi podnikateli / OSVČ a mezi studenty, přičemž platí, že v prvním z těchto segmentů je silná zejména Občanská demokratická strana a mezi studenty TOP09.

⁶⁰ Tyto veličiny byly vybrány jako testovací, protože mají významně souviset s *preferencí volnočasových aktivit*.

⁶¹ Vliv nemá jen mezi studenty, kde jsou minimálně zastoupeny respondenti preferující domácí volnočasové aktivity.

⁶² Jde o kategorie 35-44, 45-54 a 65-79 let. V kategorii 55-64 je chi-square test zamítá hypotézu o shodě (nesouvislosti) jen na hladině spolehlivosti 90 %. Obecně se tak dá říci, že preference volnočasových aktivit je predikujícím faktorem mezi respondenty středního a staršího věku.

⁶³ Tento jev lze vysvětlit zejména menší zakotveností a tradicí TOP 09 ve stranickém systému, která se může projevit slabším vztahem strany k cílovým sociálním skupinám (sociologický model volby) a nižší stranickou identifikací k TOP09 a tedy vyšším poměrem ad-hoc volebního chování, které za předpokladu omezené racionální analýzy programu (Caplan) preferuje rozhodování podle image, osobností a vystupování v TV debatách, atd.

TAB 2.3.5 Podoba vztahu volebních preferencí a specifických sociodemografických a lifestyleových faktorů

STRANA	TOTAL	Čistý měsíční příjem respondenta (Kč)					Čistý měsíční příjem respondentovy domácnosti (Kč)					Čistý měs. příjem domácnosti na člena (Kč)				
		do 10 tisíc	10 001 - 15 000	15 001 - 25 000	nad 25 000	odmítl(a) odpovědět	do 20 tisíc	20.001 - 30.000	30.001 - 40.000	nad 40.000	odmítl(a) odpovědět	do 10000	10001 - 15000	15001 - 20000	20001 a více	odmítl(a) odpovědět
ČSSD	31,0%	32,9%	37,5%	27,3%	20,0%	22,0%	37,9%	36,4%	29,5%	20,4%	24,8%	35,7%	33,7%	26,6%	16,7%	24,8%
ODS	20,2%	12,7%	14,1%	23,2%	40,0%	38,0%	8,8%	15,2%	21,4%	33,1%	31,1%	13,4%	16,9%	25,3%	38,5%	31,1%
TOP 09	12,9%	12,7%	7,3%	17,1%	18,1%	17,1%	7,0%	9,5%	15,0%	20,7%	15,4%	10,4%	10,9%	15,0%	23,9%	15,4%
KSČM	13,1%	18,0%	20,0%	5,7%	2,7%	5,4%	23,5%	15,6%	8,3%	4,2%	10,2%	15,9%	15,9%	7,8%	1,7%	10,2%
VV	7,5%	6,4%	5,6%	10,3%	8,8%	7,3%	5,1%	6,9%	10,0%	8,8%	7,0%	7,4%	6,8%	8,1%	10,6%	7,0%
KDU-ČSL	5,7%	8,0%	5,7%	4,3%	3,7%	6,3%	7,9%	5,3%	4,9%	4,6%	5,5%	6,6%	6,0%	4,7%	2,6%	5,5%
SPOZ	3,3%	3,4%	3,0%	4,5%	1,9%	,5%	3,7%	3,8%	3,8%	2,6%	1,0%	3,6%	3,6%	4,7%	1,4%	1,0%
SZ	3,4%	3,0%	3,5%	4,5%	2,4%	,5%	2,9%	4,1%	4,3%	2,5%	2,3%	3,3%	3,2%	4,9%	3,4%	2,3%
Ostatní	3,1%	2,8%	3,3%	3,2%	2,4%	2,9%	3,3%	3,2%	2,8%	3,0%	2,6%	3,8%	3,1%	2,9%	1,1%	2,6%

STRANA	TOTAL	Počet clenu domácnosti					Síla užívání internetu			Preference ve volnočasových aktivitách				
		1 clen	2 clení	3 clení	4 clení	5 a více clenu	Silní uživatelé	Slabí uživatelé	Neuživatelé	velmi domácí	spíše domácí	neutrální	spíše venkovní	velmi venkovní
ČSSD	31,0%	32,9%	35,5%	28,7%	27,2%	26,0%	20,1%	30,8%	37,1%	41,3%	34,1%	31,4%	27,6%	20,0%
ODS	20,2%	15,0%	16,1%	25,4%	23,8%	22,0%	33,9%	24,0%	11,0%	12,2%	17,2%	18,9%	24,4%	28,8%
TOP 09	12,9%	11,8%	9,1%	16,6%	15,1%	14,0%	22,7%	13,3%	7,3%	8,9%	9,8%	15,1%	14,4%	16,4%
KSČM	13,1%	18,1%	17,5%	9,3%	8,3%	10,0%	2,6%	6,9%	21,4%	17,2%	15,4%	12,9%	11,2%	8,3%
VV	7,5%	6,8%	6,4%	6,5%	9,4%	11,3%	9,4%	9,9%	5,5%	4,6%	7,8%	6,4%	7,9%	11,1%
KDU-ČSL	5,7%	6,1%	6,0%	5,4%	4,9%	8,0%	3,5%	3,2%	7,9%	9,1%	5,6%	4,8%	4,2%	4,6%
SPOZ	3,3%	2,7%	4,5%	1,9%	3,3%	1,3%	1,9%	2,7%	4,3%	2,5%	2,8%	3,8%	3,3%	4,0%
SZ	3,4%	3,7%	2,7%	2,9%	4,2%	3,3%	3,1%	4,4%	3,1%	2,5%	3,7%	4,4%	4,0%	2,2%
Ostatní	3,1%	2,9%	2,2%	3,3%	3,9%	4,0%	2,9%	4,8%	2,4%	1,7%	3,8%	2,3%	3,0%	4,5%

		Pracovní zařazení						
		Podnikatelé + OSVČ	Zaměstnanec	Nezaměstnaný	Důchodce	V domácn./na mateřské	Studenti, žáci, ostatní	Ostatní
CSSD	31,0%	10,0%	32,2%	41,9%	36,5%	18,8%	10,2%	33,3%
ODS	20,2%	49,6%	23,1%	5,8%	9,0%	31,3%	25,9%	11,1%
TOP09	12,9%	21,5%	14,5%	10,5%	5,8%	18,8%	34,7%	3,7%
KSCM	13,1%	2,7%	6,0%	19,8%	27,8%	3,9%	,7%	14,8%
VV	7,5%	7,3%	9,0%	9,3%	3,8%	14,8%	12,2%	11,1%
KDU-CSL	5,7%	3,8%	4,3%	3,5%	9,2%	1,6%	2,0%	3,7%
Ostatní	9,7%	5,0%	11,0%	9,3%	7,8%	10,9%	14,3%	22,2%

Oproti tomu elektorát hlavních levicových stran (ČSSD + KSČM) je situován do nižších příjmových skupin – přičemž tento vztah je silnější u KSČM. Obě strany si drží vysokou podporu zejména u respondentů s příjmem s čistým měsíčním příjmem na člena pod 15 000 Kč (33-35 % u ČSSD, 16 % u KSČM). S rostoucím příjmem pak zejména podpora KSČM upadá, zatímco ČSSD má z hlediska příjmu elektorát více heterogenní. Z hlediska pracovního zařazení jsou preference levice silné zejména mezi ekonomicky neaktivní částí populace, přičemž se zde projevuje korelovanost preferencí KSČM s věkem – ČSSD mají silnější potenciál mezi nezaměstnanými, KSČM mezi důchodci.

Z hlediska life-stylových proměnných je elektorát situován dominantně mezi lidmi silně užívající internet (denní uživatelé) a lidmi preferující volno-časové aktivity mimo domov. Struktura elektorátu levice je podle očekávání⁶⁴ opět opačná – silněji jsou zastoupeni neuživatelé internetu a lidé preferující domácí volnočasové aktivity.

Struktura elektorátu KDU-ČSL je z hlediska příjmů, pracovního zařazení, preferencí volnočasových aktivit a intenzity užívání internetu podobná struktuře elektorátu levice (výrazná podpora mezi nízkopříjmovými respondenty, lidmi s preferencí domácích volnočasových aktivit, neuživateli internetu a důchodci – nikoli však nezaměstnanými). Struktura voličů Věcí veřejných je oproti ostatním parlamentním stranám velmi diverzní, což reflektuje absenci historie u této strany (malá stranická identifikace a volba podle příslušnosti k sociální třídě a skupině).

2.3.3 Predikce volební preference pomocí regresních stromů

Zjištění z předchozích dvou podkapitol by se daly shrnout do následujícího závěru: Volební preference obyvatel ČR 2010 jsou v řadě případů více než základními sociodemografickými faktory (podle nichž probíhá kontrola reprezentativity, vážení či kvótování) ovlivněny faktory specifickými jsou – příjem respondenta, příjem jeho domácnosti, pracovní zařazení respondenta či i velikost rodiny, v níž bydlí. Na volební preference navíc významně působí faktory lifestyle – například preference ve volnočasových aktivitách či intenzita využívání internetu. Působení všech faktorů je komplexní – některé kombinace sociodemografií posilují vliv faktorů jednotlivě (viz TAB 2.3.3), lifestyle faktory ovlivňují volební preference uvnitř některých věkových skupin, atd.

Pro vytvoření lepšího obrázku o skutečném vzájemném působení faktorů je tedy nutno aplikovat pokročilejší statistické metody – například multinomickou regresi (závislá proměnná

⁶⁴ To, že je struktura elektorátů opačná nelze považovat za překvapivé zjištění. Čtyři hlavní pravicové / levicové strany (ODS, TOP09, ČSSD, KSČM) mají ve volebním modelu dohromady 77 % hlasů. Vychýlení na jedné straně spektra se tak komplementárně pochopitelně projevuje opačným jevem na straně druhé.

je nominální) či metodu klasifikačních a regresních stromů (MCART). Zvolili jsme druhé řešení a pro vytvoření rozhodovacích stromů jsme použily analytický software společnosti MEDIAN – Data Analyzer (MEDIAN 2011). Ten pracuje na bázi obdobné klasifikačním a regresním stromům (C&RT) v software SPSS Answer Tree, ale umožňuje větší uživatelské nastavení a pokročilejší systém tvorby stromů.⁶⁵

“Výhodou klasifikačních a regresních stromů je, že pružně postihují vztahy mezi různými typy veličin, nelineární závislosti, interakce proměnných (...) Ve srovnání s klasickými parametrickými metodami dosahují často srovnatelné přesnosti, ale poskytují přitom daleko přehlednější a názornější modely,” popisují přínosy metody Klaschka a Kotrč (2004, s. 179).

Nevýhodou MCART je podle Klaschky a Kotrče nestabilita stromů (při malé změně vstupních parametrů⁶⁶ lze vytvořit zcela odlišný regresní strom pro téže data se srovnatelnou chybou odhadu). Lze doplnit, že další nevýhodou regresních stromů je v našem případě fakt, že mají omezené možnosti odhadu málo pravděpodobných kategorií.⁶⁷ Proto byla MCART aplikována na tříhodnotovou proměnnou volebních preferencí s hodnotami – parlamentní pravice (ODS, TOP09, VV), parlamentní levice (ČSSD, KSČM) a neparlamentní strany. Analýza byla aplikována na respondenty vstupující do volebního modelu (18 let a více, avizují účast u voleb a vědí, koho by volili). Jako prediktory vstupují všechny jednorozměrné kategorie analyzované výše, a to v maximálně podrobné formě. (Viz Příloha 6)

Obrázek v Příloze 6 ukazuje podobu stromu, který ze zvolených 12 proměnných nejlepší predikce. První štěpení probíhá podle proměnné *kraje*, druhé podle proměnných *pracovního zařazení* a *vzdělání*, třetí štěpení podle *proměnných pracovního zařazení* a *čistého měsíčního příjmu respondenta*, atd. Na nižších úrovních analýza pro zlepšení predikce zařadila v některých větvích i údaj o *počtu členů domácnosti* a *preferencích respondenta v trávení volno-časových aktivit*.

⁶⁵ Uživatelské nastavení umožňuje např. definovat maximální hloubku stromu (počet úrovní), minimálního počtu respondentů (n) v uzlu a jeho synech (štěpeních). MCART v Data analyzer navíc pracuje s pokročilejší metodou tzv. prořezávání stromů (Klaschka & Kotrč 2004, s. 178), kdy při prořezávání aplikuje štěpení, které sice na úrovni synů produkují horší kriteriální statistiky (splitting criterion), ale v dalších úrovních mohou umožňují výhodnější štěpení a tedy vést k celkově vyšší kvalitě predikce (menší chybě).

⁶⁶ V naší analýze jsme rovněž zkoušeli zařazování různých typů proměnných (např. věk kardinálně / ordinálně, atd.) – vybrán byl model s nejmenší chybou predikce. Maximální hloubka stromu byla pro všechny stromy zafixována na 8 úrovní a maximální počet tzv. synů ve štěpení na 6. Minimální počet respondentů v uzlu štěpení byl stanoven na 80 a v synu štěpení na 40.

⁶⁷ Např. neexistuje jakákoli kombinace sociodemografických a dalších celkem 12 proměnných zmíněných v kapitole 2.3, při níž by byla největší pravděpodobnost, že respondent bude volit např. KDU-ČSL. Na tak podrobné úrovni se dají regresní stromy úspěšně aplikovat, jen pokud do prediktorů zařadíme proměnnou strany volené v minulých volbách. V tom případě ovšem z predikce vypadne drtivá většina dalších faktorů a ztratí se smysl zmapování vlivu sociodemografických a lifestyleových faktorů na volební preference.

Potvrzuje se tedy naše zjištění z kontingenční analýzy, že tyto proměnné působí na volební preference (či řečeno z druhé strany – přinášejí více informací pro jejich predikci) více než některé základní sociodemografické faktory.⁶⁸ Naopak model nevyužívá proměnnou *intenzity využívání internetu*.⁶⁹ Celková úspěšnost predikce je pro tříhodnotovou překódovanou závislou proměnnou 61,76 % - tedy cca o necelých 18 procentních bodů úspěšnější, než by bylo optimalizované náhodné tipování bez jakýchkoli informací o respondentovi (tj. tipování nejfrekventovanější kategorie – parlamentní levice, která má v modelu 44,1 %).

Při zařazení pouze pěti základních demografických faktorů (pohlaví, věk, vzdělání, kraj, VMB) by byla úspěšnost predikce podstatně nižší (47 %).

2.3.4 Důsledky zjištění pro preferenční výzkumy

Jak bylo uvedeno v kapitole 2.3.2.1 hlavní agentury používající kvótní výzkum (STEM, CVVM, Factum) využívají podle otevřených zdrojů pro určování kvótních znaků výhradně základní sociodemografické veličiny a společnost MEDIAN před volbami v roce 2010 tyto veličiny využívala (jednoduše a sdruženě) pro post-stratifikační vážení.

Z informací agentur Factum a STEM (Chábová 2011, s.42) (STEM 2010a) (Factum 2010b) vyplývá, že tyto společnosti využívají jistý systém váženého zařazování respondentů do volebního modelu, podle toho jak respondent odhaduje svoji účast u voleb. Factum i STEM mimo to zohledňují výsledky minulých voleb (Chábová 2011, s.42) (STEM 2010a) (Factum 2010b), ačkoli z otevřených zdrojů není zřejmé v jakém kroku a jakým způsobem tyto operace aplikují.⁷⁰

Z otevřených zdrojů (weby a tiskové zprávy agentur) ani pracích zabývajících se aktuální praxí volebních výzkumů v ČR (Chábová 2011) však nijak nevyplývá, že by agentury používající kvótní výzkum (STEM, CVVM, Factum), aplikovaly jakékoli další vážení dat podle specifických sociodemografických a základních life-stylových proměnných.

⁶⁸ Například proměnnou velikost místa bydliště regresní strom nevyužívá vůbec a pohlaví jen v jedné nižší větvi stromu.

⁶⁹ Ve specifických kombinovaných kategoriích podle kraje, věku, pracovního zařazení, atd. už tato proměnná nemusí přinášet dostatečné informace pro signifikantní zlepšení modelu. To však neznamená, že preference neovlivňuje – do regresního stromu by se mohla dostat při vyšším počtu respondentů (větší N v nižších potomcích stromu) či při zařazení dalších proměnných, s kterými může interagovat, atd. Do testů alternativních vážení ji proto zařadíme.

⁷⁰ Např. zda toto zohlednění pracuje vážením na úrovni respondentů, či zásahem do výsledného modelu. U STEM (díky publikaci stranických preferencí a odhadu zisku mandátů) navíc nelze rozeznat zda agentura tento zásah dělá při tvorbě volebního modelu či až při odhadu zisku mandátů. Je zde tedy nemožné rozeznat odchylky způsobené tvorbou modelu a zjednodušením přepočtu podle d'Hondtova systému.

Neexistuje přitom relevantní důvod, proč volební výzkum označovat za „obecně“ reprezentativní (viz kapitola 1.3), když:

- a) populačním četnostem odpovídá (díky kvótám či vážení) ve veličinách, které s volební preferencí souvisí velmi málo (např. pohlaví) či omezeně (velikost obce)
- b) není jasné, zda populačním četnostem odpovídá ve veličinách, které s volebními preferencemi souvisí silněji (pracovní status, příjem respondenta a jeho domácnosti, intenzita používání internetu) či srovnatelně silně (preferenze volnočasových aktivit, počet členů domácnosti)
- c) zároveň z analýz vyplývá, že minimálně některé tyto proměnné diverzifikují preference respondentů uvnitř kategorií základních sociodemografických proměnných a jsou prediktivními faktory s nenahraditelnou informační hodnotou.

Toto zjištění přitom může být pro kvótní výzkumy ještě závažnější než pro výzkumy s pravděpodobnostním designem výběru, které se dají označit za reprezentativní v druhém smyslu tohoto slova (viz kapitola 1.3).

Pokud totiž některé z nekvótních znaků souvisí s pravděpodobností zastižení respondenta v domácnosti (např. preference volnočasových aktivit, pracovní aktivita, atd.), redukuje toto riziko pravděpodobností výběr postupy náhodného výběru respondenta na adrese Kishovou tabulkou a opakovanou návštěvou pro jeho zastižení (např. min. 4 návštěvy). Při kvótních výzkumech je díky absenci těchto institucí (náhodný výběr respondenta + opakovaná návštěva) úroveň diskvalifikace respondentů, kteří mají díky nekvótním znakům menší šanci na přítomnost v domácnosti, ceteris paribus nutně vyšší.

Na toto úskalí volebních výzkumů přitom pamatují i metodologické doporučení asociace ESOMAR:

„Bylo by dobrou praxí v předvolebních výzkumech, zaměřit se ve sběru na data, které mohou s volebními volebními korelovat. Mělo by být známo, jak silně se tyto položky v alternativních zdrojích vyskytují. (...) Pokud demograficky reprezentativní výzkum (ve Velké Británii, pozn. autora) produkuje moc velký odhad členství v odborech, je pravděpodobné, že tento výzkum také nadhodnotí zisk Labouristické strany.“ (Esomar 2003, s. 18, překlad autor)

2.3.5 Srovnání preferencí nadhodnocených / podhodnocených stran podle faktorů, které nejsou předmětem sociodemografického vážení

Před analýzou nakolik lze reprezentativitu v klíčových faktorech zlepšit specifickými typy vážení a kombinovanou metodikou sběru dat (kapitola 3), se podívejme, zda specifické

sociodemografické a life-stylové faktory souvisí s vychýlením, které výzkum MEDIANu před volbami 2010 produkoval.

Výrazně (a opakovaně) nadhodnocené strany před volbami 2010 ve volebním modelu společnosti MEDIAN byly ČSSD, KDU-ČSL a SPOZ a výrazně podhodnocené byly naopak TOP09 a VV. Křížení (TAB 2.3.6) ukazuje statisticky významný vztah⁷¹ typu respondenta s preferencí nadhodnocených resp. podhodnocených stran.

Nadhodnocené strany mají elektorát dominantněji situovaný do nízkopříjmových skupin respondentů / domácností, mezi největší sektory ekonomicky neaktivního obyvatelstva (důchodci, nezaměstnaní), mezi neuživatelé internetu a mezi lidi s preferencí domácích volnočasových aktivit. Naopak strany modelem podhodnocené mají voliče vychýleny opačným směrem.

Podhodnoceny jsou tedy strany, které preferují skupiny respondentů, u nichž může předpokládat těžší zastižitelnost v domácnosti i při opakované návštěvě (ekonomicky aktivní + lidé preferující venkovní volnočasové aktivity) a lidé, pro které má finanční / hmotná motivace k účasti ve osobním sběru relativně menší hodnotu (nízkopříjmoví), což může ovlivňovat míru odmítnutí.

V kapitole 3 prozkoumáme, jaká je reprezentativita vzorků tvořených osobním a telefonickým sběrem v těch kritériích, kde existuje relevantní informace o struktuře v populaci (pracovní zařazení, příjmy, užívání internetu, atd.).

Výzkumnou otázkou je, zda by se reprezentativita a predikční schopnost volebního modelu zlepšila spojením s telefonním sběrem (CATI) dat, který může mít odlišný charakter „non-response bias“. A zda by se predikční vlastnost volebního modelu z osobního (CAPI) sběru zlepšila, pokud bychom do post-stratifikačního zařadili nejen standardní sociodemografie, ale i specifické faktory (příjmy, pracovní zařazení, užívání internetu, atd.)

71 Hodnota Pearsonova Chi-sq 139 při 8 stupních volnosti se splněním podmínek testu. Tj. H0 o nesouvislosti zamítnutelná na 99% hladině spolehlivosti.

TAB 2.3.6 Vztah vychýlení volebního modelu a specifických sociodemografických a lifestylových faktorů

		Čistý měsíční příjem domácnosti (R)					Čistý měsíční příjem respondenta (R)					Příjem domácnosti na člena				
		do 20 tisíc	20.001 - 30.000	30.001 - 40.000	nad 40.000	odmítl(a) odpovědět	do 10 tisíc	10 001 - 15 000	15 001 - 25 000	nad 25 000	odmítl(a) odpovědět	do 10000	10001 - 15000	15001 - 20000	20001 a více	odmítl(a) odpovědět
nadhodnocené strany	40,0%	49,4%	45,6%	38,2%	27,7%	31,3%	44,4%	46,2%	36,0%	25,6%	28,8%	45,9%	43,2%	36,0%	20,7%	31,3%
strany bez vychýlení	39,7%	38,5%	38,0%	36,8%	42,8%	46,2%	36,5%	40,9%	36,6%	47,5%	46,8%	36,3%	39,1%	40,9%	44,8%	46,2%
podhodnocené strany	20,4%	12,1%	16,4%	25,0%	29,5%	22,5%	19,2%	12,9%	27,4%	26,9%	24,4%	17,8%	17,7%	23,0%	34,5%	22,5%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

		Pracovní aktivita							Preference volnočasových aktivit				
		Podnikatel + OSVČ	Zaměstnanec	Nezaměstnaný	Důchodce	V domácnosti / na mateřské	Studenti, žáci, ostatní	Ostatní	velmi domácí	spíše domácí	neutrální	spíše venkovní	velmi venkovní
nadhodnocené strany	40,0%	14,2%	39,7%	45,3%	49,6%	26,6%	13,6%	44,4%	52,9%	42,4%	40,0%	35,1%	28,7%
strany bez vychýlení	39,7%	56,9%	36,8%	34,9%	40,7%	39,8%	39,5%	40,7%	33,6%	40,0%	38,5%	42,6%	43,8%
podhodnocené strany	20,4%	28,8%	23,4%	19,8%	9,7%	33,6%	46,9%	14,8%	13,5%	17,5%	21,5%	22,3%	27,5%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

		Síla užívání internetu			Počet členů domácnosti				
		Silní uživatelé (denně, téměř denně)	Slabí uživatelé (méně než jednou denně)	Neuživatelé	1 člen	2 členi	3 členi	4 členi	5 a více členů
nadhodnocené strany	40,0%	25,4%	36,7%	49,3%	41,7%	46,0%	36,0%	35,4%	35,3%
strany bez vychýlení	39,7%	42,5%	40,1%	37,9%	39,7%	38,5%	40,9%	40,2%	39,3%
podhodnocené strany	20,4%	32,1%	23,2%	12,8%	18,6%	15,5%	23,1%	24,4%	25,3%
TOTAL	100%	100%	100%	100%	100%	100%	100%	100%	100%

Zdroj: Data volebního sběru MEDIAN červenec – prosinec 2010. Nevážená. Respondenti avizující účast u voleb, pokud by se konaly nyní, preferující jednu stranu (n=3974)

Pozn.: Červeně vybarvená pole o více než 25 % (nikoli p.b.) překračují zisk skupiny stran v celkovém volebním modelu. Modře vybarvená o více než 25 % nedosahují hodnoty v celkovém modelu. Červený / modrý font značí statisticky významnou odlišnost (p=0,05) od hodnoty v celkovém modelu.

3. Optimalizace volebního modelu na příkladu společnosti MEDIAN

3.1 Optimalizace volební predikce kombinací metod sběru dat (CAPI + CATI)

3.1.1 Teoretický úvod k řešení

Kombinace více metodik kvantitativního sběru dat (mixed-mode data collection) je v sociologické literatuře doporučována zejména proto, že:

- a) Při předpokladu fixní velikosti vzorku se může jednat o cenově efektivnější variantu sběru dat, než použití jedné techniky sběru. Přičemž důležitější je obrácená interpretace – lze dosáhnout podstatného rozšíření vzorku za omezeného růstu nákladů. (Martin 2011, s.8)
- b) Může vést redukci systematického vychýlení výběru způsobeného nezastižením respondentů (non-contact rate) a odmítnutím účastnit se dotazování (refusal rate), tedy celkově k redukci systematického vlivu tzv. non-response bias (více o těchto kategoriích viz Krejčí 2007). Nejde přitom jen o prosté zvýšení návratnosti, která nemůže být finálním a jediným cílem výběru metodologie:

„Pokud může být reprezentativita dosažena při nižší návratnosti a pokud vzorek dané velikosti může být sesbírán s nižšími náklady metodou, která má nižší návratnost (...), pak by neexistoval jediný důvod trvat na samoúčelné maximalizaci míry návratnosti,“ píše například Martin a navrhuje, aby se z tohoto důvodu výzkum European Social Survey více otevřel kombinovaným metodám sběru dat. (Martin 2011, s.8)

Další autoři se shodují, že kombinace více metod sběru dat může být při dostatečných oporách výběru pro obě metody a dodržení určitých podmínek (deRada 2010, s.49) vést k eliminaci systematických vychýlení, které jsou způsobeny rozdílnou strukturou non-response rate u různých metod sběrů dat (Bowers and Ensley 2003).

Konkrétně v případě kombinace sběru dat osobním a telefonickým dotazování došla řada výzkumníků (Wessell 2000) (deRada 2010) (Bowers and Ensley 2003) například ke zjištění, že *„[...] osobní sběr vede k výběru vyššího množství lidí s nízkým stupněm vzdělání (...), zatímco telefonní sběr dat zasáhne vyšší počet lidí se sekundárním a terciálním vzděláním;“* (deRada 2010, s.54).

deRada v případě španělských voleb dále zjistil, že v telefonním výzkumu respondenti častěji avizují pravděpodobnou účast u voleb než ve výzkumu osobním dotazováním (deRada 2010, s. 56). Ve srovnání s výsledky voleb telefonní výzkum podhodnocoval hlavní pravicovou stranu PP a nadhodnocoval nacionalistické (BNG) a levicové strany (PSdeG/PSOE), přičemž face-to-face sběr dat se choval naprosto opačně.⁷²

deRada při analýze dalších faktorů dále shrnuje, že: „*Zkráceně řečeno, telefonický výzkum dokáže postihnout větší počet aktivních lidí (tím myslíme pracovně aktivních), (...)*“ (deRada 2010, s.54) Volební model z osobního sběru dat ve Španělsku predikoval výsledky voleb přesněji než volební model z telefonního sběru. Platilo ale, že „*(...) oba výzkumy produkují lepší odhad dohromady, než každý z nich jednotlivě.*“ (deRada 2010, s.60)

V případě našeho srovnání je specifikum v tom, že CATI výzkum obsahoval složku podobnou kvótnímu výběru. Výběr začíná stratifikovaným náhodným výběrem telefonních čísel z databáze pevných linek a prostým náhodným výběrem (generování) mobilních kontaktů (v přesně daném poměru obou opor). Sekundárně je ale do výběru zapojeno i eliminování respondentů, kteří by výrazně narušili reprezentativitu podle znaků určených základními sociodemografiemi (pohlaví, věk, vzdělání, kraj, VMB).

Využití telefonního výzkumu musí být podepřeno existencí kvalitní opory výběru, která by měla zasahovat min. 85 % cílové populace (Esomar/Wapor 2003, s. 16). Podle údajů ČSÚ v roce 2008 užívalo mobilní telefon 87,9 % obyvatel ČR ve věku 16+.⁷³ (ČSU 2009a) Od roku 2008 penetrace mobilními telefony dále pokračovala. Přesto, že obecný požadavek pokryvu 85% populace je překročen, dá se populace zastižitelná mobilem považovat stále za sociodemograficky vychýlenou.⁷⁴ Pevnou telefonní linku v roce 2008 mělo 31,1 % českých domácností. (ČSU 2009b) Distribuce jejího vlastnění je přitom z hlediska sociodemografie takřka opačná než u mobilních telefonů.⁷⁵ Tato specifičnost obou výběrových opor si

⁷² Jde de facto o opačný jev, než v ČR, kde telefonický výzkum vede k podhodnocení levice a nadhodnocení pravice (viz níže). Je však nutné si uvědomit, že Španělsko má velmi rozdílnou historii, než Česká republika. Okolo 40 let strávilo v pravicové Frankově diktatuře, zatímco Československo v „levicové“ komunistické totalitě. Z toho vyplývá velmi rozdílná politická orientace intelektuálních a kulturních elit, mládeže, apod. v obou těchto zemích.

⁷³ A bylo tedy zařaditelných do výběru náhodným generováním mobilních kontaktů respektujícím strukturu operátorů (předvoleb) v ČR pro dosažení stejné pravděpodobnosti zařazení jednotek.

⁷⁴ Mezi nad 64 let používalo v roce 2008 mobilní telefon jen 52,1 % lidí a i díky korelaci s věkem je omezený počet uživatelů mezi lidmi se základním vzděláním (61,9 % uživatelů) a mezi starobními důchodci (59,6 % uživatelů). (ČSU 2009a)

⁷⁵ Pevnou linku vlastní jen 6,6 % domácností s hlavou domácnosti do 29 let, ale naopak 36,5 % a 52,4 % domácností s hlavou domácnosti mezi 60-69 lety respektive 70 a více lety. Díky korelaci s věkem vlastní pevnou linku více domácností důchodců (41,3 %). Na druhou stranu je málo zastoupena v rodinách ekonomicky neaktivních hlav domácnosti (jen 15,9 % domácností s nezaměstnanou hlavou rodiny). (ČSU 2009b)

vynucuje jejich kombinované využití, ale také je argumentem pro sledování (a eliminování respondentů podle) sociodemografických znaků při sběru dat.

Jak však vyplývá ze zjištění v kapitole 2.3 i zahraniční studie uvedené výše, sledování základních sociodemografických znaků však nezaručuje reprezentativitu vhodnou pro správný odhad volebních výsledků. Telefonický sběr (i s kombinovanou a dostatečnou oporou výběru) může díky jiné struktuře „non-response bias“ produkovat oproti osobnímu sběru rozlišnou reprezentaci skupin obyvatelstva podle specifických ukazatelů (pracovní status, příjem, užívání médií, volnočasová aktivita, atd.), které jsou pro volební preference zásadní a základní sociodemografické znaky je plně nepostihují.

3.1.2 Vliv kombinace osobního a telefonního sběru na reprezentativitu

Možnosti analýzy reprezentativity jsou limitovány přítomností a charakterem relevantní datové opory pro definování četností vybraných charakteristik v populaci. Při použití dat z Výběrového šetření pracovních sil, výzkumu SILC, exit-pollu SC&C a Eurostatu pro definování teoretických četností předpokládáme, že tato výběrová šetření mají minimální systematické vychýlení a díky velkému počtu respondentů malé konfidenční intervaly okolo zkoumaných četností a mohou tak sloužit jako datová opora pro definici teoretických četností. V analýze se zaměříme postupně na standardní sociodemografické faktory, specifické sociodemografické faktory a life-stylové faktory analyzované v kapitole 2.3.

3.1.2.1 Reprezentativita podle základních sociodemografických ukazatelů – plné vzorky

Rozdílné empirické studie se neshodují, zda za předpokladu aplikace pravděpodobnostního výběru dochází k větší reprezentativitě telefonního sběru dat (deRada 2010), či osobní sběr (Bowers and Ensley 2003). Lze předpokládat, že na to má vliv i přesná podoba pravděpodobnostního výběru⁷⁶ a míra, s jakou se výběrové modely od pravděpodobnostního výběru odklánějí.⁷⁷

V této kapitole srovnáváme vlastnosti nevážených vzorků z CAPI sběru a CATI sběru dat společnosti MEDIAN na měsících duben 2010, květen 2010⁷⁸ a červenec 2010 a jejich spojení. Porovnávat reprezentativitu celých vzorků pro odpovídající populaci (12-79 let)

⁷⁶ Tedy: zda je stratifikovaný, v kolika výběrových stupních probíhá, kvalita opor, systém eliminování nadreprezentace respondentů z máločlených domácností adresním výběrem, zda je navýšení u adresního výběru aplikováno paušálně či regionálně specificky, atd.

⁷⁷ Např. možnost substitue, sekundární sledování kvótních znaků, atd.

⁷⁸ Pro účely srovnání sociodemografických údajů (tj. ne poslední volené strany) byl tento měsíc zařazen do srovnání celý – tj. i včetně volebního a povolebního období 28.5. - 1.6.2010.

z hlediska základních sociodemografických znaků lze jen s vědomím, že u CATI sběru obsahoval prvek eliminování respondentů podle těchto základních znaků a reprezentativita tedy bude nutně vyšší.⁷⁹

Ze srovnání vychází, že mezi základními sociodemografickými veličinami CAPI výzkum *vícestupňový stratifikovaný náhodný* adresní výzkum produkoval srovnatelnou reprezentativitu v kritériu *velikosti místa bydliště*.⁸⁰ V ostatních základních sociodemografických kritériích (*pohlaví, věk, kraj, vzdělání*) produkuje vyšší reprezentativitu CATI sběr. (Viz TAB 3.1.1). Platí přitom, že v jednom případě (*vmb*) produkuje lepší reprezentativitu spojení sběrů než obě metody sběru samostatně. V dalších 4 případech (*pohlaví, kraj, vzdělání, věk*) spojení s CATI vylepšuje hodnoty z CAPI sběru a z toho ve 2 případech (*kraj, vzdělání*) se hodnoty takřka blíží lepšími z obou sběrů. (Viz TAB 3.1.1)

TAB 3.1.1 Reprezentativita výběru pro populaci 12-79 let (suma absolutních odchylek teoretických a evidovaných procentuálních četností - v procentních bodech)

	CAPI 04/10	CAPI 05/10	CAPI 07/10	CAPI 04+05+07/10	CATI 04/10	CATI 05/10	CATI 07/10	CATI 04+05+07/10	CATI + CAPI 04+05+07/10
Pohlaví	10,6	14,3	9,9	11,6	1,0	1,9	1,5	0,8	5,6
Kraj	20,9	20,2	20,1	17,3	10,1	7,6	8,8	6,2	7,5
Vzdělání	14,3	10,5	13,2	11,5	8,8	17,6	19,3	8,8	9,4
Věk	19,1	19,4	20,6	19,7	13,3	8,0	5,5	6,1	10,3
VMB	8,8	10,3	8,4	7,2	9,6	10,4	4,6	8,4	5,4

Pozn.: Srovnávána je struktura vzorku 12-79 let se strukturou populace 12-79 let (zdroj t.č.: ČSU 2010a, ČSU 2010c, ČSU 2010e)

Spojení telefonního a osobního sběru dat se ukazuje jako žádoucí zejména pro zvýšení reprezentativity z hlediska: A) *věku* (osobní sběr bez vážení nadhodnocuje nejnižší a nejvyšší věkové skupiny 12-17 let, 55-79 let)⁸¹, B) *velikosti místa bydliště* (osobní sběr má tendenci mírně nadhodnocovat středně velká města 20 - 99 999 tisíc), C) *kraje* (osobní

⁷⁹ Odchylky v reprezentaci u CATI sběru jsou dány zejména použitím teoretických četností z nevýběrového šetření sčítání lidu 2001 (např. vzdělání) při eliminaci respondentů a relativní volnosti těchto „kvót“, zatímco pro naše srovnání používáme údaje z projekcí a výběrových šetření ČSU aktuálních v době sběru (ČSU 2010a, ČSU 2010c, ČSU 2010e).

⁸⁰ Jedná se o jednu ze strat ve stratifikovaném adresním výběru. Z výsledků lze tedy usuzovat, že velikost místa bydliště výrazně neovlivňuje ochotu respondentů na vybraných adresách účastnit se výzkumu či jejich zastizitelnost.

⁸¹ Nadhodnocení nejnižší věkové skupiny není pro volební model problém, protože otázka na preference směřuje jen na občany 18+. Nadhodnocení vysoké věkové skupiny se dá zdůvodnit vyšší zastihnutelností těchto respondentů v domácnosti, vyšší ochotou odpovídat (viz kapitola 2.4), i případnou nedostatečností mechanismů vynechávání málopočetných starších domácností ve vzorku vybraným náhodným adresním výběrem (protože vyšší věkové skupiny mají větší pravděpodobnost života v máločlenných domácnostech a Kishova tabulka v takových domácnostech vybírá z omezenějšího okruhu respondentů žijících na dané adrese, zvyšuje se pravděpodobnost zařazení takových respondentů do vzorku).

dotazování bylo úspěšnější na Moravě). Tyto tři faktory se spojením s CATI vzorkem zmírňují či neutralizují. Podrobněji viz Příloha 3.

3.1.2.2 Reprezentativita podle základních sociodemografických ukazatelů – vzorky voličů

Z hlediska správnosti predikce volebního výsledku je kromě reprezentativity celého výběrového vzorku pro českou populaci 12-79 let důležité zejména to, jaká je struktura části vzorku, u níž průzkum odhaduje účast u voleb a kalkuluje z ní volební model. V ideálním případě by měla tato struktura v kategoriích, které statisticky souvisí s volebními preferencemi respondentů, co nejvíce odpovídat struktuře lidí, kteří šli v 28. a 29. května k volbám. Pro účely této analýzy jsme četnosti tří základních sociodemografických kategorií (pohlaví, vzdělání, věk) v sub-samplech, z nichž MEDIAN počítá volební model (avizují účast u voleb – spíše či určitě), s četnostmi těchto kategorií v exit-pollu SC&C (2010b).⁸² Ve srovnání struktury voličů z hlediska krajů jsme vycházeli přímo z tohoto údaje ČSU (2010b).⁸³

Ze srovnání se ukazuje, že CAPI sběr dat dochází k shodné či nepatrně lepší reprezentaci struktury voličů v jedné kategorii (vzdělání), CATI sběr ve dvou kategoriích (kraj, věk) a reprezentace z hlediska pohlaví je přibližně srovnatelná. Spojením obou metod sběru docházíme ve dvou případech (pohlaví, kraj) vylepšuje odhad struktury voličů oproti oběma metodám jednotlivě, v jednom případě oproti čistému osobnímu sběru (věk) a v druhém případě jen oproti čistému telefonnímu sběru (vzdělání) – ačkoli zde jsou rozdíly minimální. Viz TAB 3.1.2

TAB 3.1.2 Reprezentativita výběru pro voliče z května 2010 (suma abs. odchylek mezi teoretickými a evidovanými procent. četnostmi jednotlivých kategorií proměnné)

	CAPI 04/10	CAPI ⁸⁴ 04-05/10	CAPI 05*/10	CAPI 04+05*	CATI 04/10	CATI 04-05/10	CATI 05*/10	CATI 04+05*	CAPI+CATI 04+05*
Pohlaví	1,6	10,9	12,9	6,7	9,5	6,8	1,1	5,7	0,5
Kraj	30,3	22,5	23,4	22,6	9,9	11,4	14,6	11,5	10,9
Vzdělání	30,4	30,9	27,5	28,7	19,9	33,2	43,4	30,6	29,8
Věk	23,6	28,9	24,1	23,8	16,8	13,3	12,0	11,3	16,4

Srovnávána je struktura voličů z května 2010 (zdroj t.č.: SC&C 2010, ČSU 2010b) a voličských částí vzorků (18 let +, avizovali účast u voleb). * Z květnových vzorků jsou použita jen data do začátku voleb (do 28.5. 14:00)

⁸² Zde tedy předpokládáme, že sociodemografie vzorku úspěšného exit-pollu SC&C (SC&C 2010b) s výrazně neodlišuje od sociodemografie skupiny obyvatel, kteří ve volbách v květnu 2010 odevzdali platné hlasy.

⁸³ Srovnání z hlediska velikosti obce není možné, protože ČSU tento údaj ve svých výstupech eviduje v kategoriích zcela odlišných od kategorií používaných společnostmi MEDIAN.

⁸⁴ Označení 04-05 / 10 znamená tzv. plovoucí data – tj. data vzniklá spojením druhé poloviny dubnového sběru a první poloviny květnového sběru.

Spojení telefonního a osobního sběru dat se ukazuje pro odhad skutečné struktury voličů jako žádoucí zejména z hlediska proměnných:

A) *kraje*, kde osobní sběr mírně nadhodnocuje počet voličů z Moravy a podhodnocuje voliče z ČR a telefonní sběr se chová opačným způsobem.

B) *pohlaví*, kde osobní sběr mírně nadhodnocuje počet žen a telefonní počet mužů, čímž opět dochází k neutralizaci vychýlení produkovaných oběma metodami.

C) *věku*, kde se sice obě metody vychylují podobným směrem (nahodnocení počtu voličů 60+ a podhodnocení počtu voličů mladších 29 let), ale telefonní sběr produkuje vychýlení menší, čímž zlepšuje reprezentaci při spojení vzorků z obou metod.

3.1.2.3 Reprezentativita podle speciálních sociodemografických ukazatelů

Jak ukázala analýza v kapitole 2.3, volební preference jednotlivců závisí kromě základních sociodemografických faktorů i na aktivitě respondenta (komponované z pracovní aktivity a volnočasové aktivity) a příjmu domácnosti, z které respondenta pochází. Tento vztah je přitom silnější než vztah mezi preferencemi a některými základními sociodemografickými faktory (pohlaví). V analýze reprezentativity vzorků vzniklých jednotlivými typy sběru dat se tedy musíme zaměřit i na tyto faktory, které navíc v případě telefonního sběru nejsou předmětem kvartálně udržovaných kvót.

TAB 3.1.3 Reprezentativita vzorků z hlediska speciálních sociodemografických kritérií (suma abs. odchylek mezi teoretickými a evidovanými procent. četnostmi jednotlivých kategorií proměnné)

	CAPI 04/10	CAPI 05/10	CAPI 07/10	CAPI 04+05+07/10	CATI 04/10	CATI 05/10	CATI 07/10	CATI 04+05+07/10	CATI + CAPI 04+05+07/10
Pracovní status	15,7	13,4	10,8	11,9	11,4	7,9	11,0	7,6	9,2
Pracovní aktivita	11,1	13,5	9,5	11,3	8,7	3,2	9,9	6,0	7,4
Příjem domácnosti na člena ⁸⁵	16,8	18,7	14,1	15,7	21,3	25,5	29,7	24,3	18,5

Pozn.: U kategorií pracovní status, pracovní aktivita je srovnáván vzorek 15-79 let s odpovídající populací (zdroj t.č.: ČSU 2010f), v kategorii příjem domácnosti na člena je srovnávána vzorek 12-79 (bez odmítnutí) s údaji pro celou populaci (zdroj t.č. ČSU 2010d). V intenzitě užívání internetu je srovnáván vzorek 16-74 s odpovídající populací (zdroj t.č. Eurostat 2010).

Srovnání (viz TAB 3.1.3) vzorků s údaji Českého statistického úřadu ukazuje, že telefonní sběr produkuje reprezentativnější výběr v otázce pracovního statusu respondenta a výsledně i pracovní aktivity respondenta⁸⁶ - jeho připojením k osobnímu sběru se tedy kvalita výběru z tohoto hlediska zlepšuje. CATI konkrétně ve srovnání s CAPI, méně nadreprezentovává

⁸⁵ Byla použita proměnná kalkulovaná v kapitole 2.3.

⁸⁶ Pracovní status je sedmihodnotová proměnná (zaměstnanec, podnikatel se zaměstnanci, podnikatel bez zaměstnanců / OSVČ, nezaměstnaný, důchodce, student / žák, jiné), pracovní aktivita je rekódovaný pracovní status do tříhodnotové proměnné (aktivní, neaktivní, jiné) způsobem použitým v kapitole 2.4.

kategorii důchodců, méně podrepresentovává OSVČ a podnikatele, ale naopak více podrepresentovává kategorii zaměstnanců. Spojení je tedy z hlediska distribuce odchylek žádoucím, protože zmenšuje odchylky očekávaných a evidovaných četností v těchto nejvíce zastoupených pracovních kategoriích. (viz TAB 3.1.4 a Příloha 3)

TAB 3.1.4 Reprezentativita vzorků z hlediska pracovního statusu respondenta

	Teoretické četnosti	Odchylka evidovaných a teoretických četností (v p.b.)		
		CAPI (04+05+07/10)	CATI (04+05+07/10)	CAPI+CATI (04+05+07/10)
zaměstnanec	46,5%	-0,4	-2,8	-1,8
soukromý podnikatel (spolu)vlastník firmy bez zam.	8,0%	-2,8	-0,1	-1,3
soukromý podnikatel (spolu)vlastník firmy se zam.	2,1%	-0,7	-0,1	-0,3
nezaměstnaný	4,4%	0,0	-0,3	-0,1
důchodce (nepracující)	23,7%	5,9	1,6	3,5
v domácnosti, mateřská dovolená	4,1%	-0,3	0,9	0,4
student, žák, učeň	10,4%	-1,7	-0,5	-1,0
jiné	0,7%	0,0	1,3	0,7

Zdroj teoretických četností: (ČSU 2010f) Srovnávány jsou údaje pro populaci 15-79 let a vzorky 15-79 z obou typů sběru dat a jejich spojení. Pro redukci dat ze zdroje (15-100) byl použit předpoklad, že všichni lidé nad 79 let jsou důchodci. Barevně označeny jsou kategorie s odchylkou evidovaných četností od teoretických více než +/- 2,5 procentních bodů.

Naopak z hlediska ukazatele příjmu domácnosti přepočteného na člena⁸⁷ vede k mírně lepší reprezentaci osobní sběr dat a spojení tak vylepšuje výsledky telefonního sběru. Jedinou slabinou osobního sběru oproti telefonnímu sběru v tomto ohledu je, že vede k vyššímu nadrepresentování kategorie respondentů s čistým měsíčním příjmem domácnosti na člena 8 001 – 10 000 Kč, což je způsobeno vyšší úspěšností sběru v případě důchodců, a že vůbec nezachycuje úzkou skupinu populace s čistým příjmem domácnosti na člena nad 50 tisíc Kč měsíčně. Tento typ výchyly je spojením redukován. (viz Příloha 3)

3.1.2.4 Reprezentativita podle life-stylových faktorů a minulého volebního chování

Volební chování souvisí kromě sociodemografických charakteristik s řadou dalších faktorů jako životní styl respondenta a způsob trávení volného času (viz kapitola 2.3), míra vystavení médiím (media exposure) a konzumace jednotlivých médií (Bergar & Gerber & Karlan 2007), ideologické afinity respondenta i psychologické vlastnosti jako extroverze/introverze,

⁸⁷ Data MEDIANu neobsahují příjem domácnosti přepočtený na člena, ale poměrně ordinální proměnnou příjmu domácnosti s poměrně detailními kategoriemi (15 kategorií příjmu). Pro účely této analýzy jsme tuto ordinální proměnnou nahradili průměrem dané kategorie (tj. například příjmová kategorie 8 001 – 10 000 Kč byla nahrazena hodnotou 9 000 Kč) a tuto proměnnou jsme vydělili počtem členů domácnosti respondent. Takto vypočtená kvazi-kardinální proměnná byla nakonec rekódována do kategorií, které v evidenci příjmů domácnosti na člena používá ČSU (2010d). Popis používaných proměnných.

otevřenost, svědomitost, neurotičnost, atd. (Schoen & Schumann 2007). Ideální volební výzkum by měl pracovat se vzorky reprezentativními podle všech takových kritérií.

Zaměříme-li se na tyto specifické faktory, ukazuje se, že spojení CAPI a CATI sběru má zásadně pozitivní vliv na reprezentativitu v případě ukazatele intenzity využívání internetu. Osobní sběr totiž vede k nadreprezentování neuživatelů a telefonní k výraznému nadreprezentování silných uživatelů. Spojením sběrů se reprezentativita výrazně zvyšuje. (Viz TAB 3.1.5)

TAB 3.1.5 Reprezentativita vzorků z hlediska intenzity využívání internetu

	Teoretické četnosti	Odchylka evidovaných a teoretických četností (v p.b.)		
		CAPI (04+05+07/10)	CATI (04+05+07/10)	CAPI+CATI (04+05+07/10)
Silní uživatelé (denně) ⁸⁸	38 %	-9,2	14,5	4
Střední uživatelé (méně než denně)	31 %	-3,2	-10,5	-7,3
Neuživatelé	31 %	-12,4	-4	3,3
SUMA ABS. ODCHYLEK	0	24,9	29,1	14,5

Zdroj t.č.: (Eurostat 2010). Pozn.: Srovnávány jsou četnosti pro populaci/vzorek ve věku 16-74, kterou popisují data Eurostatu.

U většiny ostatních specifických atributů je ovšem problém, že relevantní datová opora popisující rozložení těchto charakteristik v populaci buď zcela chybí či je metodologicky přinejlepším stejně důvěryhodná jako výzkumy volebních preferencí. Proto také nelze zkoumat, nakolik analyzované typy sběru dat přispívají k udržení reprezentativity výběru z hlediska těchto speciálních kritérií.

Víme však, že aktuální volební preference souvisí s volebním chováním v minulosti.⁸⁹ Za předpokladu, že některé ze speciálních life-stylových a mediálních faktorů na volební chování trvale působí, se analýza jejich působení dá částečně obejít právě analýzou, jak dobře výzkum reprezentuje strukturu populace z hlediska volebního chování v minulých volbách.

88 Data Eurostatu a Medianu se mírně liší v definici silných uživatelů (Eurostat – „denně“, MEDIAN – „denně, či téměř denně“). To může způsobit mírnou diskrepanci mezi distribucí silných a středních uživatelů. Distribuce uživatelů tím ovšem zůstává nenarušena.

89. Při crosstabulaci proměnných o aktuálních voličských preferencích a strany volené v minulých volbách dosahuje kontingenční koeficient hodnoty 0,836 při maximální časové distanci od voleb (dubnu 2010, data z osobního sběru) a 0,896 při minimální distanci od voleb (červen 2010, data z osobního sběru) – v obou případech se dá hypotéza o neexistenci vztahu vyloučit na 99 % hladině spolehlivosti.

Výchyly mezi výsledkem dotazování na stranu volenou v minulých volbách do Sněmovny a skutečným výsledkem těchto minulých voleb, nelze pochopitelně zdůvodňovat primárně a výhradně metodikou výzkumu (metodou sběru dat, atd.). Existují zde faktory reálně působící faktory způsobující vychýlení:

- a) zapomnění vlastního volebního chování respondentem, které může mít zejména v posledním desetiletí v ČR velký vliv díky poměrně slabé afinitě voličů k politickým stranám⁹⁰ (CVVM 2011c) a celkově klesající afinitě k tradičním ideologiím, které rozlišují stranické spektrum (CVVM 2011d).
- b) neochota přiznat volbu strany, jejíž volební výsledek je mediálně prezentován jako prohra (tj. například neúčast ve vládě či neúčast Sněmovně) a naopak příklon k vítězům. Tj. efekt spirály mlčení Elisabeth Noelle-Neumannové (1974) a tzv. band-wagon efektu.⁹¹
- c) neochota přiznat neúčast u voleb u respondentů, kteří preferovali stranu, jejíž výsledek je mediálně interpretován jako prohra (viz výše), ale k volbám nešli. Tento jev může vést ke zpětnému nadhodnocení zejména u méně úspěšných stran s neukotvenými voliči – například ČSSD (v roce 2006 i 2010) a Strany zelených (v roce 2010). (CVVM 2011c) (STEM 2010c)
- d) v delším časovém horizontu po volbách neochota přiznat volbu aktuálně vládnoucích stran, u nichž se v horizontu 1 až 3 let po volbách standardně propadá afinita voličů (CVVM 2011c) i důvěra populace (CVVM 2011e)

Předpokládáme-li však, že tyto fenomény primárně souvisí s reflexí vlastního volebního chování a nikoli s metodikou, kterou je tato reflexe dotazována, nemělo by se vychýlení přiznané volby od skutečných výsledků v minulých volbách v jeden časový okamžik výrazně lišit podle různých metod sběru dat. Takovou odlišnost však v našem srovnání osobního a telefonního sběru pozorujeme.

Ve vzorku sebraným osobním sběrem jsou v dubnu a květnu 2010 je podreprezentována volba ODS (- 6 p.b.) a neparlamentních stran (- 5,6 p.b.) a nadreprezentovány preference všech ostatních parlamentních stran kromě KSČM (tedy ČSSD, KDU-ČSL, SZ). Vychýlení způsobené telefonním sběrem je takřka opačné – přiznaná volba ODS je oproti reálnému

⁹⁰ V dlouhodobé perspektivě poslední dekády měla necelá polovina voličů silný vztah (souhlasí ve všem či ve většině postojů) ke straně, kterou (by) volili. (CVVM 2011c, s.3). Dá se tedy předpokládat, že i výrazná část minulých voleb respondentů byla volbami výběrem menšího zla z aktuální politické nabídky, či volbou založenou na aktuálních osobnostech a rétorice strany – a tedy volbou z dlouhodobého hlediska snáze zapomenutelnou.

⁹¹ Lze předpokládat, že tento jev bude existovat zejména v době následující těsně po volbách, protože později je převážen klesající důvěrou ve strany, které sestavily vládní koalici. Viz bod c.

volebnímu výsledku z roku 2006 silně nadreprezentována (+8,8 p.b.) a ostatní strany (zejména KSČM a malé strany) jsou podreprezentovány.

TAB 3.1.6 Reprezentativita vzorků z hlediska strany volené ve volbách do PSP ČR 2006

Strana	Teoretické četnost (volby 2006)	Odchylka evidovaných od teoretických četnosti (v p.b.)						
		CAPI 04/10	CAPI 05*/10	CAPI 04+05*/10	CATI 04/10	CATI 05*/10	CATI 04+05*/10	CAPI+CATI 04+05*/10
ODS	35,4	-7,2	-4,5	-6,0	8,6	8,9	8,7	2,5
ČSSD	32,3	2,9	5,2	3,9	0,0	-3,0	-1,3	0,9
KSČM	12,8	2,4	-1,4	0,7	-4,1	-3,5	-3,9	-1,9
KDU-ČSL	7,2	3,1	3,9	3,5	-2,0	-0,3	-1,2	0,8
SZ	6,3	4,1	2,8	3,5	1,3	2,0	1,6	2,4
Ostatní strany	6,0	-5,3	-6,0	-5,6	-3,9	-4,1	-4,0	-4,7
Suma abs. odchylek		24,9	23,8	23,1	20,0	21,7	20,7	13,2

**V květnu 2010 byly použity rozhovory započaté před začátkem voleb (tj. do 28.5.2010, 14:00). Na reprezentativitu květnových vzorků to může mít vliv (dotazování specifických PSU na konci osobního sběru, doplňování kvót na konci telefonického sběru).*

POZN: Pro barevné označení silných odchýlení je použit stejný systém jako v kapitola 2.1 (silné odchýlení = odchýlení mimo konfidenční interval u daného kvantilu pro referenční pravděpodobnostní výběr N=1000 s účastí 62,6 – tedy n=626).

Ve vzorku sebraným osobním sběrem v červenci 2010 jsou oproti výsledkům z voleb 2010 podreprezentovány pravicové strany (ODS a TOP09 dohromady -4,9 p.b.) a nadreprezentovány hlasy parlamentní levice (ČSSD a KSČM dohromady + 8,6 p.b.). Vychýlení způsobené telefonním sběrem je opět takřka opačné – přiznaná volba koaličních stran je oproti reálnému volebnímu nadreprezentována (celkem + 8 p.b.) a volba levicové parlamentní opozice je podreprezentována (dohromady -7 p.b.).

Obě metodiky se případech obou srovnání (TAB 3.1.5 i 3.1.7) „shodují“ v podreprezentování přiznané volby neparlamentních stran, přičemž tento jev je silnější s rostoucím odstupem od voleb.⁹² Z toho lze usuzovat, že hlavním důvodem podreprezentování malých stran je zapomnění volby (fenomén a viz výše), které je u neparlamentních stran pravděpodobnější díky menší medializaci činnosti těchto stran po volbách, zánikem některých z nich, atd.

⁹² Tj., že je silnější při srovnání výsledků z roku 2006 s přiznanou volbou o 4 roky poté, než výsledků 2010 s přiznanou volbou v červenci téhož (1-2 měsíce poté).

TAB 3.1.7 Reprezentativita vzorků z hlediska strany volené ve volbách do PSP ČR 2010

Pracovní status	Teoretické četnost (volby 2010)	Odchylka evidovaných od teoretických četnosti (v p.b.)		
		CAPÍ 07/10	CATÍ 07/10	CAPÍ + CATÍ 07/10
ČSSD	22,1	6,4	-1,0	2,8
ODS	20,2	-1,9	2,8	0,4
TOP09	16,7	-3,0	1,3	-0,9
KSČM	11,3	2,2	-6,0	-1,8
VV	10,9	0,2	3,9	2,0
KDU-ČSL	4,4	0,6	-0,5	0,1
SPOZ	4,3	-0,8	-0,8	-0,8
Suverenita	3,7	-2,0	0,1	-1,0
SZ	2,4	1,5	0,8	1,1
Jiná strana nebo hnutí	4,0	-3,3	-0,7	-2,0
Suma absolutních odchylek		22,0	18,0	12,9

POZN: Pro barevné označení silných odchýlení je použit stejný systém jako v kapitola 2.1 (silné odchýlení = odchýlení mimo konfidenční interval u daného kvantilu pro referenční pravděpodobnostní výběr N=1000 s účastí 62,6 – tedy n=626).

Celkově se tedy dá konstatovat, že telefonní sběr dat vede k otázce na minulé volební chování k vychýlení směrem k pravicově orientovaným voličům a osobní sběr naopak k levicově a středově orientovaným voličům. Tento jev je přitom ve specifických případech silněji narušen či posílen dalšími faktory (viz výše), které je však díky možnému protikladnému působení a nedostatku dotazového aparátu těžké jednoznačně identifikovat.⁹³ Rozdílné vychýlení rozdílných typů sběrů dat v popisu minulé volebního chování se dá vysvětlit:

- Základními manifestními faktory** popsanými v kapitolách 3.1.2.1 a 3.1.2.3 – tedy tím, že osobní sběr vedl k mírnému nadhodnocení starších lidí, moravských regionů a žen ve vzorku, zatímco telefonní sběr měl v některých případech mírné opačné vychýlení a jindy shodné ale méně výrazné (viz kapitoly 3.1.2.1 a 3.1.2.3).
- Specifickými manifestními faktory** popsanými v kapitole 3.1.2.3 – osobní sběr vede k nadhodnocení ekonomicky neaktivních (důchodci) a podhodnocení OSVČ a podnikatelů bez zaměstnanců. Telefonní výzkum naopak tyto skupiny reprezentuje

⁹³ Ze srovnání v TAB 3.1.6 se dá například usuzovat na neochotu respondentů přiznat dlouhatou afinitu ke KSČM, což se projevuje výrazněji poklesem přiznané volby ve velmi vzdálených volbách, než ve volbách aktuálních či právě minulých, kde se dá výběr strany argumentovat nízko-afinitním chováním (negativní volba proti ostatním stranám, atd.). Tento faktor by vysvětloval, proč je v případě srovnání s volbami 2006 u osobního sběru nadreprezentování KSČM nižší než ČSSD a u telefonního sběru podreprezentování KSČM silnější než ČSSD. Naopak nadreprezentování SZ v obou sběrech a obou srovnáních a VV v obou sběrech ve srovnání pro rok 2010 se dá vysvětlit malou ukotveností voličů těchto stran, a tedy relativně větším počtem voličů, kteří danou stranu nakonec nevolili kvůli neúčasti u voleb, již však odmítají přiznat. Ověření těchto hypotetických vysvětlení by však vyžadovala samostatný výzkum.

dobře a mírně podreprezentovává zaměstnance. Osobní výzkum rovněž vede k nadreprezentování lidí nevyužívajících internet, zatímco telefonní výzkum nadreprezentovává uživatele.

- c) **Latentními faktory**, které ovlivňují volební preference a zároveň podmiňují rozdílnou ochotu / možnosti respondenta účastnit se osobního, respektive telefonního výzkumu. Tyto faktory bohužel nelze přesně identifikovat a jejich vliv analyzovat, protože chybí v proměnných jednoho (např. volnočasová aktivita) či obou sběrů (např. psychologický profil respondenta). Dá se však předpokládat, že například tendence respondenta trávit volný čas mimo domov, která podle našeho dřívějšího zjištění souvisí s politickými preferencemi (viz kapitola 2.3), více omezuje možnost a ochotu k účasti při osobním výzkumu v domácnostech.

U **Základních manifestních faktorů** (základní sociodem. veličiny) lze předpokládat, že u vychýlení bude vyrovnáno standardním sociodemografickým vážením orientovaným na tyto ukazatele. U **Specifických manifestních faktorů** a **Latentních faktorů** však tento předpoklad plně neplatí – jak ukázala analýza v kapitole 2.3, tyto ukazatele přinášejí informace přínosné pro predikci volebních preferencí, které nejsou plně nahraditelné základními sociodemografickými faktory.

Důležitým závěry této kapitoly jsou z hlediska popisu minulého volebního chování zjištění:

- 1) **Osobní a telefonní sběr dat produkuje rozdílný typ vychýlení v různých hlediscích reprezentativity, přičemž často se jedná o vychýlení protichůdné** (intenzita využívání internetu, minulá volba, pracovní status, atd.)
- 2) **Ani kvótní předpis (CATI) plně nezaručuje reprezentativitu z hlediska specifických faktorů, které ovlivňují volební chování a nejsou součástí kvót** (suma absolutních je u CATI v těchto případech často vyšší než u náhodného výběru v CAPI).
- 3) **Spojením vzorků z obou sběrů dat se celkové velikosti vychýlení v řadě ukazatelů redukuje** (internet, minulá volba, pracovní status, atd.)
- 4) **Celkově neplatí očekávání, že s rostoucí dobou od voleb se bude výrazně zvětšovat odchylka v přiznaném minulém volebním chování od výsledků voleb.**⁹⁴ Rozdíl je hlavně v narůstajícím podreprezentování malých stran s odstupem od voleb (díky zapomnění).

⁹⁴ Tento jev se dá částečně vysvětlit faktem, že brzy po volbách jsou silnější jevy, v nichž respondent záměrně zkresluje vlastní minulé volební chování (spirála mlčení, band-wagon effect, nepřiznání neúčasti u voleb, která poškodila preferovanou stranu) – viz body b) až d) na straně výše.

Zjištění 1) až 3) jsou důležitá pro následující kapitolu (3.1.3), protože díky výrazné statistické souvislosti mezi minulým a aktuálním volebním chováním jednotlivců, lze předpokládat, že spojení obou metodik sběru dat povede také k lepší predikci volebních výsledků. Zjištění 4) je důležité pro kapitolu 3.2, v níž budeme zkoumat možnosti vážení datového souboru s ohledem na minulé volební chování.

3.1.3 Predikční kvalita volebních modelů vzniklých kombinovanou metodologií

V této kapitole ověřujeme hypotézu, že sloučením dat z osobního a telefonního sběru získáme při použití shodné a základní metodologie konstrukce volebního modelu lepší predikci výsledků voleb, než z obou metodik sběru jednotlivě. Společnost MEDIAN v dubnu a květnu před volbami 2010 používala při osobním sběru a telefonním sběru stejný dotazový aparát (viz Příloha 1), takže data jsou slučitelná.

Srovnání nevážených dat (TAB 3.1.8) nelze vykládat jako přímou analýzu predikční kvality obou typů výzkumy.⁹⁵ Zejména model vycházejícího z osobního sběru dat (*vícestupňový stratifikovaný náhodný adresní výběr*) počítá s post-stratifikačním vážením, které zde ještě neuplatňujeme.

Srovnání na nevážených datech však ukazuje, že spojení dat z osobního a telefonního výzkumu predikční kvalitu modelu jednoznačně prospívá. Všechna tři základní kritéria (suma čtverců odchylek, průměrná absolutní odchylka, maximální absolutní odchylka) v CATI+CAPI modelu klesají pod úroveň obou modelů z CATI a CAPI odděleně.

⁹⁵ Proto ještě v této kapitole nevyužijeme analýzu kvality predikce podle sady kodifikovaných, jak jsme ji definovali v kapitole 2.1.

3.1.8 Volební modely MEDIAN (duben až květen 2010) vzniklé z nevážených dat

Strana	Výsledky voleb	DUBEN 2010			DUBEN-KVĚTEN 2010 (plovoucí)			KVĚTEN 2010*		
		CATI nevážené	CAPI nevážené	CAPI+CATI nevážené	CATI Nevážené	CAPI nevážené	CAPI+CATI nevážené	CATI nevážené	CAPI nevážené	CAPI+CATI nevážené
ČSSD	22,1	25,4	30,0	27,8	23,2	26,6	24,9	19,7	23,2	21,5
ODS	20,2	31,2	15,3	22,7	26,2	13,0	19,6	23,6	12,8	18,0
TOP 09	16,7	11,6	8,0	9,7	14,3	9,5	11,9	17,8	14,3	16,0
KSČM	11,3	9,8	15,1	12,6	11,1	13,8	12,4	11,3	12,1	11,7
VV	10,9	10,2	7,8	9,0	12,4	8,4	10,4	11,0	11,9	11,5
KDU-ČSL	4,4	3,3	8,6	6,2	3,9	11,0	7,5	4,5	8,2	6,4
SPOZ	4,3	1,8	7,3	4,7	1,3	6,3	3,8	3,7	6,8	5,3
Suverenita	3,7	,7	2,2	1,5	1,3	2,4	1,8	1,3	3,1	2,3
SZ	2,4	1,1	3,7	2,5	2,0	5,2	3,6	3,1	4,4	3,8
Ostatní	4,0	4,9	2,0	3,3	4,3	3,9	4,1	3,9	3,1	3,5
Suma čtverců odchylek		179	220,3	102,5	61,0	193,8	47,3	25,5	88,7	15,3
Prům. abs. odchylka		3	4	2,4	1,8	3,7	1,7	1,1	2,2	1,1
Max. abs. odchylka		11	8,7	7	6	7,2	4,8	3,4	7,4	2,2

*V měsíci květnu byly použity rozhovory započaté před začátkem voleb (tj. do 28.5.2010, 14:00). Na reprezentativitu obou modelů to může mít vliv (dotazování specifických PSU na konci osobního sběru, doplňování kvót na konci telefonického sběru).

POZN: Pro barevné označení silně se odchylujících buněk je použit stejný systém jako v kapitole 2.1, který vychází z konfidenčních intervalů adjustovaných podle časové vzdálenosti od voleb.

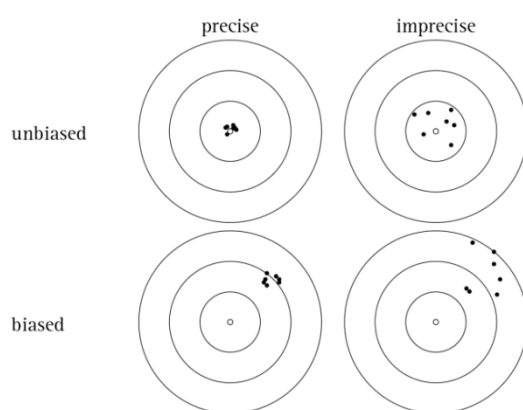
3.2 Optimalizace volebního modelu alternativním vážením

V kapitole 3.1 jsme ukázali, že úroveň reprezentativity⁹⁶ vzorků z osobního (CAPI) sběru a výsledně také predikční kvalita volebního modelu se dá zlepšit připojením datového zdroje s odlišným typem sběru dat (CATI). Alternativou či komplementem tohoto procesu je post-stratifikační vážení.

Při použití vážení pro zkvalitnění predikce odchylek volebního modelu, je třeba mít na paměti jednu věc: Vážením lze zvýšit úroveň reprezentativity vzorku a eliminovat systematické odchylky predikce (např. přidělením váhy >1 přiblížit poměr podnikatelů a OSVČ ve vzorku zastoupení v populaci a tím narovnat podreprezentování lidí preferujících pravicové strany). Děje se tak ale za cenu snižování efektivní velikosti vzorku (Ganninger 2006). Omezení efektivní velikosti vzorku přitom vede ceteris paribus k zvyšování náhodného složky výběrové chyby a tedy i k zvyšování rozptylu odhadů.

Využijeme-li Gannigerovo schéma (2011) z metodologie ESS, lze obecně říci, že vážením dat přesouváme naše odhady (volební model) směrem od stavu vychýlený-přesný (biased-precise) ke stavu nevychýlený-nepřesný (unbiased-imprecise).

OBR 3.2.1 Typy odhadů podle přesnosti a vychýlení



Zdroj: Ganninger 2011

Tento fakt se může projevit mimo jiné zvýšenou oscilací výsledků volebního modelu v čase. Statistikové zabývající se vážením (M. Anděl, P. Hebák, P. Charamza, J. Neustadt) však upozorňují, že to není důsledkem nevyhnutelným a závisí na poměru,

⁹⁶ Chápeme-li reprezentativitu jako vlastnost struktury vzorku vzhledem k cílové populaci, která může mít úroveň (Krejčí 2008, s. 9). Reprezentativitu v druhém slova smyslu (Soukup & Rabušic 2007) (tedy vlastnost výběru), samozřejmě ex-post vážením ovlivnit nelze.

jakým vážením zvyšuje náhodnou složku výběrové chyby a redukuje systematickou složku.⁹⁷ (Neustadt 200X, s. 46)

V analýze výsledků vážení jsme tedy sledovali: a) Nakolik se různé typy vážení (tedy změna struktury vzorku směrem k reprezentativitě podle různých ukazatelů) projevují redukcí systematického vychýlení a zkvalitněním odhadu volebního chování a redukcí vychýlení odhadu; b) Jaký mají vážení vliv reálný vliv na oscilaci výsledků volebního modelu v čase.

3.2.1 Zkvalitnění predikčních vlastností volebního modelu vážením

První otázku lze zkoumat zejména na datech sebraných před volbami 2010 analýzou, nakolik se při použití jednotlivých typů vážení, zlepšují predikční vlastnosti modelu, které jsme definovali v kapitole 2.1. V případě, že ideálně predikující vážení nebudou obsahovat proměnnou minulého volebního chování⁹⁸, lze druhotně zjištění ověřit i zpětně – tedy nakolik po volbách při využití jednotlivých vážení odpovídá struktura odpovědí na volenou stranu skutečnému výsledku voleb.

Společnost MEDIAN používala před volbami 2010 na datech z CAPI sběru volebních preferencí standardní sociodemografické vážení podle 5 jednorozměrných kategorií (pohlaví, věk, vzdělání, kraj NUTS3, velikost místa bydliště) a 3 sdružených kategorií (pohlaví x kraj, pohlaví x věk, velikost místa bydliště x region NUTS 2).⁹⁹

Při testování alternativních vážení budeme brát toto vážení jako referenční – minimálním požadavkem na vážení je, aby u všech zkoumaných datových úseků (a z nich vycházejících volebních modelů) kvalitu predikce oproti standardnímu sociodemografickému vážení buď zlepšovalo, či nezhoršovalo (pokud v jiných měsících kvalitu výrazně zlepšuje). Analýza byla provedena primárně na datech z CAPI sběru, které pocházejí z vícestupňového stratifikovaného náhodného adresního výběru.¹⁰⁰ Následně byla aplikována i na spojená CAPI+CATI data, která vykazují vyšší reprezentativitu ve většině ukazatelů (kapitola 3.1.1) a

⁹⁷ Výsledky modelu neoscilují jen díky náhodné složce výběrové chyby, která je přímo závislá na efektivní velikosti vzorku, ale také díky proměnlivé síle systematického vychýlení, které výzkum v čase produkuje. Pokud vážení kvalitně vyrovná toto systematické vychýlení, mohou výsledky volebního modelu v čase stabilnější i přes sníženou efektivní velikost vzorku. V porovnávání vážení se tedy nezaměřujeme na statistické ukazatele vah (např. standardní odchylku proměnné vah), ale na reálné důsledky vážení, tedy na oscilaci volebního modelu v čase.

⁹⁸ Zde pochopitelně s výjimkou vážení, které obsahuje vážení podle minulého volebního chování, a pokud je dobře provedeno musí minulým volbám odpovídat.

⁹⁹ Vážení podle volebního chování v minulých volbách do PSP společnost v průběhu volebního období 2006-2010 redukovala posléze zcela opustila, a to díky v dané době poměrně oprávněným pochybnostem, zda se jedná o správný prediktor současných preferencí, vycházejících z faktů: 1) nízké a dlouhodobě klesající úrovně identifikace a věrnosti voličů k politickým stranám (CVVM 2011c), 2) výrazné proměny volebního spektra mezi lety 2006 a 2010 (dvě nové silné strany, rozpad KDU-ČSL vznikem TOP09), 3) poměrně výrazné personální obměny v čele stran (jiný předseda ODS, SZ, KDU-ČSL), 4) účast nových silných osobností ve volbách do Sněmovny (Karel Schwarzenberg, Radek John), 5) přítomností 3 odlišných voleb (krajské volby 2008, volby do EP 2009, Senátní volby 2006/2007/2008) v době od voleb do PSP ČR 2006, která mohly narušit vědomí respondentů, jak hlasovali v těchto posledních volbách do Sněmovny.

¹⁰⁰ Pro pravděpodobnostního designu výběru, pro který se vážení doporučuje (Soukup & Rabušic 2007).

i bez vážení docházejí k výrazně lepší predikci volebních výsledků, než oba typy sběrů samostatně (3.1.2).

Vyzkoušeno celkem 8 alternativních typů vážení, které nevyužívají ukazatel minulého volebního chování (strana volená ve volbách do PSP v roce 2006). Při konstrukci vážení jsme vycházeli ze zjištění z kapitoly 2.3 a preferovali ty sociodemografické (jednoduché i kombinované) a life-stylové faktory, které nejvíce souvisí s volebními preferencemi a nejlépe je vysvětlují. Jak však připomínají Soukup a Rabušic (2007, s. 392), problémy vážení nejsou jen teoretické, ale i praktické – vážení je omezeno jen ukazatele a jejich kombinace (sdružené kategorie), u nichž jsme schopni najít relevantní a aktuální popis populace.¹⁰¹

Kromě základních sociodemografických proměnných (pohlaví, věk, vzdělání, kraj, VMB), které jsou jednorozměrně použity v každém vážení, jsme tedy do jednotlivých vážení zařazovali kombinovaně ukazatele:

- Intenzita užívání internetu (zdroj t.č.: Eurostat 2010) (3 kategorie + mimo rozmezí 16-74 let)¹⁰²
- Příjem domácnosti na jednoho člena (zdroj t.č.: ČSU 2010d)¹⁰³ (9 kategorií + neví/odmítl odpovědět)¹⁰⁴
- Počet členů domácnosti (zdroj t.č.: ČSU 2010d) (4 kategorie¹⁰⁵)
- Pracovní zařazení (zdroj t.č.: ČSU 2010f) (8 kategorií)
- Věk x region (zdroj t.č.: ČSU 2010g) (4 kategorie věku x 3 regionů¹⁰⁶)
- Věk x pohlaví (zdroj t.č.: ČSU 2010e) (7 kategorií věku x 2 kategorie pohlaví)

101 Dalším problémem je nerozpornost zejména křížených tabulek související s jejich aktuálností. Například sdružené kategorie velikosti místa bydliště a věku ČSÚ jen ze Sčítání lidu – tedy v letech 2001 a 2011. I tato ne zcela aktuální (2001) či příliš aktuální (2011) informace by mohla být pro vážení přínosná. Sdružené teoretické četnosti věku a VMB bychom ale museli poměrně složitě přepočítávat, aby nebyly v rozporu s teoretickými četnostmi věku a VMB jednotlivě pro rok 2010. Pro uchování věrohodnosti a relativní jednoduchosti vážení je tak lepší vůbec kombinaci věk x VMB do vážení nezařadit.

102 Opора Eurostatu popisuje populaci 16-74 let. Respondentům ze vzorku mimo tuto věkovou hranici je intenzita užívání internetu překódována na 4 hodnotu „mimo rozmezí 16-74 let“. Tabulka teoretických četností pro vážení kopíruje zastoupení této možnosti v nevážených datech, které jsou právě předmětem vážení, a ostatní 3 kategorie jsou rozpočteny, aby celková suma tvořila 100 %. Respondenti mimo rozpětí 16-74 let tak nejsou podle této proměnné váženi.

103 Ostatní příjmové ukazatele, které s volebními preferencemi souvisí (příjem respondenta, příjem domácnosti – viz kapitola 2.3) nejsou využitelné. ČSÚ sice publikuje údaje o příjmu domácností, ale nekříží je přes počet členů v domácnosti – četnosti tedy nelze převést na populaci jedinců. (SILC 2010) Údaje o příjmu jednotlivců jsou zase roztrženy mezi řadu zdrojů (ČSÚ, MPSV), odděleně pro různé části populace (zaměstnaní, podnikatelé, důchodci, atd.) a navíc většinou publikovány pomocí průměru, mediánu či kvantilového rozdělení. Vytvořit tak relevantní tabulku teoretických četností pro příjmy jednotlivce by bylo téma na samostatnou práci.

104 Tabulka teoretických četností vážení v tomto případě obsahuje i kategorii „neví/odmítl“. Ta svou hodnotou kopíruje zastoupení této možnosti v nevážených datech, které jsou právě předmětem vážení, a ostatní kategorie jsou rozpočteny, aby celková suma tvořila 100 %. Respondenti, kteří neuvedli relevantní informaci o příjmu tedy nejsou podle této proměnné váženi. Stejně postupujeme i u odmítnutí a zapomnění při vážení na minulé volební chování.

105 1 člen, 2 členové, 3 členové, 4 a více členů.

106 Praha + Stř. Čechy, zbytek Čech, Morava. Podrobnější členění by tvořilo pro vážení měsíčních dat přílišné množství sdružených kategorií (např. 54 při použití NUTS3).

- Věk x vzdělání (zdroj t.č.: ČSU 2010c, ČSU 2001) (4 kategorie věku x 2 kategorie vzdělání)¹⁰⁷

Následně byla do vážení zařazena i proměnná v20 (strana volená v posledních volbách do PSP ČR v roce 2006), a to jako dodatečná jednorozměrná tabulka ke standardnímu sociodemografickému vážení a dvěma alternativním vážením, které produkovaly nejlepší výsledky (viz níže). Celkově tak bylo s referenčním standardním sociodemografickým vážením porovnáno 11 druhů alternativního vážení a to vždy na třech datových úsecích (duben 2010, plovoucí data duben-květen 2010, květen 2010¹⁰⁸). Všechna vážení byla provedena vážícím softwarem používaným společností MEDIAN, jehož autorem je statistik Jan Neustadt (Neustadt 1997) (MEDIAN 2003). Pro udržení srovnatelnosti všech vážení byly zafixovány základní kritéria vážení, tak jak byly využity při standardním sociodemografickém vážení před volbami 2010.¹⁰⁹

Podmínku zpřesnění předvolební predikce na všech zkoumaných datových úsecích volebního modelu vycházejícího z osobního (CAPI) sběru dat splňují celkem čtyři typy vážení, z nichž dvě nevyužívají ukazatel minulého volebního chování (strana volená ve volbách do PSP v roce 2006) a dvě ho využívají. Popis těchto vážení, kterými se budeme zabývat, zobrazuje TAB 3.2.1. Popis všech vážení viz Příloha 5.

Vážení typu *Alternative_6* i *Alternative_8* kromě standardních sociodemografií zařazují do vážení jednorozměrné speciální sociodemografické a life-stylové faktory, u nichž jsme v kapitole 2.3 prokázali souvislost s volebními preferencemi (intenzita užívání internetu, pracovní zařazení, čistý měsíční příjem na člena domácnosti). Vzájemně se liší v tom, že vážení *Alternative_6* plně využívá sdružená sociodemografické kategorie z referenčního vážení, zatímco vážení *Alternative_8* jednu z nich (pohlaví x kraj), která mají podle zjištění z kapitoly 2. 3. nejmenší souvislost s volebními preferencemi, nahrazuje silněji souvisejícími sdruženými kategoriemi sociodemografií (věk x oblast, věk x vzdělání). Vážení *Alternative_V6* a *Alternative_V8* poté vznikla přidáním ukazatele strany volené v posledních volbách do PSP ČR do obou alternativních vážení.

¹⁰⁷ Stejně jako v kapitole 3.1. vycházíme z údajů VŠPS pro populaci 15+ a na cílovou populaci 12+ je přepočítáme pomocí předpokladu, že všichni lidé 12-14 mají základníškolské vzdělání a lidé nad 80 let mají strukturu vzdělání shodnou jako ve sčítání lidu 2001 (ČSU 2001).

¹⁰⁸ Rozhovory sebrané do začátku voleb, tedy do 28.5. 14:00.

¹⁰⁹ Omezení hodnoty velikosti vah na 0,3 až 3, které vede k redukci vlivu vážení na zmenšení efektivní velikosti vzorku (Neustadt 1997, s.20-24); suma vah odpovídá počtu respondentů; zastavovací pravidlo max. pěti iteračních cyklů vážení. Vážen je celý vzorek pro teoretické četnosti populace 12+. Z váženého vzorku je pro výstupy týkající se preferencí posléze separována část 18+ a váhy znormovány koeficientem tak, aby subvzorek oprávněných voličů 18+ měl stejnou velikost jako v neváženém souboru. Tento postup je nutný zejména proto, že řada charakteristik populace použitých pro vážení (např. údaje o vzdělání a pracovním statusu z VŠPS 2010) se týká populace 15+, z čehož lze jednoduše přepočítat teoretické četnosti pro populaci 12+ (např. předpokladem, že všichni lidé 12-14 let nemají základní vzdělání a jsou studenty), ale jen obtížně lze provést přepočet na populaci 18+.

TAB 3.2.1 Popis vážení (referenční a alternativní) a jeho vlivu na datech z CAPI sběru

Č.	Název	Vstupující proměnné	Vliv na zlepšení predikce
1	Socio_1	1D – pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 2D - pohlaví X věk, pohlaví X kraj, region (8) X VMB	nulový (referenční vážení standardní sociodemografické v.)
2	Socio_V1	1D - pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 1D - chování ve volbách 2006 (9) 2D - pohlaví X věk, pohlaví X kraj, region (8) X VMB	Zlepšení oproti (1) na všech třech úsecích
3	Alternative_6	1D - pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D - věk (7) x pohlaví, pohlaví X kraj, region (8) X VMB	Zlepšení oproti (1) na všech třech úsecích, zlepšení oproti (2) na dvou úsecích
4	Alternative_8	1D - pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D – věk (4) x oblast (3), věk (7) x pohlaví, věk (7) x vzdělání, region(8) X VMB	Zlepšení oproti (1) na všech třech úsecích, zlepšení či udržení kvality (2) na dvou úsecích
5	Alternative_V6	1D - pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 1D - chování ve volbách 2006 (9) 2D - věk (7) x pohlaví, pohlaví X kraj, region (8) X VMB	Zlepšení oproti (1) i (2) na všech třech úsecích
6	Alternative_V8	1D - pohlaví (2), vzdělání (4), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 1D - chování ve volbách 2006 (9) 2D – věk (4) x oblast (3), věk (7) x pohlaví, věk (7) x vzdělání, region (8) X VMB	Zlepšení oproti (1) i (2) na všech třech úsecích

Právě tyto dvě metody vážení produkují nejlepší výsledky, co se týče zlepšení celkové predikční kvality modelů v jednotlivých měsících. Na rozdíl od spojení osobního (CAPI) sběru s telefonickým (CATI), kterým do modelu reálně dostáváme další respondenty/informace z výzkumu produkující jiné typy vychýlení dotázaných vzorků vzhledem k populaci (viz kapitola 3.1), se však projevuje to, že vážení je čistě statistickou korekční metodou.

Vliv na zlepšení kvality predikce je omezený, ale přesto ne zanedbatelný. Ve všech zkoumaných obdobích dochází k redukci nadhodnocení výsledků SPOZ a KDU-ČSL. Zároveň ve všech obdobích dochází k realističtějšímu odhadu celkové podpory pravice (ODS+TOP09+VV) ve společnosti – celková hodnota pravice při alternativním vážení, které se zaměřuje na reprezentativitu z hlediska faktorů souvisejících s volebním chováním, posiluje o 3 až 5 procentních bodů. (viz TAB 3.2.2)

Posledního volebního modelu z osobního sběru, který se zakládá na úplných měsíčních datech (model z plovoucího období duben-květen 2010) tak už při alternativních váženích správně predikuje, že zisk budoucí pravicové koalice ve volbách bude přinejmenším stejný jako zisk levicových stran ČSSD a KSČM a středně velké strany KDU-ČSL se budou pohybovat okolo 5% kvalifikační hranice do Sněmovny.

TAB 3.2.2 Srovnání volebních modelů ze standardního sociodemografického a nejvhodnějších alternativních vážení

Strana	Volby	CAPI duben 2010			CAPI duben-květen 2010			CAPI květen 2010*		
		Socio_1	Alternative v6	Alternative v8	Socio_1	Alternative v6	Alternative v8	Socio_1	Alternative v6	Alternative v8
ČSSD	22,1	26,2	27,8	27,7	23,7	24,3	24,5	20,7	18,8	20,1
ODS	20,2	19,0	21,6	21,5	15,3	18,2	18,8	11,9	14,1	13,7
TOP 09	16,7	10,7	11,0	10,5	13,9	13,7	13,4	17,5	18,4	19,0
KSČM	11,3	13,3	11,4	12,1	11,9	14,5	14,2	9,5	13,5	13,7
VV	10,9	7,6	8,2	7,7	7,4	6,8	6,9	10,2	11,8	11,1
KDU-ČSL	4,4	7,5	5,7	6,1	9,0	6,1	6,4	10,4	5,4	5,5
SPOZ	4,3	6,8	6,2	6,3	5,8	4,8	4,4	7,4	7,6	5,9
Suverenita	3,7	3,1	2,8	2,7	3,0	2,3	2,4	2,8	2,6	2,7
SZ	2,4	3,5	3,3	3,2	5,3	4,4	4,2	5,0	4,6	4,9
Ostatní	4,0	2,2	2,0	2,2	4,8	4,8	4,8	4,6	3,1	3,3
Hodnocení chybovosti										
Součet čtverců odchylek		90,0	84,8	95,0	80,0	54,2	52,4	128,5	75,4	68,5
Prům. abs. odchylka		2,6	2,3	2,5	2,4	2,1	2,0	2,6	2,3	2,0
Max. abs. odchylek		6,0	5,7	6,2	4,9	4,1	4,0	8,3	6,1	6,5
Průměrná relativní odchylka		34,6%	27,9%	29,8%	38,3%	29,5%	27,8%	42,9%	32,7%	29,1%
Maximální relativní odchylka		70,4%	50,4%	46,4%	121,3%	81,7%	75,0%	137,4%	91,5%	106,0%
Rozdíl mezi lídry		5,3	4,3	4,3	6,5	4,2	3,8	6,8	2,8	4,5
Chyby v kvalifikaci stran		1	0	0	1	0	1	2	1	0
Kategorie kvality predikce										
Součet čtverců odchylek		4	4	4	4	3	3	5	4	3
Průměrná abs. odchylka		4	3	3	3	3	3	4	3	3
Maximální abs. odchylka		4	4	4	3	3	3	5	4	4
Průměrná relativní odchylka		4	4	4	5	4	4	5	4	4
Maximální relativní odchylka		4	3	3	7	5	4	7	5	6
Rozdíl mezi lídry		6	5	5	7	5	4	7	3	5
Chyby v kvalifikaci stran		3	1	1	3	1	1	5	3	1
Celkový index kvality predikce		4,1	3,4	3,4	4,6	3,4	3,1	5,4	3,7	3,7

Pozn.: Pro barevné značení silně vychýlených buněk (modře podreprezentované, červeně nadreprezentované) jsou použita kritéria z kapitoly 2.1.

*Z května 2010 byla do modelu zařazena pouze data sebraná do 28.5.14:00. To resultuje ve zmenšení vzorku (celkem 719 respondentů 18+), které je navíc systematické, protože se projevuje výpadkem oblastí stratifikovaného výběru (VMB x kraj) dotazovaných na konci měsíční fáze. **Proto se z květnová CAPI data nedají považovat za plnohodnotný odhad volebního výsledku a slouží jen pro srovnání výsledků vážení.** Rovněž při jejich vážení nebyla použita sdružená kategorie VMB x kraj, protože díky úplné absenci některých sdružených kategorií se nelze přiblížit teoretickým četnostem.

V reálném volebním výsledku se zisk pravice a levice lišil ještě dramatičtěji (47,8 % v. 33,4 %). Rozdíl mezi odhadem produkovaným v CAPI modelu z dubna-května s optimalizovaným vážením lze přičítat jednak omezeným možnostem vážení zlepšovat kvalitu modelu, ale zejména rozhodovacímu procesu voličů v posledních 14 dnech. To, zda mohl finální vliv rozhodnutí dosud nerozhodných voličů částečně předvídat, budeme zkoumat v kapitole 3.3.

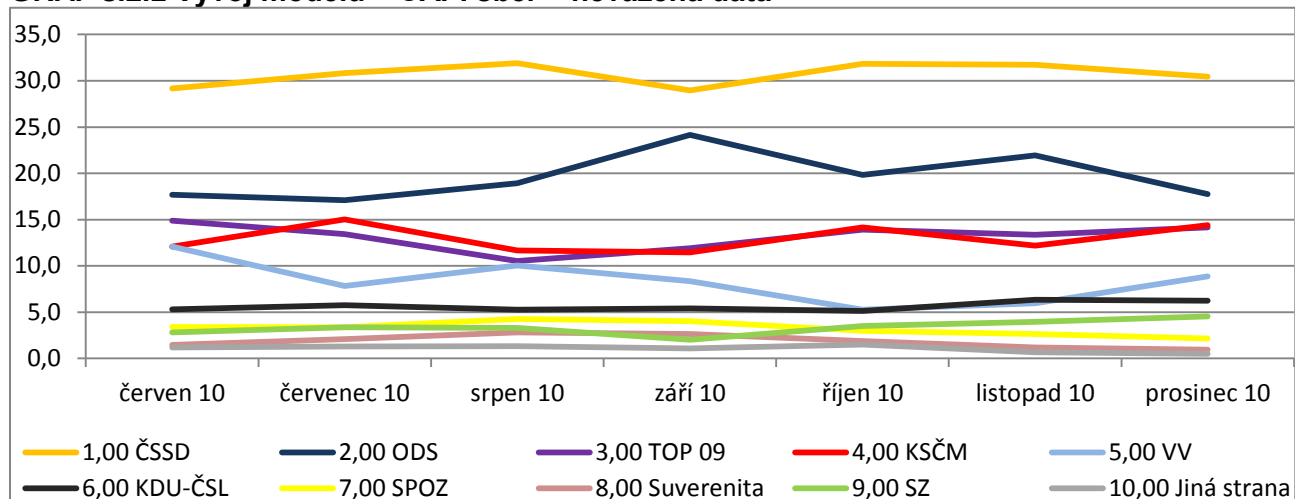
3.2.2 Vliv vážení na oscilaci volebního modelu v čase

Druhá základní otázka této kapitoly je, nakolik se jednotlivá vážení projevují zvýšením oscilace volebního modelu v čase. Zaměření na tento jev přitom nemá jen akademické důvody. Řada analýzy z oblasti sociologie žurnalistiky (Galtung & Rugeová 1968) (Shulz 2006) (Westerstahl & Johansson 1994) (Prokop 2007) se totiž shoduje, že zejména masová zpravodajská média určená nejširšímu publiku mají tendenci preferovat (priming) události se zpravodajskými hodnotami překvapivosti, jednoznačnosti a negativity či konfliktu a rámcovat je tak, aby byly tyto hodnoty v mediovaném textu posíleny (framing). Je tedy logické, že v oblasti volebních výzkumů mají média tendenci upřednostňovat výzkumy ukazující překvapivé posuny preferencí a rámcovat je jako epizodicky jako odhalení náhlého výrazného poklesu/nárůstu reálné podpory, a to zhusta s využitím konfliktualistického framingu pomocí konfrontace politiků dané strany apod. (Wirnitzer 2011) a se sníženým ohledem na výběrovou chybu či na dlouhodobé trendy v preferencích, jakožto faktorů, které konfliktualistický epizodický framing narušují.¹¹⁰

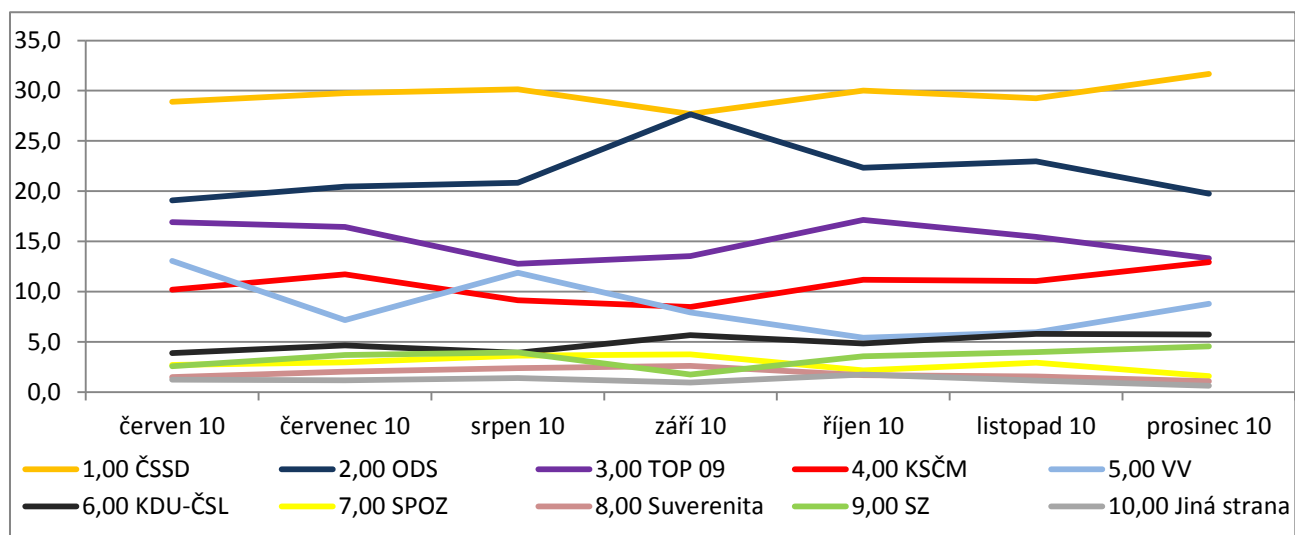
Protože důraz na zpravodajské hodnoty je inherentní charakteristikou většiny masových médií působících v tzv. liberální modelu médií (Siebert & Peterson & Schramm 1956), měly by s ní agentury publikující volební modely a prognózy počítat - nabízet novinářům informace, které redukují přeceňování informační hodnoty náhodné oscilace modelu, a upřednostňovat takové modely, které méně oscilují v čase a dávají méně příležitostí k chybným interpretacím. Grafy 3.2.1 až 3.2.4 ukazují vývoj volebního modelu z osobního (CAPI) sběru v období od prvního měsíce po volbách (červen 2010) do konce tohoto roku (prosinec 2010). Již z vizuální stránky je patrná poněkud mírnější oscilace neváženého modelu.

¹¹⁰ Nakolik je tento postoj vnesouladu s logikou mediální produkce -

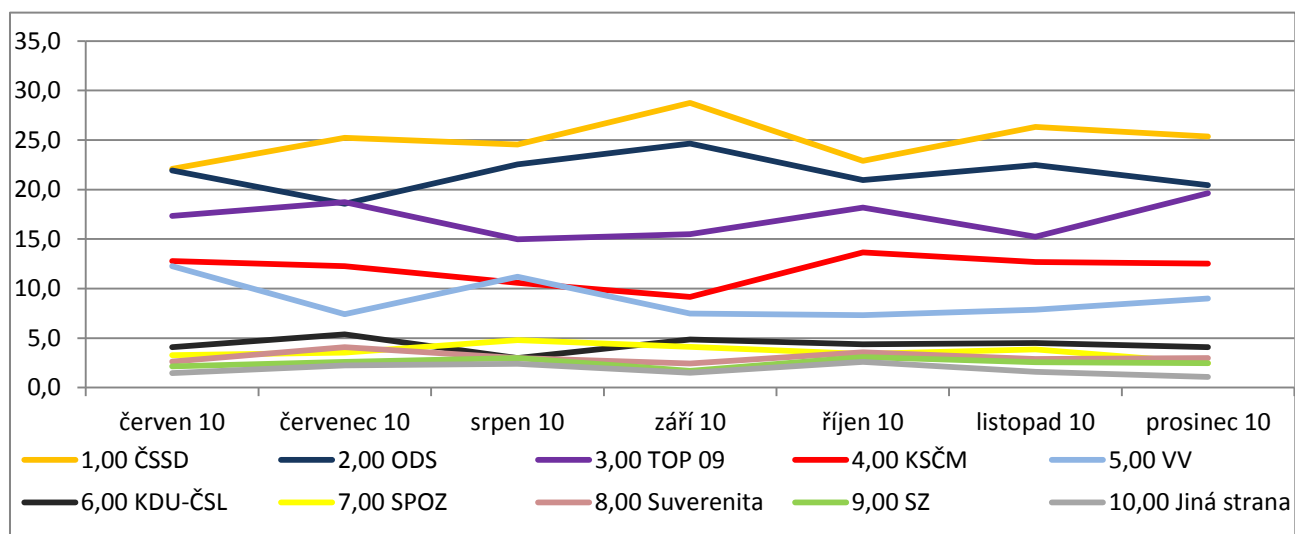
GRAF 3.2.1 Vývoj modelu – CAPI sběr – nevážená data



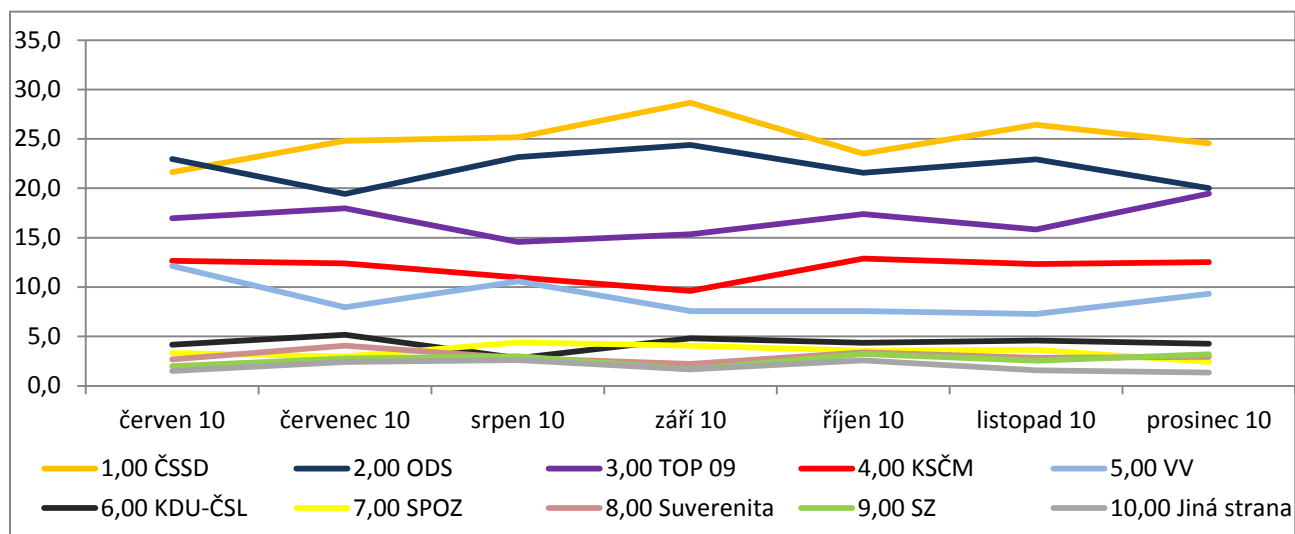
GRAF 3.2.2 Vývoj modelu – CAPI sběr – standardní sociodemografické vážení



GRAF 3.2.3 Vývoj modelu – CAPI sběr – vážení *Alternative_v6*



GRAF 3.2.4 Vývoj modelu - CAPI sběr – vážení *Alternative_v8*



Pro statistický popis oscilace modelů jsme zvolili ukazatele - průměrné absolutní meziměsíční odchyly¹¹¹, průměrného čtverce odchyly¹¹² a průměrné relativní odchyly.¹¹³ (TAB 3.2.3)

Ukazuje se, že vážení, které jsme pojmenovali jako *Aternative_v8* (obsahuje základní jednorozměrnou sociodemografii, region NUTS2 x velikost místa bydliště, intenzitu užívání internetu, pracovní zařazení, příjem domácnosti na člena a volební chování v roce 2010) produkuje podle všech ukazatelů mírně nižší úroveň oscilace než model ze standardního sociodemografického vážení.

TAB 3.2.3 Statistický popis oscilace modelu z osobního sběru (červen – prosinec 2010)

	Bez červen 2010 - volby 2010			S červen 2010 - volby 2010		
Název vážení	Průměrná abs. odchylna	Průměrný čtverec odchylny	Průměrná relativní odchylna	Průměrná abs. odchylna	Průměrný čtverec odchylny	Průměrná relativní odchylna
Bez vážení	1,35	3,41	14,9 %	1,45	4,01	16,5 %
Socio_1	1,51	4,55	18,4 %	1,56	4,89	19,4 %
Alternative_ V6	1,65	4,78	19,2 %	1,56	4,33	18,9 %
Alternative_ V8	1,48	3,72	18,3 %	1,43	3,47	18,2 %

¹¹¹ Odchylnou přitom rozumíme to, nakolik se hodnota jedné strany v modelu v měsíci N liší od hodnoty téže strany v měsíci N-1. Neplatí tedy, že by optimální model měl dosahovat nulové průměrné odchylny, protože pak by musely být preference stran zcela konstantní. Oscilace hodnoty pro jednotlivé strany se ale projeví do zvýšené hodnoty průměrné absolutní meziměsíční odchylny.

¹¹² Není zde vhodné aplikovat klasickou sumu čtverců, protože ta by se zvětšovala úměrně s rostoucím analyzovaným obdobím (počtem měsíců).

¹¹³ Relativní odchylnu kalkulujeme jako: $e = (\text{hodnota strany v měsíci N} - \text{hodnota strany v měsíci N-1}) / \text{zisk strany ve volbách 2010}$.

Pokud započítáme do charakteristiky oscilace jako výchozí hodnotu i odchylky prvního povolebního sběru (červen 2010) od skutečného výsledku voleb, produkuje model s vážením *Alternative_v8* dokonce srovnatelnou či nižší oscilaci než model z nevážených dat.

Zahrnutí této odchylky prvního povolebního modelu od voleb do popisu oscilace je relevantní díky předpokladu, že v mediálním a politickém diskursu budou výsledky povolebních modelů interpretovány jako změna podpory stran od voleb. Penalizuje ovšem v hodnocení oscilace modely, které jsou systematicky vychýlené (nevážený model, model se standardním sociodemografickým vážením).

Alternativní vážení *Alternative_v6* produkuje vyšší oscilaci než vážení *Alternative_v8*.¹¹⁴ Tento výsledek se přitom dal očekávat, vážení *Alternative_v6* totiž obsahuje sdruženou sociodemografickou kategorii (pohlaví x kraj), které s volebními preferencemi souvisí méně než alternativní sdružené sociodemografické kategorie ve vážení *Alternative_v8* (věk x oblast, věk x vzdělání). První vážení obsahuje prvek, u něhož se dá předpokládat posílení náhodné složky výběrové chyby bez většího vlivu na složku systematickou. Proto, přestože celkově není silnější,¹¹⁵ zřejmě způsobuje vážení *Alternative_v6* vyšší oscilaci modelu v čase.

Protože oba typy vážení produkují zhruba srovnatelné zlepšení predikční kvality modelů z CAPI dotazování v měsících těsně před volbami v roce 2010 (viz TAB 3.2.2), lze konstatovat, že nejefektivnějším vážením je při zvážení obou hledisek (zkvalitnění predikce, zmírnění či udržení oscilace modelu v čase) vážení *Alternative_v8*.

Ve srovnání s referenčním standardním sociodemografickým vážením, toto alternativní vážení vyvažuje vzorek podle charakteristik, které s volební preferencemi silněji souvisí (viz kapitola 2.3), zlepšuje volební predikci (kapitola 3.2.1) a vede i k menší oscilaci modelu v čase. Nenabízí se tedy významný relevantní důvod, proč by nemělo být proti standardnímu sociodemografickému vážení preferováno.

¹¹⁴ Bez zahrnutí odchylky prvního povolebního modelu od voleb produkuje vážení *Alternative_v6* také mírně větší oscilaci než referenční sociodemografické vážení. Rozdíl je ovšem tak malý, že by musel být ověřen na delším časovém vývoji modelu.

¹¹⁵ Směrodatná odchylka souboru vah na spojených datech červen 2010 – prosinec 2010 je 0,86 u vážení *Alternative_v6* a 0,87 u vážení *Alternative_v8*.

3.2.2 Aplikace vážení na data ze spojeného sběru (CATI + CAPI)

Při aplikaci vybraných vážení na data ze kombinovaného sběru CAPI+CATI (TAB 3.2.4), se jednoznačně zlepšení predikční kvality. V modelech z dubna 2010 má vážení ve většině ukazatelů srovnatelné výsledky se sociodemografickým, v modelu z května 2010 alternativní model predikci mírně zhoršují - snížení preferencí ODS a zvýšení u KSČM (platí pro duben-květen 2010 i květen 2010).¹¹⁶

Vzhledem k velmi reprezentativní struktuře dat ze spojného CAPI+CATI modelu (viz kapitola 3.1) vážení nemá možnost příliš zlepšit systematické vychýlení a může se tedy projevit jeho posilování náhodné složky vychýlení odhadu.

V případě aplikace vážení na data z kombinovaného sběru však **nelze dělat ukvapené závěry před analýzou, jak lze volební modely optimalizovat zařazováním preferencí dosud pevně nerozhodnutých voličů.** Těch je totiž výrazně větší množství v telefonním (CATI) sběru a toto zařazování může výsledky modelu ze spojených dat významně ovlivnit.

¹¹⁶ Model z plovoucích dat duben-květen 2010 je pak podle očekávání kombinuje vlivy přítomné v dubnu a v květnu.

TAB 3.2.4 Výsledky alternativního vážení ve spojených datech (CAPI + CATI)

Strana	Volby	CAPI + CATI duben 2010				CAPI + CATI duben-květen 2010				CAPI + CATI květen* 2010			
		bez vážení	Socio_1	Alternative _V6	Alternative _V8	bez vážení	Socio_1	Alternative V6	Alternative V8	bez vážení	Socio_1	Alternative v6	Alternative v8
ČSSD	22,1	27,8	26,5	25,9	26,5	24,9	23,6	22,7	22,5	21,5	20,4	20,4	20,7
ODS	20,2	22,7	24,1	22,2	22,4	19,6	21,3	19,7	19,5	18,0	17,9	15,4	15,3
TOP 09	16,7	9,7	10,8	11,6	11,0	11,9	13,4	13,4	13,0	16,0	16,8	16,5	16,0
KSČM	11,3	12,6	11,8	13,8	13,9	12,4	12,1	16,0	15,9	11,7	11,8	14,2	14,3
VV	10,9	9,0	9,2	8,4	8,2	10,4	9,8	9,0	9,1	11,5	10,6	11,1	10,9
KDU-ČSL	4,4	6,2	5,7	5,5	5,4	7,5	6,7	5,7	5,9	6,4	7,0	5,6	5,6
SPOZ	4,3	4,7	4,6	4,8	4,8	3,8	3,2	3,3	3,6	5,3	5,1	5,6	6,1
Suverenita	3,7	1,5	1,5	1,4	1,4	1,8	1,8	1,8	1,9	2,3	2,2	2,6	2,5
SZ	2,4	2,5	2,4	2,4	2,4	3,6	3,7	3,1	3,3	3,8	4,4	4,2	4,0
Ostatní	4	3,3	3,4	4,0	4,0	4,1	4,3	5,4	5,4	3,5	3,8	4,4	4,5
Hodnocení chybovosti													
Součet čtverců odchylek		102,6	78,7	63,9	76,9	47,3	28,2	46,1	47	15,3	22,4	42	44,2
Prům. abs. odchylka		2,4	2,1	2	2,1	1,7	1,5	1,7	1,7	1,1	1,2	1,6	1,6
Max. abs. odchylek		7	5,9	5,1	5,7	4,8	3,3	4,7	4,6	2,2	2,6	4,8	4,9
Průměrná relativní odchylka		24,00%	20,50%	20,30%	21,10%	24,30%	24,20%	25,40%	25,70%	20,30%	23,40%	23,00%	24,20%
Maximální relativní odchylka		60,50%	58,90%	62,40%	61,50%	69,70%	54,80%	51,20%	48,40%	57,40%	84,10%	73,20%	67,40%
Rozdíl mezi lídry		3,2	0,6	1,8	2,2	3,4	0,4	1,1	1	1,6	0,6	3,1	3,5
Chyby v kvalifikaci stran		0	0	0	0	1	0	0	0	0	1	0	0
Kategorie kvality predikce													
Součet čtverců odchylek		5	4	3	4	3	2	3	3	2	2	3	3
Průměrná abs. odchylka		3	3	3	3	3	2	3	3	2	2	3	3
Maximální abs. odchylka		5	4	4	4	3	3	3	3	2	2	3	3
Průměrná relativní odchylka		3	3	3	3	3	3	4	4	3	3	3	3
Maximální relativní odchylka		4	3	4	4	4	3	3	3	3	5	4	4
Rozdíl mezi lídry		4	1	2	3	4	1	2	2	2	1	4	4
Chyby v kvalifikaci stran		1	1	1	1	3	1	1	1	1	3	1	1
Celkový index kvality predikce		3,6	2,7	2,9	3,1	3,3	2,1	2,7	2,7	2,1	2,6	3	3

Pozn.: Pro barevné označení silně vychýlených stran jsou použity kritéria z kapitoly 2.1.

* Výsledek květnového modelu je ovlivněn uříznutím dat k začátku voleb (28. 5. 14:00), které se projevilo systematickým výpadkem oblastí (region x VMB), které jsou dotazovány na konci osobního sběru dat a segmentů populace, které jsou kontaktovány na konci měsíčního telefonického sběru. Výsledky tedy nelze považovat za plnohodnotný odhad voleb. **Posledním plnohodnotným a publikovatelným odhadem je model z plovoucích dat duben-květen.**

3.3 Optimalizace volebního modelu zařazováním nerozhodnutých voličů

Dotazový instrument MEDIANu používaný před volbami 2010 v CAPI i CATI sběru umožňoval respondentům odpovědět, že neví, jakou stranu budou volit. V tom případě byli dotázáni, aby jmenovali až 3 strany, které nejvíce zvažují. (viz Příloha 1)

Jak ukazuje statika popsaná v tabulce 3.3.1, v kombinovaném sběru dat z dubna a května 2010 je dohromady 34,2 % respondentů, kteří deklarovali nerozhodnost.¹¹⁷ To zhruba odpovídá údajům SC&C, že až 33,7 % voličů učinilo finální volební rozhodnutí v posledních 2 týdnech před volbami.

Tyto údaje však nelze vykládat tak (což se po volbách často dělo), že nerozhodnutí respondenti byly pro výzkum nutně nepredikovatelným faktorem a důvodem, proč volební výsledky nešlo odhadovat. (Válková 2010) Více než 2/3 nerozhodnutých voličů totiž bylo schopno specifikovat 1-3 strany, o které uvažují. Prostým odříznutím segment nerozhodnutých byly tyto informace ignorovány.

TAB 3.3.1 Zastoupení nerozhodnutých respondentů podle metodiky sběru

	CAPI 04+05 / 2010	CAPI + CATI 04+05 / 2010
Počet respondentů 18+	1633	3976
Počet avizujících účast u voleb	1100	2664
Počet rozhodnutých	923	1753
Počet nerozhodnutých	177	911
Počet nerozhodnutých neschopných specifikovat	28	269
% nerozhodnutých voličů	16,1%	34,2%
% nerozhodnutých voličů neschopných specifikovat	2,5%	10,1%

Politické spektrum před volbami 2006 bylo výrazně přehlednější, než před volbami 2010. Pokud nepočítáme KSČM, což je dlouhodobě strana s širokým pevným voličským jádrem a omezenou možností oslovit nerozhodnuté voliče (CVVM 2011c) (SC&C 2010b, s. 14), mohli se nerozhodnutí voliči v roce 2006 rozhodovat v malém a symetrickém politickém spektru s ODS/ČSSD na pravici/levici a Stranou zelených a KDU-ČSL v pozici hodnotově liberálního a hodnotově konzervativního politického středu. Menší strany byly ve velmi marginálním

¹¹⁷ Ve vzorku sebraném čistým osobním sběrem je „nerozhodnutých“ podstatně méně, než v telefonním sběru. To se dá zdůvodnit mimo jiné tím, že: a) Telefonní výzkumy mají obecně, a to nejen v ČR, tendenci koncentrovat více odpovědí „nevím“ (Bethlehem & Cobben & Schouten 2011). b) V ČR telefonní sběr preferencí tenduje k pravici (viz kapitola 3.1) a dostává se tedy do něj více lidí, kteří se rozhodují mezi jednou z pravicových stran ve specifickém volebním spektru května 2010 (viz níže); c) Tazatelé CAPI omnibusu měli na odpověď „nemohu si vzpomenout na jméno strany“ v první otevřené preferenční otázce reagovat předložením seznamu stran a na odpověď „nevím“ (ve smyslu, nevím koho budu volit), postupem k dotazování, jaké strany respondent zvažuje. V reálné dotazové situaci však mohlo být toto zaměňováno a respondenti nepřipouštění do multi-preferenčních otázek.

postavení a do rozhodování voličů výrazně nezasahovaly - dohromady získaly neparlamentní strany ve volbách 6 procent hlasů (viz kapitola 2.1.1).

Poměrně dobře tedy zafungovaly volební modely tvořené prostým „odříznutím“ nerozhodnutých voličů ze vzorku respondentů, u nichž předpokládáme účast u voleb, čímž se preference nerozhodnutých voličů symetricky rozdělí stranám úměrně jejich zisku v celkovém vzorku voličů.

Vzhledem ke zmíněným charakteristikám volebního spektra v roce 2006 lze předem předpokládat jen to, že tato konstrukce se výrazněji projeví nadhodnocením zisku KSČM, u níž díky vyššímu zastoupení tvrdého jádra voličů v celkovém elektorátu existuje předpoklad, že „svůj díl“ nerozhodnutých z jednoduchého přepočtu ve volbách nezíská. Komunisté v roce 2006 skutečně nakonec uspěli méně, než jim všech 5 hlavních předvolebních průzkumů předpovídalo (viz kapitola 2.1.1).¹¹⁸

Před volbami v roce 2010 se ale politické spektrum ČR značně zesložilo. Výrazně posílily menší strany – suma hlasů, které u voleb „propadly“, činila 18,8 %. (viz kapitola 2.1) Zejména se však rozpadla symetrie pravé části volebního spektra - pravice se rozložila na dva silné subjekty (ODS, TOP 09), k nimž se navíc částečným programovým zaměřením na vyrovnanou rozpočtovou politiku a proti zneužívání sociálních dávek blížily Věci veřejné. Detailní výstupy exit-pollu SC&C potvrzují, že očekávání voličů VV kladená na budoucí vládu se blížila voličům TOP09 a ODS, s důrazem na zastavení zadlužování. (SC&C 2010b, s. 19)

Jak ukazuje analýza struktury nerozhodnutých voličů na plovoucích datech z dubna-května 2010 (spojený vzorek CAPI+CATI), voliči se skutečně velmi často rozhodovali mezi jednou z budoucích koaličních stran. Viz TAB 3.2.1. Ta, jak namítne kritik, může být ovlivněna zmiňovaným faktem, že v subvzorku nerozhodnutých je více lidí z telefonního sběru dat, který obecně tenduje více k pravici. Je tedy dobré zaměřit se zejména na relativní ukazatel, který jsme nazvali *Riziko ztrátovosti*. Ten např. u koalice vyjadřuje poměr nerozhodnutých, kteří na prvních dvou místech zvažují na jednu z budoucích koaličních stran, ale také jinou stranu, v celkovém počtu nerozhodnutých, kteří na prvních či druhém místě (či obou) zvažují nějakou z koaličních stran.

Hodnoty ukazatel *Rizika ztrátovosti* demonstruje (TAB 3.3.2), že nerozhodní voliči se orientovali zhusta na rozhodování mezi jednotlivými koaličními stranami a koalici tak méně hrozila celková ztráta voliče, zatímco parlamentní levicová opozice (a zde majoritně ČSSD) trpěla větším rizikem, že se od ní voliči celkově odkloní.

¹¹⁸Pro ověření hypotézy, že tomu bylo díky nezohlednění skutečných úvah nerozhodnutých voličů, bychom ovšem potřebovali další data.

TAB 3.3.2 Struktura nerozhodnutých voličů (vzorek CAPI + CATI, duben-květen 2010)

			n
A	Počet respondentů avizujících účast u voleb		1436
B		celkový počet nerozhodnutých	512
C		neschopní specifikace	136
D		schopní specifikace	376
E	% schopných specifikace se stranou koalice 1. či 2. místě		64,1%
F	% schopných specifikace se stranou opozice 1. či 2. místě		36,4%
G	% schopných specifikace, kteří při výběru z prvních dvou preferencí nutně volí stranu koalice ¹¹⁹		41,5%
H	% schopných specifikace, kteří při výběru z prvních dvou preferencí nutně volí stranu opozice ¹²⁰		16,8%
I	RIZIKO ZTRÁTOVOSTI KOALICE = 1-(G/E)		43,2 %
J	RIZIKO ZTRÁTOVOSTI OPOZICE = 1-(H/F)		59,6 %

Pozn.: Absolutní a procentuální četnosti vyjádřené z nevážených dat.

Z perspektivy znalosti exit-pollů a ex-post volebních výzkumů se nejedná o překvapivé informace. Nicméně zásadní je zjištění, že tyto informace byly obsažené už v předvolebních datech a existovala tedy možnost je částečně předvídat.

3.3.1 Aplikace zařazování nerozhodnutých na data z kombinovaného sběru (CAPI + CATI)

Výzkumná otázka tedy zní, jak tedy s nerozhodnutými voliči nakládat, abychom zejména ve specifických volebních situacích, jakou byla ta z května 2010, zcela neignorovali fakt, že nerozhodnutí nejsou nepopsanými listy, které se symetricky rozdělí mezi strany úměrně jejich síle (viz výše), ale jejich „nerozhodnutost“ ve skutečnosti většinou znamená, že se rozhodují mezi 2 až 3 politickými subjekty, které emotivně či racionálně preferují.

Na vážených datech kombinovaného sběru (CAPI + CATI), které spojení s telefonním sběrem obsahují podstatně více respondentů, kteří přiznávají nerozhodnost, jsme analyzovali, jak se mění odhady výsledků voleb, pokud volební model netvoříme odříznutím nerozhodnutých, ale započítáváme do něj různým způsobem, které strany nerozhodnutý respondent zvažuje.

Celkově bylo vyzkoušeno 8 metod zařazování nerozhodnutých voličů. Extrémy byl na jedné straně předpoklad, že volič se se 100% pravděpodobností rozhodne pro první zmiňovanou stranu, a na druhé straně předpoklad, že se může s rovnoměrnou pravděpodobností rozhodnout pro jakoukoli ze tří zmiňovaných stran. Zohlednění nerozhodnutých proběhlo pomocí váženého zařazování jejich multi-preferencí. Například v modelu 7-2-1, který

¹¹⁹ Započítání respondentů, kteří řekli, že jsou nerozhodnutí, ale specifikovali jen jednu koaliční stranu (ODS, TOP09, VV), a nerozhodnutí respondentů, kteří specifikovali min. dvě strany a na obou prvních dvou místech uvedli stranu koalice.

¹²⁰ Započítání respondentů, kteří řekli, že jsou nerozhodnutí, ale specifikovali jen jednu opoziční stranu (ČSSD, KSČM), a nerozhodnutí respondentů, kteří specifikovali min. dvě strany a na obou prvních dvou místech uvedli stranu opozice.

předpokládá, že volič se rozhodne se 70% pravděpodobností pro první uvedenou stranu, se 20% pro druhou stranu a 10% pro třetí stranu mají tyto tři odpovědi váhu 0,7 – 0,2 – 0,1. Preference rozhodnutého voliče má vždy plnou váhu 1.¹²¹ Výsledné váhy jsou pak kombinací původních vah *Alternative_v8* (viz výše) a vah podle modelu zařazování nerozhodnutých.

Prvním a základním zjištěním z analýzy na CAPI+CATI vzorcích je, že zařazování multi-preferencí nerozhodnutých obecně pomáhá zlepšit predikční kvalitu volebního modelu. Všechny modely zařazování nerozhodnutých zmenšují při zachování shodného alternativního vážení odchylky volebního modelu od skutečných výsledků voleb, a to ve všech zkoumaných datových úsecích (duben 2010, duben-květen 2010, duben 2011). Viz Příloha 4.

Zároveň se potvrzuje, že nebylo chybou aplikovat alternativní vážení *Alternative_v8* i na spojený CAPI+CATI vzorek. Bez zařazení nerozhodnutých voličů sice toto vážení resultuje například v nadměrné posílení preferencí KSČM (viz tabulka 3.2.3 – duben 2010 a duben-květen 2010). To je však zařazením nerozhodnutých, kteří komunistickou stranu zvažují minimálně, redukováno a odhad zisku komunistů se vrací na realistických 11 – 13 %. Jinými slovy – naše predikce byla zřejmě správná / nevychýlená, pouze platila na omezenou populaci již rozhodnutých voličů, mezi kterými je voličů KSČM více, než mezi voliči celkově. Stejně jako v případě KSČM, tak i u KDU-ČSL všechny modely se zařazováním multi-preference nerozhodnutých redukuje odhadovaný zisk na realistickou hladinu okolo 5 %, kterých nakonec lidovci ve volbách těsně nedosáhli.

Modely zařazování nerozhodnutých voličů se svojí úspěšností příliš neliší - v každé ze tří datových úseků (duben 2010, duben-květen 2010, květen 2010) se dá identifikovat 4 až 7 modelů, které mají z hlediska kvality predikce zcela srovnatelně kvalitní výsledky. Metoda je tedy poměrně stabilní bez ohledu na změnu parametrů. Viz Příloha 4.

Do Tabulky 3.3.3 jsme vybrali ty modely, které patří v daných fázích k nejúspěšnější, a srovnali takto optimalizované modely s modely z nevážených a standardně vážených dat, které nahrazování respondentů nepoužívaly.

¹²¹ Prakticky toho bylo dosaženo restrukturováním dat, aby z otázek v02 a v04 (viz Příloha 1) vznikla jedna preferenční proměnná. Jejím případům byla poté přidělena váha podle daného modelu.

TAB 3.3.3 Srovnání modelů z kombinované sběru (CAPI+CATI) s rozřazováním multi-preferencí nerozhodnutých X standardní modely

Fáze		CAPI + CATI duben 2010			CAPI + CATI duben-květen 2010			CAPI + CATI květen 2010*		
Vážení		Bez vážení	Socio1	Alternative_V8	Bez vážení	Socio1	Alternative_V8	Bez vážení	Socio1	Alternative_V8
Využití nerozhodnutých		vyřazení	vyřazení	metoda 7-2-1	vyřazení	vyřazení	metoda 7-2-1	vyřazení	vyřazení	metoda 10-0-0
ČSSD	22,1	27,8	26,5	25,7	24,9	23,6	22,4	21,5	20,4	20,6
ODS	20,2	22,7	24,1	22,0	19,6	21,3	20,1	18,0	17,9	17,9
TOP09	16,7	9,7	10,8	12,9	11,9	13,4	15,0	16,0	16,8	17,3
KSČM	11,3	12,6	11,8	11,4	12,4	12,1	12,9	11,7	11,8	11,7
VV	10,9	9,0	9,2	9,8	10,4	9,8	9,9	11,5	10,6	11,8
KDU-ČSL	4,4	6,2	5,7	4,7	7,5	6,7	5,3	6,4	7,0	4,8
SPOZ	4,3	4,7	4,6	4,7	3,8	3,2	3,9	5,3	5,1	5,2
Suverenita	3,7	1,5	1,5	1,3	1,8	1,8	1,8	2,3	2,2	2,1
SZ	2,4	2,5	2,4	3,1	3,6	3,7	3,6	3,8	4,4	4,3
Ostatní	4,0	3,3	3,4	4,3	4,1	4,3	5,1	3,5	3,8	4,3
Hodnocení chybovosti										
Suma čtverců odchylek		102,5	78,5	38,9	47,3	28,1	14,0	15,3	22,3	16,0
Prům. abs. odchylka		2,4	2,1	1,5	1,7	1,5	1,0	1,1	1,2	1,1
Maximální abs. odchylka		7,0	5,9	3,9	4,8	3,3	1,9	2,2	2,6	2,4
Prům. relativní odch.		23,8%	20,3%	17,6%	24,1%	24,1%	19,6%	20,0%	23,0%	19,0%
Max. relativní odch.		60,2%	58,5%	63,5%	70,1%	52,4%	50,9%	54,8%	81,1%	76,3%
Rozdíl mezi lídry		3,2	0,6	1,8	3,4	0,4	0,4	1,6	0,6	0,9
Chyby v kvalifikaci stran		0	0	0	1	0	0	0	0	0
Kategorie kvality predikce										
Součet čtverců odchylek		5	4	2	3	2	2	2	2	2
Průměrná abs. odchylka		3	3	2	3	2	2	2	2	2
Maximální abs. odchylka		5	4	3	3	3	2	2	2	2
Průměrná relativní odchylka		3	3	3	3	3	3	3	3	3
Maximální relativní odchylka		4	3	4	4	3	3	3	5	4
Rozdíl mezi lídry		4	1	2	4	1	1	2	1	1
Chyby v kvalifikaci stran		1	1	1	3	1	1	1	3	1
Celkový index kvality		3,6	2,7	2,4	3,3	2,1	2,0	2,1	2,6	2,1

* Výsledek květnového modelu je ovlivněn uříznutím dat k začátku voleb (28. 5. 14:00), které se projevilo systematickým výpadkem oblastí (region x VMB), které jsou dotazovány na konci osobního sběru dat a segmentů populace, které jsou kontaktovány na konci měsíčního telefonického sběru. Výsledky tedy nelze považovat za plnohodnotný odhad voleb. **Posledním plnohodnotným a publikovatelným odhadem je model z plovoucích dat duben-květen.**

Platí přitom, že čím více se sběr blížil termínu voleb, tím vhodnější je vsázet na modely, které dávají velkou váhu první preferované straně.¹²² (Viz Příloha 4) To odpovídá očekávání, že s přibližováním voleb se preference voličů profilují a tedy pohybují směrem k nejvážněji zvažované straně.

Je zajímavé srovnat výsledky posledního datově plnohodnotného volebního modelu (TAB 3.3.3 duben-květen 2010), na nějž bylo aplikováno alternativní vážení a zařazení odpovědí nerozhodnutých v poměru 7-2-1, s výsledky volebních modelu agentur, které ukončily poslední publikovaný sběr zhruba ve stejné době (okolo 15.-17. 5. 2010) a s výsledkem exit-pollu SC&C. Srovnej s kapitolou 2.1.

Ukazuje se, že takto optimalizovaný volební model,¹²³ se úspěšností ve všech sledovaných charakteristikách blíží prvnímu exit-pollu, který agentura SC&C publikovala v sobotu 29.5. ve 14:00 (viz tabulka 2.1.7). V odhadu výsledku dvou hlavních stran (ODS, ČSSD) a jejich rozestupu je odhad dokonce přesnější. Celkový index predikční kvality optimalizovaného modelu je 2, celkový index predikční kvality prvního exit-pollu SC&C je 1,9.

Ve srovnání s posledními a to i nejkvalitnější predikujícími modely (poslední výzkum společnosti Factum) je většině ukazatelů kvalita predikce výrazně vyšší a zhruba se shoduje jen u maximální relativní odchylky, která je daná zhruba totožným podceněním strany Suverenita (viz výše). Této chybě se dalo (na rozdíl např. od celkového podcenění pravice) výše uvedenými metodami těžko zamezit. Exit-poll SC&C ukázal, že až 2/3 voličů Suverenity se rozhodlo dát jí hlas v posledních 14 dnech před volbami a celých 27 % se ji rozhodlo volit až těsně před volbami (SC&C 2010b, s. 14). V našich vzorcích existuje i po zařazení nerozhodnutých část respondentů, kteří účast u voleb plánovali, ale vůbec nebyli schopni specifikovat, o volbě jaké konkrétní strany uvažují.¹²⁴ Jedná se právě o typ voličů bez větší afinity k politickým stranám, kteří navíc v dlouhodobějším předstihu neprovádějí racionální volbu strany na základě volebního programu (viz typy volebního chování v kapitole 2.1). Platí-li, že právě tento segment voličů rozhoduje nejsilněji na základě poměrně krátkodobých emocí a aktuálních sympatií ke konkrétním politickým představitelům, není překvapivé, že nadstandardní podíl jeho hlasů dostala Suverenita s výraznou mediální osobností Jany

¹²² Pro srovnání v Tabulce 3.1.2 jsme vybrali pro fáze duben a duben-květen 2010 model zařazení 7-2-1 a pro květen 2010 model zařazení 10 – 0 – 0, kde plně zařazujeme jen první zvažovanou stranu. Vždy ale platí, že srovnatelně kvalitní jsou i modely okolní. Systém je tedy celkově velmi stabilní – nejedná se o statistickou „alchymii“, při níž by různým typem zařazení nerozhodnutých vznikaly zcela jiné výsledky. Viz Příloha 4.

¹²³ Díky časové blízkosti k volbám a přítomnosti pokročilejších matematických optimalizačních modelů by se zřejmě označit za volební predikci. To však vždy závisí na terminologické politice agentury.

¹²⁴ V květnových CAPI+CATI datech (do voleb) to je 4,3 % respondentů z těch, kteří avizují účast u voleb (52 z 1211 respondentů avizujících účast). V datech CAPI+CATI z plovoucího období květen-duben 2010 to je 9,5 % (136 respondentů ze 1436 avizujících účast).

Bobošíkové.¹²⁵ Ta má jinak díky heterogennímu programu a nárazovému politickému působení¹²⁶ omezené možnosti vytvářet si stranickou afinitu a oslovovat voliče volící racionální volbou, což ji „penalizuje“ ve volebních modelech.

Jestliže u Věcí veřejných, TOP09 a celkově pravice lze konstatovat, že její posílení šlo díky struktuře postojů v datech volebních průzkumů predikovat, u voličů Suverenity to díky dominanci ad-hoc volebního chování neplatí.

3.3.1 Aplikace zařazování nerozhodnutých na data z osobního (CAPI) sběr

Stejně modely zařazování multi-preferencí nerozhodnutých jako v případě dat z kombinovaného sběru (CAPI + CATI) byly aplikovány na volební modely vycházející z čistě z osobního sběru (CAPI). Nešlo zde očekávat tak významný vliv této procedury, protože v datech z osobního sběru je významně méně lidí, kteří přiznávají, že nejsou rozhodnutí, koho volit.

Oproti prosté aplikaci alternativního vážení¹²⁷ bez zařazování nerozhodnutých modely se zařazováním vedou k mírnému posílení celkových preferencí budoucích koaličních stran a mírnému omezení nadhodnocené podpory KDU-ČSL a KSČM. (viz Příloha 4) Vliv je tedy obdobný jako u aplikace zařazování u dat z kombinovaného sběru. Predikční kvalitu naopak mírně zhoršuje posílení Strany zelených multi-preferencí zařazením nerozhodnutých.

Rizikem váženého zařazování nerozhodnutých je zejména stav, kdy: a) ve volebním modelu je díky náhodné složce výběrové chyby nadreprezentována strana s nejistou účastí ve Sněmovně; b) v těsně předvolebním období dojde k odlivu voličů, kteří se o této straně rozhodovali (projevují se tak při váženém zařazení multi-preferencí), k alternativám.¹²⁸ Riziku chyby tohoto typu lze předejít např. zohledněním, nakolik je pro nerozhodnuté voliče důležité riziko nepřekročení 5 % klauzule.¹²⁹

¹²⁵ Její jméno bylo před volbami v roce 2010 i v názvu strany na volebních lístcích.

¹²⁶ Taktikou Suverenity je předvolební ad-hoc slučování a kooperace s dalšími menšími stranami, které rozšíří její heterogenní a k politické scéně kritické programové zaměření a posílí ji o lokálně známé osobnosti - například s ČSNS v krajských a senátních volbách mezi lety 2005-2008, se *Stranou za důstojný život* a *Stranou zdravého rozumu* před volbami do PSP ČR či se stranou Moravistů pro krajské volby v roce 2012 (Mediafax 2010). Jinak má tato strana velmi omezený počet 86 komunálních zastupitelů a (Stran zelených 323) a v mezivolebních obdobích redukuje svoje mediální nadregionální aktivity na akce Jany Bobošíkové zaměřené proti projevům evropské integrace (viz www.suverenita.cz).

¹²⁷ Na květnových datech z osobního sběru nebyly ve vážení použity sdružené kategorie VMB x region NUTS2. Omezení dat do voleb (28.5. 14:00) totiž zcela vyřazuje některé kombinace VMB x region a vážení se tak nemůže přiblížit teoretickým četnostem.

¹²⁸ Což byl případ odlivu voličů od Strany zelených k TOP09, na který ukazuje i exit-poll SC&C (2010b).

¹²⁹ Např. podle odpovědi na důležitost rizika nepřekročení 5% pro volební rozhodnutí přeradit zvažované strany s pravděpodobností nepřekročení 5% klauzule na zadní pozice zvažovaných stran.

TAB 3.3.4 Srovnání modelů z osobního sběru (CAPI) s rozřazováním multi-preferencí nerozhodnutých X standardní modely

Fáze		CAPI duben 2010			CAPI duben-květen 2010			CAPI květen 2010*		
Vážení		bez vážení	Socio1	Alternative V8	bez vážení	Socio1	Alternative V8	bez vážení	Socio1	Alternative V8
Využití nerozhodnutých		vyřazení	vyřazení	metoda 7-2-1	vyřazení	vyřazení	metoda 7-2-1	vyřazení	vyřazení	metoda 7-2-1
ČSSD	22,10	30,0	26,2	26,5	26,6	23,7	23,7	23,2	21,5	20,8
ODS	20,22	15,3	19,0	20,1	13,0	15,3	17,6	12,8	12,0	14,5
TOP09	16,71	8,0	10,7	11,8	9,5	13,9	14,4	14,3	17,9	19,2
KSČM	11,27	15,1	13,3	10,9	13,8	11,9	12,3	12,1	12,4	12,5
VV	10,88	7,8	7,6	8,3	8,4	7,4	7,9	11,9	10,9	11,6
KDU-ČSL	4,39	8,6	7,5	5,9	11,0	9,0	5,9	8,2	6,0	4,7
SPOZ	4,33	7,3	6,8	6,2	6,3	5,8	4,7	6,8	8,0	5,2
Suverenita	3,67	2,2	3,1	2,8	2,4	3,0	2,7	3,1	2,9	2,6
SZ	2,44	3,7	3,5	4,1	5,2	5,3	5,1	4,4	3,7	5,9
Ostatní	3,99	2,0	2,2	3,2	3,9	4,8	5,7	3,1	4,8	3,0
Hodnocení chybovosti										
Suma čtverců odchylek		220,3	89,9	60,1	194,0	79,9	37,9	88,9	89,7	58,0
Prům. abs. odchylka		4,04	2,57	1,90	3,66	2,38	1,77	2,24	1,93	1,83
Maximální abs. odchylka		8,67	6,01	4,90	7,26	4,92	2,94	7,39	8,24	5,76
Prům. relativní odch.		48,3%	34,3%	26,6%	49,0%	37,9%	29,2%	33,1%	27,6%	29,0%
Max. relativní odch.		96,5%	70,7%	69,5%	150,9%	117,6%	110,1%	87,5%	83,8%	141,0%
Rozdíl mezi lídry		12,8	5,3	4,5	11,7	6,5	4,2	8,5	7,6	4,4
Chyby v kvalifikaci stran		2	1	0	1	1	0	1	1	0
Kategorie kvality predikce										
Součet čtverců odchylek		7	4	3	7	4	2	4	4	3
Průměrná abs. odchylka		5	4	3	5	3	3	3	3	3
Maximální abs. odchylka		5	4	3	5	3	2	5	5	4
Průměrná relativní odchylka		6	4	4	6	5	4	4	4	4
Maximální relativní odchylka		5	4	4	7	6	6	5	5	7
Rozdíl mezi lídry		7	6	5	7	7	5	7	7	5
Chyby v kvalifikaci stran		5	3	1	3	3	1	3	3	1
Celkový index kvality		5,7	4,1	3,3	5,7	4,4	3,3	4,4	4,4	3,9

* Výsledek květnového modelu je ovlivněn uříznutím dat k začátku voleb (28. 5. 14:00), které se projevilo systematickým výpadkem oblastí (region x VMB), které jsou dotazovány na konci osobního sběru dat a segmentů populace, které jsou kontaktovány na konci měsíčního telefonického sběru. Výsledky tedy nelze považovat za plnohodnotný odhad voleb. **Posledním plnohodnotným a publikovatelným odhadem je model z plovoucích dat duben-květen.**

Při srovnání finálního plnohodnotného volebního modelu z dubna-května 2010 (alternativní vážení, metoda 7-2-1) se standardním modelem z tohoto období (standardní sociodemografické vážení, vyřazení nerozhodnutých), však aplikace námi testovaných optimalizačních modelů výrazně zlepšuje (viz TAB 3.3.4), a to zejména ve statistických ukazatelích měřících celkové vychýlení modelu (pokles sumy čtverců odchylek ze 79,9 na 37,9, pokles průměrné abs. odchylky z 2,38 na 1,77).

Ve srovnání s ostatními agenturami (viz tabulka 2.1.7) se původně vychýlený model celkovou kvalitou predikce dostává na úroveň nejlepších odhadů ze stejné doby (Factum Ingenio).¹³⁰

3.3.2 Výběr nejlepšího modelu zařazování a alternativa „donucení“ k single-response

Při výběru nejvhodnějšího modelu zařazování nerozhodnutých respondentů je vhodné zvážit:

a) Jak časově vzdáleny budou volby od posledního publikovaného výzkumu. Modely silněji zahrnující druhou a třetí zvažovanou stranu výrazněji zlepšují predikční kvalitu modelů u časově vzdálenějších výzkumů (v době méně vyprofilovaných preferencí). Modely s velkým důrazem na první stranu jsou výhodnější v těsné blízkosti voleb (viz výše) – tedy např. u předvolebních bleskových ad-hoc výzkumů pro média, nebo u kontinuálních výzkumů, s kratším trváním fáze sběru, která končí těsně před zákazem publikace.

b) Další faktory, kterými bude zařazování ovlivňovat výsledky – například, jak se projeví do oscilace modelu v čase a tedy zvěrohodnění výsledků dané agentury. Dá se přitom očekávat redukce oscilace, protože zařazováním nerozhodnutých respondentů zvětšujeme efektivní velikost vzorku (viz kapitola 3.2.). Zahraniční studie (Walsh & Dolphin & DiNardo 2009) také ukazují, že dynamicky klesající a rostoucí podpora stran je způsobena tím, že výzkumné agentury nerespektují, že elektorát strany A se při poklesu podpory přesouvá právě do sektoru pevně nerozhodnutých voličů, kteří by sice ve volbách stranu A s určitou pravděpodobností stejně volili, ale při ignorování nerozhodnutých se projeví jako kompletní výpadek preferencí.

Zajímavou otázkou je, proč jsou výsledky českých volebních průzkumů tak často v rozporu s realitou v tom, kolik identifikují nerozhodnutých voličů, a jaké to má důsledky. Podle posledního předvolebního průzkumu CVVM z května 2010 bylo mezi respondenty, kteří nevylučovali účast u voleb, 10,9 % nerozhodnutých. (CVVM 2010a)¹³¹ STEM má v posledních stranických preferencích 7,8 % nerozhodnutých, což činí 8,2 % z respondentů,

¹³⁰ Dostí se však liší v distribuci slabých a silných stránek predikce – srovnej TAB 3.3.2 a 2.1.7.

¹³¹ Ve stranických preferencích bylo odpovědí „nevím“ 8,5 %. Při přepočtu na respondenty nevylučující účast $8,9/(1-0,22) = 10,9$ %.

kterí neuvedli, že by nevolili „žádnou stranu“. (STEM 2010a) Nelze však určit jaké je procento nerozhodnutých mezi respondenty, u kterých agentura počítá s účastí u voleb.¹³² Za předpokladu, že lidé plánující účast u voleb mají preference vyprofilovanější než nevoliči, by však toto procento nemělo být vyšší. MEDIAN měl v posledním předvolebním průzkumu (z dubna 2010) při tehdy použitém standardním sociodemografickém vážení ve vzorku voličů 14 % odpovědí „nevím“. (Median 2010a) Factum stranické preference neuvádělo. (Factum 2010a, 2010b)

Bez ohledu na komplikace s přepočtem pro vzorek voličů (CVVM, STEM) je zjevné, že zastoupení lidí, kteří v průzkumech přiznávali, že „nevím“, bylo nižší než 33,7 % voličů, kteří v exit-pollu SC&C uvedli, že se rozhodli až v posledních 14 dnech před volbami či ve dnu volby. (SC&C 2010b, s.14)

Podreprezentování „nerozhodnutých“ může být způsobeno nezamýšleným důsledkem metodologie a formulace otázek či přímou strategií agentury: *„Tazatelé STEM jsou školeni tak, aby se s touto odpovědí snadno nespokojovali, zatímco tazatelé TNS Factum a CVVM se snadněji s odpovědí "nevím" smíří.“* (STEM 2011b)

Výsledek v obou případech je, že výzkumníci počítají za rozhodnuté voliče i část těch, kteří ve skutečnosti nijak pevně rozhodnutí nejsou a byli pouze „donuceni“ single-response charakterem otázky vybrat jednu stranu. Tento postup je de facto podobný našemu zařazování nerozhodnutých voličů metodou 10-0-0 (viz výše), kdy se první zmíněná zvažovaná strana zařadí do souboru s vahou 1 a k dalším se vůbec nepřihlíží.

Jak naznačila naše analýza na vzorcích z kombinovaného sběru (CATI + CAPI) i osobního sběru (CAPI) dat, tento postup je vhodný situaci, kdy reálné volby jsou velmi blízko, preference nerozhodnutých jsou už vyprofilované a často ve výsledku volí skutečně první zmíněnou/zvažovanou stranu. (viz výše) Při větší vzdálenosti od voleb (například i na modelu MEDIANu z plovoucího období duben-květen 2010 se sběrem dat mezi 17. 4. a 16. 5. 2010), kdy jsou preference méně vyprofilované, ale může tento postup vést k méně optimálním výsledkům.

Matematici Walsh, Dolphinová a DiNardo ve své podnětné práci z roku 2009 dokonce navrhnou, aby se agentury od začátku neptaly otázkou typu *„Jakou stranu byste volila(a), kdyby se volby do Poslanecké sněmovny konaly příští týden,“* která respondenta nutí do vyjádření jedné odpovědi. Ale aby nechaly respondenta přiřadit pravděpodobnosti, s jakou

132 Z formulace STEM ve zprávách z roku 2010 lze vydedukovat, že podobně jako Factum pracuje s nějakým způsobem váženého zařazování respondentů podle toho, jak odhadují svoji volební účast. (STEM 2010a) Procento nerozhodnutých ve vzorku voličů tedy opět spočítat nelze.

budou dané subjekty volit. To vede k odlišení přesvědčeného podpůrce G.W. Bushe (100 %) a jeho nejistého voliče (Bush – 60 %, Kerry – 40 %). V matematických modelech ukazují, k jakým odchylkám odhadů od skutečného stavu může „nucení“ do single-response odpovědím vést. (Walsh & Dolphin & DiNardo 2009)¹³³

3.4 Optimalizace prognózou odhadu účasti respondenta u voleb

Dotazníky volebních výzkumů všech hlavních agentur obsahují otázku, zda se respondent by se respondent zúčastnil voleb, pokud by se konaly v blízké budoucnosti, ačkoli přesná formulace této otázky se liší.¹³⁴ Všechny agentury přitom volební účast odhadly poměrně správně: CVVM 61 %, STEM 62 %, MEDIAN 64 %, Factum 66,8 %. Vzhledem k reálné účasti 62,6 % (ČSU 2010b) se dá říci, že odhad byl významněji¹³⁵ vychýlení jen v případě společnosti Factum.

Zástupci některých agentur se shodují, že pravděpodobnost, zda se respondent avizující účast u voleb hlasování skutečně zúčastní, je pro relevanci odhadů jedním z nejdůležitějších prvků (Čro 2010). Postup agentur byl v roce 2010 v této otázce rozdílný:

Společnosti MEDIAN (2010a) důvěřovala voličům, že jejich deklarativní odpověď má v otázce volební účasti v daných podmínkách nejlepší vypovídací hodnotu. Volební model tak konstruovala pouze z respondentů avizujících účast (určitě ano + spíše ano). CVVM (2010a) do volebního modelu zahrnuje všechny respondenty, kteří nevyloučili účast u voleb. Zařazuje tedy i respondenty, kteří odpověděli, že se „spíše nezúčastní“ a v následné otázce byli schopni uvést stranu, kterou by v případě účasti volili. Ze srovnání stranických preferencí a volebního modelu CVVM přitom vyplývá, že model vzniká prostým neváženým přepočtem (respondenti avizující „spíše neúčast“ mají v modelu stejnou váhu jako respondenti avizující účast). CVVM se tím dostává do trochu paradoxní situace, kdy odhaduje volební účast na 61

¹³³ Taková metoda je zřejmě ve volbách do PSP ČR nerealizovatelná, protože přiřadit pravděpodobnosti volby v prezidentských volbách Bush-Kerry-Nader, zvláště když uvážíme značnou politickou polarizaci americké společnosti, je jednodušší, než procenta rozdělit mezi 9 subjektů, které se mohou dostat do Sněmovny či další menší strany.

¹³⁴ CVVM, Factum a STEM se ptaly na účast, pokud by se volby konaly příští týden (CVVM 2010a) (STEM 2010a) (Factum 2010b), MEDIAN pokud by se konaly zítra (MEDIAN 2010a). Median a CVVM používají 4 možnosti odpovědi (určitě ne, spíše ne, spíše ano, určitě ano). Factum k nim navíc přidává možnost „nevím“. STEM přesné znění dotazníku nepublikoval, ale v prezentaci dělí avizovanou pravděpodobnost účasti na tři typy – „Voleb se chce podle svého vyjádření zúčastnit 62 % občanů, 14 % svou účast jasně odmítá, nerozhodnuto je 24 % obyvatel ČR.“ (Stem 2010a, s. 2)

¹³⁵ Jako referenci zde můžeme brát to, že statistická odchylka činí na kvantilu 67 % zhruba u vzorku n=1000 zhruba +/- 2,7 procentního bodu.

%, ale volební model¹³⁶ kalkuluje s předpokladem, že „hypotetických voleb“ by se zúčastnilo minimálně okolo 70 % jeho vzorku.¹³⁷

Z vyjádření Jana Herzmana z agentury Factum (Chábová 2011, s. 42) vyplývá, že agentura zařazuje do modelu respondenty váženě podle toho, jak avizují účast, a přitom „(...) *ti, kdo odpověděli, že přijdou určitě, mají méně než 100 % váhy a ti, kteří odpověděli, že určitě nepřijdou, mají více než 0 % váhy.*“ (tamtéž) Na podobném principu váženého zařazování postupuje STEM (Chábová 2011, s. 42) (STEM 2010a). U ani jedné z těchto agentur ale není z veřejných výstupů zřejmé, na základě jakých expertních znalostí a analýz váhy určuje a zda stejnou váhu dostávají všichni voliči avizující jistou pravděpodobnost účasti (např. odpověď „*spíše ne*“), či je tato váha specifická podle preferované strany a dalších charakteristik respondenta.

Vážený postup zařazování čistě podle avizované pravděpodobnosti účasti na první pohled připomíná metodu použitou v kapitole 3.3 – vážené zařazování multi-preferencí nerozhodnutých. Ze statistického hlediska v obou zvyšujeme efektivní velikost vzorku. V případě zařazování nerozhodnutých prostě tím, že při přidáváme aditivní informace o nerozhodnutých respondentech, které jsme předtím vyřazovali. V případě zařazování podle volební účasti tím, že proti nulovému stavu zrovnoměrňujeme distribuci vah. Postup MEDIANu totiž lze označit za krajní případ zařazování podle avizované účasti u voleb s vahami nabývajících pro čtyři typy odpovědi hodnoty (1 - 1 - 0 - 0), zatímco při váženém zařazování, které popisuje Herzmann (výše), mohou nabývat hodnot například (0,9 - 0,6 - 0;4 - 0;1). Suma vah je v obou případech stejná, ale efektivní velikost vzorku se mění:

Vycházejme z rovnice, kterou pro odhad efektivní velikosti vzorku využívají Ganninger (2006) a další, kde¹³⁸:

$$n_{eff} = \frac{n_{def}}{a}$$

$$def = n * \frac{\sum_i^n w_i^2}{(\sum_i^n w_i)^2}$$

136 Který podle definice CVVM: „ (...) simuluje reálné rozvržení voličských hlasů v hypotetických volbách.“ (CVVM 2010a, s. 3) Jedná se tak vlastně o hrubý volební odhad k aktuálnímu časovému okamžiku.

137 V posledním volebním modelu CVVM je 700 respondentů. Celkově přitom bylo osloveno 1016 respondentů nad 15 let, tedy při předpokládaném rovnoměrném zastoupení lidí ve věku 15-17 let je vzorek 18+ menší než n=1000 (CVVM ve zprávě velikost vzorku 18+ neuvádí). Tím se dostáváme k číslu minimálně 70 %.

¹³⁸ V rovnici je n počet respondentů, n_{eff} efektivní velikost vzorku a w_i jsou váhy jednotlivých případů.

Mějme vzorek 1000 respondentů, kde 33 % říká, že se voleb rozhodně zúčastní, 28 % spíše zúčastní, 10 % spíše nezúčastní a 22 % rozhodně nezúčastní, což je poměr v posledním předvolebním výzkumu CVVM (2010a). Pokud těmto typům odpovědí přiřadíme váhy zmíněné výše, dostáváme s v prvním případě (váhy 1 - 1 - 0 - 0) efektivní velikost vzorku $n_{eff} = 610$, v druhém případě (váhy 0,9 - 0,6 - 0;4 - 0;1) efektivní velikost vzorku $n_{eff} = 719$.

Čistě statistického hlediska je tedy zařazovat respondenty váženě podle avizované pravděpodobnosti účasti výhodné, protože oproti prostému postupu CVVM a MEDIANu vždy (ačkoli rozdílnou měrou podle struktury vah) zvyšuje efektivní velikost vzorku a tím například omezuje náhodnou složku rozptylu odhadů. Zde ale podobnost s váženým zařazováním nerozhodných končí. Rozdíl je jednak ve způsobu a intenzitě s jakou zvyšujeme efektivní velikost vzorku:

V zařazování nerozhodnutých pouze zařazujeme aditivní informace/respondenty bez redukce absolutní informační hodnoty vzorku rozhodnutých voličů (váha zůstává 1). Rozšiřujeme vzorek, z kterého kalkulujeme preference, tak aby pokrýval pokud možno celou část vzorku voličů.

Při zařazování lidí avizujících neúčast (spíše či *určitě*) naopak, chceme-li zachovat odhad volební účasti odpovídající výsledkům z deklarativní odpovědi, musíme váhami snižovat hodnotu informací, kterou jsme získali od respondentů avizujících účast (k uchování průměru vah 1).

Jinými slovy, při zařazování nerozhodnutých nejdeme expertním úsudkem proti informační hodnotě deklarativních odpovědí a nepředpokládáme, že rozhodnutí voliči se „nakonec rozhodnou jinak“. Naopak, u zařazování podle avizované účasti jdeme expertním úsudkem proti informační hodnotě deklarativních odpovědí – předpokládáme, že část voličů avizujících účast k volbám nepůjde.

Kromě rozdílného charakteru vztahu expertního úsudku a deklarativních odpovědí je také rozdíl v síle těchto expertních úsudků. U zařazování preferencí nerozhodnutých jsou naše obecné předpoklady poměrně slabé.¹³⁹ Naopak u váženého zařazování respondentů podle avizované účasti u voleb musíme zvážit řadu navzájem souvisejících faktorů. Například:

- a) Stranická identifikace a věrnost voličů se u stran poměrně významně liší (CVVM 2011c). Zejména střední odpovědi k volební účasti („spíše ano“, „spíše ne“) mohou pro voliče jiných stran mít odlišný význam.

¹³⁹ Tedy, že: a) že alespoň část nerozhodnutých respondentů umí pravdivě popsat, mezi kterými stranami se rozhoduje; b) že respondent se pravděpodobně rozhodne pro jednu ze stran, kterou zvažuje, a proto jsou jeho preference lepším ukazatelem budoucí volby než předpoklad, že se nerozhodnutí rozloží mezi strany úměrně jejich síle.

- b) Existují naprosto rozdílné důvody neúčasti z voleb, které analyzoval například Lukáš Linek na základě valenční teorie chování (Linek 2010). Pokud například respondent ví, že se nebude moci zúčastnit kvůli okolnostem (vysoký věk, dlouhodobá nemoc, nepřítomnost v republice, atd.) či kvůli úplnému odmítání voleb (nikdy nehlasuje), nemá valný smysl jeho hypotetickou volbu do volebního modelu váženě zařazovat. U lidí avizujících neúčast z důvodů, které se mohou změnit (např. neví, pro kterou stranu by měl hlasovat), lze naopak účast části respondentů předpokládat. O důvodech volební neúčasti podrobněji viz Linek (2010).
- c) Podle komparativních zahraničních analýz ovlivňuje deklarativní odpovědi o účasti u volbách sociální desirabilita, která se však nelze jednoduše vyčíslit jedním koeficientem (např. 20 % z lidí avizujících „spíše účast“ k volbám nepřijde), protože úroveň odpovídání dle sociální desirability silně závisí na řadě konkrétních faktorů – síle stranické identifikace, vzdělání, míře občanské zodpovědnosti a informovanosti, přesvědčení občana, že individuální akce může přinést změnu (Karp 2005, s. 835). Tyto faktory jsou navíc různé v různých zemích a síla vlivu sociální desirability se mění i díky okolnostem – např. je vyšší u voleb s tradičně vyšší volební účastí. (tamtéž)

Regulérní konstrukce vážení, které vyjádří rozdílnou míru důvěry v to, zda se respondent zúčastní voleb, by tedy měla být založena na komplexní analýze, která bude v českém politickém prostředí analyzovat: vliv síly stranické identifikace na důvěryhodnost odpovědí o účasti ve volbách, vliv dalších faktorů ovlivňujících sociální desirabilitu, korelaci účasti ve volbách s účastí s volební účastí minulosti, intenzitu různých typů dobrovolné a vynucené neúčasti (Linek 2010) či závislost sociální desirability odpovědí na typu voleb, atd.

Model by musel přidělovat váhy určující pravděpodobnost účasti u voleb respondentům jednotlivě a volební průzkum by musel obsahovat dostatečný aparát pro popis respondentů, který přidělení váhy umožňuje. Vývoj takového aparátu samozřejmě není vyloučen, nicméně data společnosti MEDIAN přístupná pro tuto práci dostatečně podrobnou analýzu v této oblasti neumožňují.

Experimentováním na předvolebních datech bez teoretických základů a zahrnutí dostatečného množství faktorů ovlivňujících vztah mezi avizovanou a skutečnou účastí, bychom sice mohli dospět k závěru typu, že nejlepší volební predikci získáváme při zařazení respondentů podle avizované účasti přidělením vah 0,9 (určitě ano), 0,6 (spíše ano), 0,4 (spíše ne), 0,1 (určitě ne). Ve volbách 2014 by však při celkové změně výchozí politické situace, změně stranických identifikací, zájmu o volby, atd. model vůbec nemusel platit. Proto je podle názoru autora této práce za podmínek nedostačujících ke konstrukci skutečně

robustního modelu lepší zůstat u důvěry v deklarativní odpovědi respondentů ohledně volební účasti a nesnižovat jejich informační hodnotu nedostatečně relevantním expertním úsudkem.

3.5 Optimalizace modelu potlačením oscilace v čase

Jak bylo uvedeno již v kapitole 3.2.2, ukazatelem kvality volebního modelu nemůže být jen to, nakolik se v období těsně před volbami jeho odhady blíží výsledkům voleb. Z hlediska prestiže výzkumné agentury a omezení mediálních desinterpretací, které přisuzují přehnanou informativní roli náhodným výkyvům modelu, je důležitým ukazatelem kvality také míra oscilace výsledků volebního modelu v čase. Nelze totiž předpokládat, že z masových médií se vytratí důraz na zpravodajské hodnoty překvapivosti a konfliktnosti (Galtung & Rugeová 1968) (Shulz 2006) (Westerstahl & Johansson 1994), kvůli nimž je pro média zajímavé rámcovat výsledky volebních výzkumů jako „dostihový závod“ politických stran (Weimann 1990) plný zvratů a jako součást tohoto rámcování interpretovat i náhodné oscilace modelu jako reálné procesy na politické scéně.

V kapitole 3.2.2 jsme ukázali, že oscilaci modelu v čase lze snížit už tím, když místo standardního sociodemografického vážení aplikujeme alternativní vážení, které zahrnuje proměnné, jež s volebními preferencemi nejvíce souvisí (vážení *Alternative_8*). K dalšímu omezení oscilace by zřejmě došlo při zvolení konstantního typu zařazování nerozhodnutých voličů, ačkoli zde se dá očekávat zejména u osobního (CAPI) sběru díky menšímu zastoupení nerozhodnutých spíše minimální vliv.

Ve výsledku je ale vždy na posouzení publikující agentury, jakou míru oscilace modelu bude považovat za přijatelnou. Existují totiž metody, jak oscilaci modelu v čase snížit zohledněním předchozích trendů ve vývoji preferencí.

Možnostmi tzv. vyhlazování volebních dat se hojně zabývají například experti na volební výzkum, statistici i publicisté v USA v posledních 7 letech. Většinou se orientují na vyhlazování pomocí agregace výzkumů, a to různými metodami od prostého neváženého průměrování posledních výzkumů v dané geografické oblasti (www.realclearpolitics.com), přes mnohonásobně sofistikovanější metody, kteří používají statistici, jež jsou autory agregátorů *Pollster.com* či *FiveThirtyEight*. Například poslední jmenovaný používá vážené průměrování průzkumů, které bere v potaz aktuálnost výzkumu, velikosti vzorků i historickou úspěšnost agentury a data navíc vyhlazuje vlastním nulovým odhadem, který vychází z kombinace minulých výsledků voleb a demografické projekce v daném státě USA (FiveThirtyEight 2011). Spolehlivost výsledků těchto agregátorů je v dnešní době již taková,

že pokud si chce žurnalista či nezávislý pozorovatel udělat obrázek o vývoji politické scény v USA a jednotlivých státech, může jednotlivé agenturní výsledky s klidem ignorovat.¹⁴⁰

Jiní autoři (Laver 2005) se zabývají přímo možnostmi vyhlazovat časové řady výsledků datech jedné agentury. Účel je u všech těchto metod stejný – poskytnout důvěryhodné informace o trendech a aktuální podpoře stran, v nichž pozorovatele nebude rušit „šum“ způsobený nevyhnutelnou náhodnou oscilací modelu.

Mezi tzv. adaptivní metody používané v ekonomii, demografii a dalších vědách, které zkoumají časové řady a zabývají se predikcí, patří například metoda exponenciálního vyhlazování. Při exponenciálním vyhlazování prvního stupně je časová řada vyhlazena podle rekurentního vzorce (Hančlová & Tvrdý 2003, s. 21):

$$\hat{y}_t = (1 - \alpha) * y_t + \alpha * y_{t-1}$$

Přičemž koeficient α je hodnota mezi $<0,1>$, y_t je současné pozorování (volební model z aktuální fáze sběru dat) a y_{t-1} je minulé pozorování (volební model z minulé fáze sběru dat).

Pokud tedy například určíme koeficient $\alpha = 0,7$ ¹⁴¹ bude aktuální hodnota (publikovaný model po vyhlazení) počítán jako $0,3 * \text{aktuální model} + 0,7 * \text{minulý model}$. Díky rekurentnosti operace se tím do finální hodnoty promítá i model sebraný o dvě fáze napřed ($0,7 * 0,7$), atd. Pokud zvolíme hodnotu $\alpha = 0$, nebudeme vyhlazovat vůbec – minulá data se do publikovaných výsledků vůbec nezahrnou a budeme plně spoléhat na aktuální model.

Pro mírné vyhlazení volebního modelu lze zvolit například $\alpha = 0,25$. Publikovaný model tedy bude vždy ze 75 % zahrnovat model ze současných dat a z 25 % minulý publikovaný model (který z 25 % zahrnuje předminulý publikovaný model, atd.) Lze jednoduše spočítat, že takový model si sleduje historii dat 2-3 měsíce nazpět (příspěvek dalších dat je zcela zanedbatelný), přičemž klade majoritní důraz na aktuální data. Viz TAB 3.5.1.

3.5.1 Struktura dat v modelu publikovaném v prosinci 2010 (exp. vyhlazení $\alpha = 0,25$)

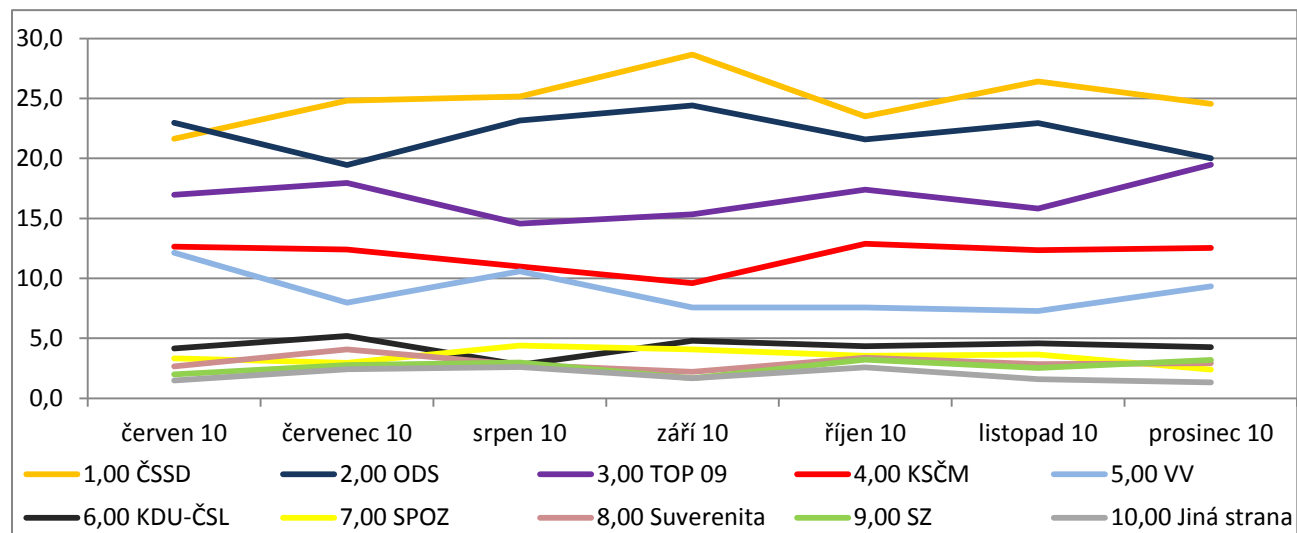
Model publikovaný v prosinci 2010	Složení z dat sebraných do prosince 2010				
	Prosinec 2010	Listopad 2010	Říjen 2010	Září 2010	Srpen 2010
100 %	75 %	18,8 %	4,7 %	1,2 %	0,3 %

¹⁴⁰ Statistik Nate Silver (autor agregátoru FiveThirtyEight) byl mimo jiné jeden z mála expertů v USA, kteří tipovali tak výrazný úspěch Ricka Santoruma v prezidentských primárkách v Iowě, 3.1.2011, které se konaly těsně před odevzdáním této práce. (FiveThirtyEight 2011)

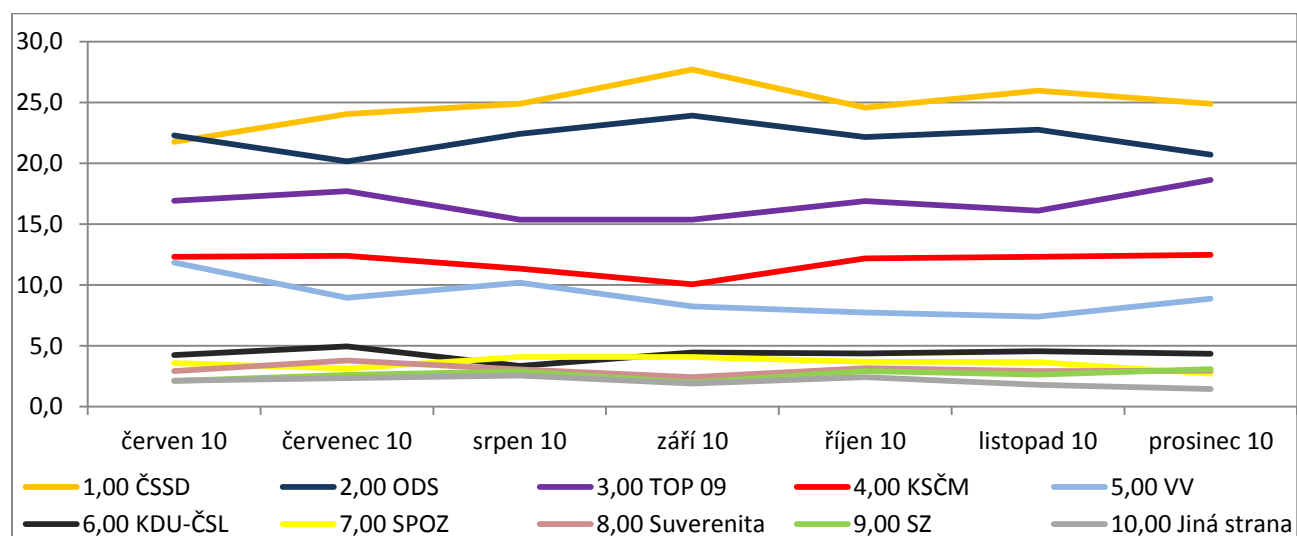
¹⁴¹ Což je minimální doporučená hladina vyhlazování v některých oblastech aplikace této metody, kde je cílem poměrně velké vyhlazení (Hančlová & Tvrdý, s. 18).

Výsledkem je to, že vyhlazování mírně zarovnává dočasné skoky ve vývoji modelu, při nichž jedna ze stran “posílí” na měsíc o 3-5 procentních bodů. Ponechává však celkové vývojové trendy modelu (ačkoli je může zpomalovat, jak si ukážeme později).

GRAF 3.5.1 Vývoj modelu (CAPI sběr – vážení *Alternative_v8*) – bez vyhlazení ($\alpha = 0$)



GRAF 3.5.2 Vývoj modelu (CAPI sběr – vážení *Alternative_v8*) – s vyhlazením ($\alpha = 0,25$)



Vyhlazení je mírné a na aktuálním zisku strany se více projevuje právě jen při výrazném meziměsíčním nárůstu či poklesu. (srovnej Graf 3.5.1 a 3.5.2) Například je zarovnán velký výkyv preference ČSSD mezi zářím a říjnem 2010 (pokles o 5,1 % v neváženém modelu, o 3,6 % ve váženém modelu). Celkové statistiky popisující oscilaci se snižují (viz TAB 3.5.2), vývoj modelu je důvěryhodnější a jednorázové skoky dané náhodnou složkou výběrové chyby neodvádějí tolik pozornost od dlouhodobějších trendů.

TAB 3.5.2 Statistický popis oscilace vyhlazeného a nevyhlazeného modelu

Typ modelu (vážení / vyhlazení)	Červen 2010 – Prosinec 2010			
	Průměrná abs. odchylka	Maximální. abs. odchylka	Průměrný čtverec odch.	Průměrná rel. odchylka
Alternative_v8 / bez vyhlazení	1,48	5,1	3,72	18,4 %
Alternative_v8 / vyhlazení $\alpha = 0,25$	1,06	3,6	2,06	11,9 %

Jako u většiny čistě statistických optimalizačních opatření není ani exponenciální vyhlazování prvního stupně bez rizika. Zpomaluje totiž i rozpoznání skutečných trendů modelem. Ukazuje to hypotetický příklad v tabulce 3.5.3 – pokud by 5 měsíců před volbami vznikla nová strana a stabilně posilovala o 4 procentní body měsíčně, odhaduje jí model na konci místo správných 20 % jen 18,7 %. Na správnou úroveň se dostává, až když se trend růstu preferencí zastaví, nebo preference strany začnou klesat.

TAB 3.5.3 Exponenciální vyhlazení strany, která za 5 měsíců posílí na 20 % ($\alpha=0,25$)

	N-5	N-4	N-3	N-2	N-1	N
Bez vyhlazení	0%	4,0%	8,0%	12,0%	16,0%	20,0%
Exp. vyhlazení 1. stupně	0%	3,0%	6,8%	10,7%	14,7%	18,7%

V těsně předvolebním období tedy může exponenciální vyhlazování kvalitu predikce pomoci i uškodit. Vyhladí skoky způsobené náhodnou či systematickou složkou výběrové chyby (např. nadreprezentování KDU-ČSL a SPOZ v posledním publikovaném modelu MEDIAN z osobního sběru s použitím sociodemografického vážení, podreprezentování ODS v květnovém modelu dané výpadkem některých kombinací VMB x region při uříznutí dat do voleb). Naopak zpomalí vnímání reálných trendů modelem – například výrazné posilování TOP09.

Tento problém však není neřešitelný. Za předpokladu, že strana nestoupá jen v poslední fázi sběru, lze použít tzv. exponenciální vyhlazování druhého stupně. To mimo hodnot strany v minulých fázích zahrne i trend mezi těmito minulými fázemi. Rozpozná tedy, jestli posílení v posledním měsíci je skok či pokračováním trendu posilování strany (více viz Hančlová & Tvrdý 2003, s. 22 – 28) Toto řešení s dostatečně intenzivním sběrem dat, který může trend v posledních 2 měsících před volbami popsat.

Jako další řešení lze navrhnout, aby výzkumná agentura v předvolebním sběru obohatila dotazník o otázky, které identifikují, zda nárůst dané strany je krátkodobou náhodnou či systematickou výchylnou, nebo skutečným trendem v posilování dané strany. Nabízí se například zařazení otevřených otázek, kdy se respondent pro danou stranu rozhodl a jaké

strany zvažoval předtím. Na základě analýz těchto odpovědí pro strany se silnou výchylnou oproti minulému měsíci a obecných trendů mezi nerozhodnutými lze rozhodnout, jaký typ vyhlazení (pokud nějaký) v poslední předvolební fázi aplikovat.

4. Závěr

4.1 Shrnutí

Cílem této práce bylo ukázat možnosti optimalizace volebních výzkumů v České republice. Výsledky analýz provedených na datech společnosti MEDIAN by se daly shrnout do několika tezí:

- 1) Předpoklad, že sledování základních sociodemografických údajů (pohlaví, věk, vzdělání, kraj, VMB), a to při definici kvótních znaků či při konstrukci post-stratifikačního vážení, je dostatečnou zárukou reprezentativity vhodné pro volební výzkum, není plně oprávněný. Volební preference souvisí více než s některými z těchto základních znaků se specifickými sociodemografickými atributy (např. příjem, pracovní zařazení) a atributy životního stylu a vystavení médiím (např. intenzita užívání internetu, preference volnočasových aktivit), které nejsou základními znaky plně nahraditelné.
- 2) Odlišné formy *non-response bias* u osobního sběru a telefonního sběru dat se mohou projevovat odlišnou strukturou vzorků v manifestních specifických sociodemografických (pracovní zařazení, příjem) či mediálních attributech (používání internetu) a potenciálně i v odlišné distribuci latentních faktorů,¹⁴² které souvisí s volebními preferencemi respondentů. Minimálně v případě předvolební situace v roce 2010 se přitom ukazuje, že použití kombinované metodiky sběru (CAPI + CATI) pomáhá k dosažení vyšší reprezentativity u klíčových faktorů ovlivňujících volební preference a ve výsledku i ke kvalitnější predikci výsledku voleb.
- 3) Alternativou či komplementem kombinované metodiky sběru dat může být použití alternativních vážení, které kromě zafixování základních sociodemografických faktorů váží data podle sdružených sociodemografických kategorií a specifických faktorů (intenzita používání internetu, příjem, pracovní zařazení) nejvíce souvisejících s volebními preferencemi. Využití zacíleného vážení v případě použití dalších

¹⁴² Tedy neověřitelně, protože chybí srovnatelné údaje ve výzkumech a opora pro popis populace.

optimalizací tvorby volebního modelu (zařazení nerozhodnutých - bod 4) vyplatí v případě CAPI sběru i CAPI+CATI sběru.

- 4) Rozhodování nerozhodnutých voličů před volbami 2010 bylo standardními výzkumnými instrumenty neodhadnutelné jen do jisté míry. Výzkum mohl zohledněním faktu, mezi kterými stranami se tito voliči dominantě rozhodují, předpokládat posílení pravicových stran ve volbách. Konkrétně v případě společnosti MEDIAN by vážené zahrnutí nerozhodnutých pomohlo k redukci nadhodnocení KDU-ČSL, které částečně vyplývalo z toho, že ignorování nerozhodnutých voličů strany s pevným voličským jádrem posiluje. Váženým zahrnutím multi-preferencí nerozhodnutých voličů, se predikční kvalita modelů zvyšuje, přičemž platí obecné pravidlo, že s blížícím se termínem voleb se zvyšuje výhodnost spoléhat na volbu první zmíněné zvažované strany.

Nebezpečím váženého zařazování nerozhodnutých je zejména to, že nezachycuje některé systematické vlivy na rozhodnutí voličů váhající mezi několika stranami v posledních 14 dnech. Některé z takových vlivů však existují v předvolebních situacích obecně (např. riziko nepřekročení 5% klauzule; volba zvažované strany, která nemá jisté vítězství, atd.) a pro účely poslední předvolební prognózy na ně lze zaměřit specifické dotazové instrumenty.

- 5) Matematické vyhlazovací metody, které jsou ekonomie a další vědní obory používají při sledování časových řad a predikcích zatížených „šumem“ (náhodnou složkou výběrové chyby), jako např. exponenciální vyhlazování jsou využitelné i při optimalizaci volebního modelu a jeho oscilace v čase. Je však úkolem pro další analýzu zvolit nejvhodnější metodu vyhlazování (stupeň exponenciálního vyhlazování, síla koeficientů), aby model dostatečně eliminoval oscilaci a zároveň výrazně nezpomaloval zachycování trendů.

Díky poměrně univerzální povaze těchto zjištění a podobné praxi řady výzkumných agentur před volbami 2010 ve sledovaných oblastech (viz kapitola 2.2), lze předpokládat, že dané závěry disponují jistou mírou zobecnitelnosti i mimo sledovaný datový zdroj (výzkum MEDIAN před volbami 2010).

Zároveň uvedená zjištění podle názoru autora této studie nabourávají představu, že „výzkumy byly dobře“ a volby se od nich odchýlily nepredikovatelným vývojem v posledních 14 dnech před volbami. Volební výzkum jistě nemůže nikdy plně překročit omezení vyplývající z náhodné složky výběrové chyby, z nejistoty v odhadu volební účasti respondentů, z nepredikovatelného chování politicky zcela neukotvených nerozhodnutých

voličů či z možnosti významných politických událostí posledních předvolebních dnů. Podle zjištění této práce však existují možnosti zkvalitnění predikce výsledku voleb v redukci systematického vychýlení předpovědí.

Řešené otázky samozřejmě nejsou kompletním výčtem zdrojů systematického vychýlení a jeho možných řešení – optimalizace lze v praxi každé agentury směřovat i do terénní praxe (kontrola tazatelů, zvyšování oblastí – geografické heterogenity sběru, atd.) i do praxe zpracování dat (např. využití statistických metod jako bootstrapping pro optimalizaci výstupů).

4.1 Diskuse

Dvě zásadní obecné námitky proti výsledkům této práce mohou být:

A) Základní epistemologická námitka, že přeceňujeme schopnost kvantitativního výzkumu popsat a předpovídat chování lidí pomocí pozitivistické redukce jejich postojů a vnímání světa do několika kategorií kvantitativního šetření, které navíc nemají dostatečnou rozlišovací schopnost, aby sledovaly, nakolik se jedná o postoje ukotvené a nakolik jen o odpovědi ad-hoc formulované pro účel zodpovězení dotazníku. (Staněk 2011)

Tyto výhrady, které v různých formulacích opakuje řada sociologů (Krejčí 2003, s. 17) lze samozřejmě oprávněně použít proti většině kvantitativního výzkumu veřejného mínění. Volební výzkum má oproti jiným, netestovatelným zjištěním „nevýhodu“ v okamžité přítomnosti objektivního indikátoru (volby), který ukazuje (ne)kvalitu zjištění. Přítomnost objektivního indikátoru správnosti ovšem nutno chápat z metodologického hlediska také jako pozitivní faktor volebních průzkumů – umožňují totiž reflexi, zda výsledky šetření nejsou systematicky vychýlené.¹⁴³ Účelem takové reflexe je pak nutně výběr design, o kterých víme, že jsou vychýlené méně (např. sledují reprezentativitu podle vhodných kritérií, zohledňují názory nerozhodnutých, atd.).

B) Optimalizační opatření identifikovaná na volbách 2010 mohou fungovat pro tyto konkrétní volby a v roce 2014 selhat, stejně jako v roce 2006 a 2010 uspěly rozdílné metodologie tvorby volebního modelu/volební predikce.

Tomuto skutečnému riziku lze plně předejít jen testováním zjištění na co možná nejvíce datových zdrojích (více voleb, dat více agentur), ale také apriorní opatrností, např.:

¹⁴³ U řady jiných sociologických i marketingových šetření tento prvek konfliktu s realitou chybí, proto se při analýze možnosti systematického vychýlení mohou snáze spokojit s obecnějšími požadavky na metodologické standardy.

- neaplikování vážení a spojování metodologií sběru ad-hoc bez ohledu na to, jak zvyšují reprezentativitu z hlediska faktorů ovlivňujících volební preference,¹⁴⁴
- neaplikování optimalizačních opatření, pokud nás k tomu neopravňují dostatečně slabé předpoklady a výsledky opatření nejsou dostatečně stabilní,¹⁴⁵ či pokud víme, že nedisponujeme dostatečnými informačními zdroji pro zanesení expertního úsudku (např. odhad (ne)účasti proti deklarativní odpovědi respondenta)
- vyvarovat se považování dvou neutralizujících se chyb za optimalizaci a naopak odmítat předčasně odmítat optimalizační opatření, jehož vhodnost se projeví až v kombinaci s jiným.¹⁴⁶

V předkládané práci jsme se podobné postupy, které mohou vést k falešné optimalizaci „na jedno použití“ snažili minimalizovat.

Summary

Analyzing data of the MEDIAN research agency and open sources the study identified some of the possible fields for election-polls methodologic optimizations. In particular it suggest using mixed-mode data collection (CAPI+CATI) or focusing data weighting on specific sociodemographic and media-exposure factors which influence voting behavior. It also suggest methods of weighting involment of undecided voters preferences in the elections predictions. On the other hand the autor suggest to be cautious in considering to predict turn-out of the voters who say they will not (likely or definetely) participate in elections. The study also discuss possibilities of election polls results time-series smoothing. Due to general nature of some findings and simmlarities in the Czech election-polls practices can be some basic findings taken as general suggestions (not only a case-study).

¹⁴⁴ Hrozilo by riziko, že výzkum systematicky vychýlíme, čímž vyrovnam protichůdnou náhodnou složku výběrové chyby nebo jiný druh chyby (např. nezahrnutí nerozhodnutých) přítomnou v omezeném množství testovaných dat. A poté budeme mylně považovat model za optimalizovaný.

¹⁴⁵ Viz předpoklad, že nerozhodnutý respodnent se s velkou pravděpodobností rozhodne pro strany, které zvažuje, a zjištění, že při váženém zařazování nerozhodnutých v každém měsíci a typu sběru dat dochází 4-6 modelů zařazení dochází ke srovnatelnému výsledku. Viz Příloha 4.

¹⁴⁶ Např. u KSČM aplikace optimalizovaného *Alternative_v8* vážení na CATI+CAPI datech preference komunistů nadhodnotila, ale jen dokud jsme nezapočítali nerozhodnuté, mezi kterými KSČM neměli volební potenciál. Z opačného úhlu lze naopak lze hledat vysvětlení, proč k výraznému nadhondocení KDU-ČSL před volbami 2010 došlo u společnosti MEDIAN a nikoli u STEM, CVVM či Factum. Žádná ze společností totiž do volebního modelu podle otevřených zdrojů nezahrnovala nerozhodnuté voliče, ale STEM, CVVM a Factum v jisté míře zařazovali voliče avizující neúčast. U stran s vysokou věrností jádra voličů (velká volební účast příznivců, málo potenciálních voličů mezi nerozhodnutými) se dva podobné kroky mohou vynulovat.

Bibliografie:

Bergar, A. – Gerber, D. – Karlan, D. (2007). *Does The Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions*. USA, Virginia 2007

Bethlehem, J. – Cobben, F. – Schouten, L. (2011). *Handbook in Nonresponse in Household Surveys*. John Wiley & Sons, 2011

Bowers, J. – Ensley, M.J. (2003). *Issues in Analyzing Data from the Dual-Mode 2000 American National Election Study*. Washington, DC: NES Technical Report

CVVM. (2011c). *Strany a volici*. Centrum pro výzkum veřejného mínění. Sociologický ústav AV ČR, Praha 2011

Caplan, B. (2007). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*, Princeton University Press. 2007 Princeton

CVVM (2006). *Voličské preference v květnu 2006*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2006

CVVM (2010a). *Stranické preference a volební model v květnu 2010*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2010

CVVM. (2011c). *Strany a volici*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2011

CVVM (2011d). *Politická orientace českých občanů – září 2011*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2011

CVVM (2011e). *Důvera politickým stranám – září 2011*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2011

CVVM (2011f). *Stranické preference a volební model v říjnu 2011*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha 2011

Český statistický úřad. (2009a). *Informační společnost v číslech 2009, tabulka C1*. Praha 2009. On-line: [http://www.czso.cz/csu/redakce.nsf/i/c_jednotlivci_is2009/\\$File/is09_c.xls](http://www.czso.cz/csu/redakce.nsf/i/c_jednotlivci_is2009/$File/is09_c.xls) (přístupné 14.12.2011 19:00)

Český statistický úřad. (2009b) *Informační společnost v číslech 2009, tabulka B1*. Praha 2009. [http://www.czso.cz/csu/redakce.nsf/i/b_domacnosti_is2009/\\$File/is09_b.xls](http://www.czso.cz/csu/redakce.nsf/i/b_domacnosti_is2009/$File/is09_b.xls) (přístupné 14.12.2011, 19:00)

Český statistický úřad (2010a). *Věkové složení obyvatelstva 2009*. Údaje k 31.12.2009. Online: <http://notes3.czso.cz/csu/2010edicniplan.nsf/p/4003-10> (přístupné 14.12.2011, 19:00)

Český statistický úřad (2010b). *Volby do Poslanecké sněmovny Parlamentu České republiky konané ve dnech 28.05. – 29.05.2010*. Výsledky za volební kraje. Praha 2010. Online: <http://www.volby.cz/pls/ps2010/ps6?xjazyk=CZ> (přístupné 14.12.2011, 19:00)

- Český statistický úřad. (2010c). *Úroveň a obory vzdělávání ve věkových skupinách*. In Zaměstnanost a nezaměstnanost v ČR podle výsledků VŠPS 2010. Údaje pro 1. Kvartál roku 2010. Online: [http://notes3.czso.cz/csu/2011edicniplan.nsf/t/64003169EE/\\$File/31151107.pdf](http://notes3.czso.cz/csu/2011edicniplan.nsf/t/64003169EE/$File/31151107.pdf) (přístupné 14.12.2011, 19:00)
- Český statistický úřad. (2010d). *Příjmy a životní podmínky domácností 2010*. Tab 1.C. Výstup z výzkumu SILC (Statistics on Income and Living Conditions) pro celý rok 2010. Online: <http://www.czso.cz/csu/2011edicniplan.nsf/p/3012-11> (přístupné 14.12.2011, 19:00)
- Český statistický úřad. (2010e). *Počet obyvatel podle pohlaví a jednotek věku*. In Věkové složení obyvatelstva 2010. Údaje k 1.1.2010. Online: [http://notes3.czso.cz/csu/2011edicniplan.nsf/t/28003FC42A/\\$File/401911ri01.pdf](http://notes3.czso.cz/csu/2011edicniplan.nsf/t/28003FC42A/$File/401911ri01.pdf) (přístupné 14.12.2011, 19:00)
- Český statistický úřad. (2010f). *Výběrové šetření pracovních sil*. Průměrné hodnoty za rok 2010. Podrobný výstup zaslaný ČSU v Praze 30.11.2010
- Český statistický úřad. (2010g). *Počet obyvatel v obcích k 1.1.2010*. In Informace o regionech, městech a obcích. Praha 2010.
- Čro. (2010). *Rozhovor se sociologem Janem Hartlem o předvolebních výzkumech*. Český rozhlas 2010. Online: http://www.rozhlas.cz/cro6/stop/_zprava/739043 (přístupné 30.12.2011, 19:00)
- ČTK (2010a). *ČSSD vede o 10 procent, tvrdí průzkum, který si strana zadala*. in Novinky.cz 13.4.2010 . Online: <http://www.novinky.cz/domaci/197421-cssd-vede-o-10-procent-tvrdi-pruzkum-ktery-si-strana-zadala.html> (přístupné 6.1.2012 14:00)
- deRada, V.D. (2010). *Effects (and Defects) of the Telephone Survey in Polling Research: Are We Abusing the Telephone Survey?* In Bulletin of Sociological Methodology 2010, S. 108: 4
- Eurostat. (2010). *Internet use and activities*. In Benchmarking Digital Europe: Key performance indicators. Údaje pro ČR 2010. Eurostat 2010
- Esomar/Wapor (2003). *ESOMAR/WAPOR Guide to Opinion Polls*. Metodologické standardy sdružení Esomar a Wapor. 2003. On-line: http://www.opinamexico.org/docs/Esomar_Wapor_Guide-Opinion.pdf (přístupné 6.1.2012 14:00)
- Factum (2010a). *Podpora Věcí veřejných nadále roste*. Tisková zpráva agentury Factum. Praha 12.5.2010. Online: http://www.factum.cz/395_podpora-veci-verejnych-nadale-roste
- Factum (2010b). *Šest stran, pravostředová většina*. Tisková zpráva agentury Factum. Praha 24.5.2010. Online: http://www.factum.cz/images/zpravy/400/400_sest-stran-pravostredova-vetsina.pdf (přístupné 18.12.2011 11:00)
- Factum (2010c). *Rozhodování českých voličů*. Tisková zpráva agentury Factum. Praha. 25.06.2010. Online: http://www.factum.cz/403_rozhodovani-ceskych-volicu (přístupné 18.12.2011 11:00)

- Factum (2010d). *Voliči se chystají kroužkovat*. Tisková zpráva agentury Factum. Praha. 21.05.2010. Online: http://www.factum.cz/398_volici-se-chystaji-krouzkovat (přístupné 18.12.2011 11:00)
- FiveThirtyEight. (2011). *FiveThirtyEight. Nate's Silver Political Calculus*. Online: <http://fivethirtyeight.blogs.nytimes.com/> (přístupné 7.1.2011 11:00)
- Galtung, J. & Ruge, M. Holmboe (1968): *The Structure of Foreign News. The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers*, [Journal of Peace Research](#), vol. 2, s. 64-91
- Ganninger, M. (2006). *Effective Sample Size - Theoretic Perspectives and Empirical Insights from the ESS*. In QMSS Workshop on 'Measurement, Data Collection and Data Quality Issues for Cross-National Survey Data, Lugano, 2006. Online: <http://www.s3ri.soton.ac.uk/qmss/documents/MatthiasGanninger.pdf> (přístupné 6.1.2012 14:00)
- Ganninger, M. (2011). *Weighting in the ESS. Metodologická příručka*. ESS 2011. Online: <http://essedunet.nsd.uib.no/cms/topics/weight/> (přístupné 6.1.2012 14:00)
- Gaziano, C. (2005). *Comparative Analysis of Within-Household Respondent Selection Techniques*. in Public Opinion Quarterly, Vol 69. No 1. (2005), pp. 124-157. On-line: <http://www.irss.unc.edu/odum/content/pdf/S2%20GazianoPOQ2005.pdf> (přístupné 18.12.2011 11:00)
- Hančlová, J. – Tvrdý, L. (2003). *Úvod do analýzy časových řad*. Ekonomická fakulta VŠB-TU Ostrava. 2003
- Chábová, K. (2011). *Komparace metodických přístupů předvolebních průzkumů používaných v ČR*. Bakalářská práce, FF UK, Praha 2011
- Karp, J. A. (1974). *Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries*. THE JOURNAL OF POLITICS, Vol. 67, No. 3, August 2005, Pp. 825–840
- Kleinová, N. (2010). *Šoková doktrýna – vzestup a pád kalamitního kapitalismu*. Argo/Dokořán. Praha 2010
- Kohler, U. (2007). *Surveys from inside: An assessment of unit nonresponse bias with internal criteria*. in Survey Research Methods, 2007, Vol.1 , No.2 , pp.55-67
- Krejčí, J. (2003). *Výzkumy stranických preferencí, jejich uplatnění ve společnosti a jejich kvalita*. Sociologický ústav Akademie věd České republiky, Praha, 2003
- Krejčí, J. (2008). *Co je to návratnost u výběrových dotazníkových šetření*. In Socioweb SOCIoweb_12_2008. Sociologický ústav AV ČR. Online: http://www.socioweb.cz/upl/editorial/download/161_socioweb_12_08.pdf (přístupné 18.12.2011 11:00)
- Krejčí, J. (200?). *Předvolební průzkumy a média*. In Socioweb. On-line: <http://www.socioweb.cz/index.php?disp=teorie&shw=154&lst=112> (přístupné 18.12.2011 11:00)

Klaschka, J. – Kotrč, E. (2004). *Klasifikační a regresní lesy*. In Sborník konference ROBUST 2004. JČMF 2004

Kotrba, Š. (2006). *Česká televize o koze*. In Britské listy, 10.6.2006. On-line: <http://blisty.cz/art/28887.html> (přístupné 18.12.2011 11:00)

Laver, M. (2005). *Policy and the Dynamics of Political Competition*. In American Political Science Review. Vol. 99, No. 2. 2005.

Lebeda, T.. (2001). *Hlavní proměnné proporčních volebních systémů*. Sociologický časopis, 2001, roč. 37, čís. 4, s. 436.

Lebeda, T. - Hartl, J. – Herzmann, J. – Vinopal, J. – Čech, P. (2010). *Diskuse o výzkumech stranických preferencí*. In Parlamentní volby 2010: Otázky a analýzy. Konference Sociologického ústavu AV ČR, Praha 4.10.2010.

Lebeda, T. - & Krejčí J. & Leontiyeva (2004). *Volební preference, jak jim správně porozumět*. Sociologický Ústav AV ČR, on line: http://www.cvvm.cas.cz/upl/nase_spolecnost/100049s_lebeda-vyzkumy.pdf (přístupné 4.1.2011 11:00)

Lock, K. – Gelman, A. (2008). *Bayesian Combination of State Polls and Election Forecasts*. Department of Statistics, Harvard University 2008

Lukeš, J. – Machek, J. (1988). *Matematická statistika*, SNTL Praha 1988

Martin, P. (2011) *What makes a good mix? Chances and challenges of mixed mode data collection in the ESS*. Centre for Comparative Social Surveys Working Paper Series. City University London, London 2011

Medifax. 2010. *Suverenita Jany Bobošíkové se slučuje s Moravisty, spolu chtějí krajské zastupitelstvo*. On-line: <http://www.mediafax.cz/politika/3287441-Suverenita-Jany-Bobosikove-se-slucuje-s-Moravisty-spolu-chteji-krajske-zastupitelstvo> (přístupné 1.1.2012 14:00)

Médea research (2010a). *VOLEBNÍ PROGNÓZA A STRANICKÉ PREFERENCE*. Tisková zpráva agentury Médea research, Praha, Květen 2010

MEDIAN. (2003). *Software Vážení 4.00*. MEDIAN s.r.o, Praha 2003

MEDIAN. (2010a). *VOLEBNÍ PREFERENCE za duben 2010*. Tisková zpráva agentury MEDIAN. Praha, Květen 2010. On-line: <http://www.median.cz/index.php?lang=cs&page=4&sub=5> (přístupné 18.12.2011 14:00)

MEDIAN. (2011). *Data Analyzer - program pro analýzu dat v médiálních a reklamních agenturách, médiích a marketingových odděleních firem*. Aktuální verze programu. MEDIAN s.r.o, Praha 2010

Neustadt, J. (1997). *Problematika vah ve výběrových šetřeních*. Matematicko-fyzikální fakulta UK. Diplomová práce. Praha 1997

Noelle-Neumann, E. (1974). *The spiral of silence: a theory of public opinion*. in Journal of Communication, 24. 1974. s. 43-51

Parlamentní listy (2010). *Dvanáct procent pro Bobošíkovou*. STEM agenturu nezná. Praha 24.5.2010. Online: <http://www.parlamentnilisty.cz/arena/monitor/Dvanact-procent-pro-Bobosikovou-STEM-agenturu-nezna-166204> (aktivní 18.12.2011 14:00)

Popper, K. 1985. *Two Kinds of Definition*. in David Miller (ed.), Popper Selections, 1985

SANEP (2010a). *Volební preference politických stran - květen 2010*. Tisková zpráva společnosti SANEP. Praha, květen 2010

SC&C (2006). *Volby 2006. Zpráva společnosti SC&C k předvolebním výzkumům a exit-pollu v roce 2006*. Online <http://www.scac.cz/vyzkumy/archiv/96-volby-2006> (přístupné 18.12.14:00)

SC&C (2010b). *Závěrečná zpráva. EXIT POLL SC&C a SPSS CR pro Českou televizi*. Praha, 1. června 2010

SCIO (2010). *100 chytrých*. Praha, 16.8.2010. On-line http://www.100chytrych.cz/o_projektu#volby (aktivní 18.12.2011 14:00)

Shamir, J. (1986). *Preelection Polls in Israel: Structural Constraints on Accuracy*. Public Opin Q (1986) 50 (1): 62-75.

Shulz, W. - Reinemann, C. (2006). *Introduction to the special issue: News decisions and news values*. In Communications. Volume 31, Issue 1. 2006. P. 1-4

Schoen, H. – Schumann, S. (2007). *Personality Traits, Partisan Attitudes, and Voting Behavior*. Evidence from Germany. Political Psychology 28, 2007, p. 471-498

Staněk, M. (2011). *Osobní rozhovor s PhDr. Milošem Staňkem*. Praha 20.10.2011

STEM (2006). *Porovnání předvolebních výzkumů a výsledků voleb*. Tisková informace společnosti STEM. Praha, 5.6.2006.

STEM (2010a). *Preference politických stran – květen 2010*. Tisková informace společnosti STEM. Praha, 17.5.2010.

STEM (2010b). *Které motivy byly ve volbách pro voliče nejdůležitější*. Informace z výzkumu STEM Trendy 6/2010. Praha. 21.7.2010

STEM (2010c). *Polovina nevoličů se o své neúčasti rozhodla až v posledním měsíci před volbami. O nejvíce hlasů přišla sociální demokracie*. Informace z výzkumu STEM Trendy 6/2010. Praha. 21.7.2010

STEM (2011a). *Nabídka výzkumu – popis metodologie omnibusu TRENDY*. Online: <http://www.stem.cz/staticpages/ofirme-nabidka> (přístupné 6.1.2012 19:00)

STEM (2011b). *Proč se liší preference politických stran u různých firem pro výzkum veřejného mínění?* Sekce v metodologické rubrice Zeptali jste se nás na stránkách STEM. Online: <http://www.stem.cz/pages/faq.php> (přístupné 6.1.2012 19:00)

- SÚ (2010). *PARLAMENTNÍ VOLBY 2010: OTÁZKY A ANALÝZY*. Konference Sociologického ústavu AV ČR, Praha 4.10.2010. Online: <http://www.soc.cas.cz/articles/cz/12/5223/KONFERENCE-PARLAMENTNI-VOLBY-2010-OTAZKY-A-ANALYZY.html> (přístupné 6.1.2012 19:00)
- Válková, H. (2010). *Průzkumy selhaly, agentury hledají jinou cestu do hlav voličů*. In iDNES.cz. 3.6.2010. Online: http://zpravy.idnes.cz/pruzkumy-selhaly-agentury-hledaji-jinou-cestu-do-hlav-volicu-put-/domaci.aspx?c=A100603_152313_domaci_hv (přístupné 6.1.2012 19:00)
- Vlachová, K. (2000). *Stranická identifikace v České republice*. Sociologický ústav AV ČR, Praha 2000
- Vinopal, J. (2006). *Povaha předvolebních průzkumů a jejich role v rozhodování voličů*. Centrum pro výzkum veřejného mínění, Sociologický ústav AV ČR, Praha
- Walsh & Dolphin & DiNardo. (2009) *Lies, Damn Lies, and Pre-Election Polling*. University of Michigan, Michigan 2009
- Weimann, G. (1990). *THE OBSESSION TO FORECAST: PRE-ELECTION POLLS IN THE ISRAELI PRESS*. Public Opin Q (1990) 54 (3): 396-408
- Westerstahl, J. - Johansson, F. (1994). *Foreign News: News Values and Ideologies*. In European Journal of Communication, Vol. 9, No. 1, 71-89 (1994)
- Wirnitzer. (2011). *Historická remíza s komunisty ODS neděsí, viní z ní reformy i sebe*. In iDNES.cz, 1. Srpna 2011. Online: http://zpravy.idnes.cz/historicka-remiza-s-komunisty-ods-nedesi-vini-z-ni-vv-reformy-i-sebe-11h-/domaci.aspx?c=A110801_143349_domaci_jw (přístupné 6.1.2012 19:00)
- WITTGENSTEIN, L. *Tractatus logico-philosophicus*, OIKOYMENH, Praha 2008
- Wessell D, R. – Rudolph, T. (2000). *An Analysis of the 1998 NES Mixed-Mode Design*. NES (National Election Studies) Technical Reports. Ann Arbor: University of Michigan.
- Zahradníček, S. (2010). *Rozhovor - Stanislav Zahradníček (SIMAR)*. Výzkumy.cz, 2010. Online: <http://www.vyzkumy.cz/clanky/225-rozhovor-stanislav-zahradnicek-simar> (přístupné 6.1.2012 19:00)

Příloha 1: Metadata - metodologie CAPI a CATI výzkumu, dotazový aparát

Metodologie zdrojového výzkumu pro CAPI volební data MEDIAN:

- metoda dotazování = osobní dotazování vyškoleným tazatelem v domácnosti respondenta s použitím laptopu, dotazování 7 dní v týdnu
- délka fáze = 1 měsíc (analýza a publikace volebních dat), ½ měsíce (výběrová metoda)
- velikost měsíčního vzorku = 950-1100 respondentů ve věku 12-79 let
- výběrová metoda = víceetapový stratifikovaný náhodný adresní výběr
- strata = velikost místa bydliště (5 kategorií) a regiony NUTS 3 (14 kategorií)
- opora výběru = průběžně aktualizovaná konkretizovaná databáze nemovitostí ČSÚ
- metoda náhodného výběru jednotek:
 - o náhodný systematický výběr (v 1. i 2. výběrovém stupni)
 - o výběr respondenta v domácnosti podle Kishovy tabulky
- počet návratů při nezastižení náhodně vybrané jednotky = min. 3
- možnost substituce:
 - o na úrovni respondenta = NE
 - o na úrovni adresy = omezená, standardizovaná¹⁴⁷
- eliminace skupin s větší pravděpodobností zařazení do vzorku¹⁴⁸
- kontrola = min 30 % rozhovorů (regule SIMAR)

Metodologie zdrojového výzkumu pro CATI volební data MEDIAN:

- metoda dotazování = telefonické dotazování vyškoleným tazatelem z telefonického studia v Praze, dotazování 7 dní v týdnu
- délka fáze = 1 měsíc
- velikost měsíčního vzorku = 1000-1250 respondentů ve věku 12-79 let
- výběrová metoda:
 - o prostý náhodný výběr (výběr mobilních kontaktů)
 - o stratifikovaný náhodný výběr (výběr pevných linek)
 - o sledování základních sociodemografických znaků (eliminování respondentů z náhodného výběru, jejichž dotázání by výrazně narušilo strukturu vzorku měsíční fáze podle zákl. sociodemografických znaků – věk, pohlaví, vzdělání, kraj, VMB)
- opora výběru = databáze pevných linek + generování mobilních kontaktů (voláno v poměru 40:60)
- možnost substituce na úrovni respondenta = NE
- eliminace skupin s větší pravděpodobností zařazení do dat (členové domácnosti častěji zvedající pevné linky)
- kontrola = náslech min 5 % rozhovorů

¹⁴⁷ Tazatel má při splnění minimálního počtu rozhovorů výše popsanou metodikou na uvedených adresách možnost omezený počet neúspěšných adrese nahradit standardizovaným postupem výběru sousední adresy. Takto nahrazené adresy tvoří většinou jen 5-6 % vzorku.

¹⁴⁸¹⁴⁸ Výběrová metoda standardizovaně a s jistou pravděpodobností vyřazuje respondenty z jedno- a dvoučlenných domácností (zvýhodnění náhodným adresním výběrem) v závislosti na věku (pro úplné nepotlačení mladých méněčlenných domácností).

Dotazník volebního výzkumu (shodný pro CAPI i CATI):

IF AGE>17.

V01. Zúčastnil(a) byste se voleb do Poslanecké sněmovny, kdyby se konaly nyní?¹⁴⁹

- 1) určitě ano
- 2) spíše ano
- 3) spíše ne
- 4) určitě ne

IF V01=1 or 2.

V02. Kterou stranu nebo hnutí byste asi volil(a), kdyby se volby do Poslanecké sněmovny konaly nyní?¹⁵⁰ *(otevřená otázka)*

IF V01=3.

V02. Pokud byste přeci jen k volbám šel(šla), jakou stranu nebo hnutí byste asi volil(a)? *(otevřená otázka)*

- X) nevím, zatím nejsem rozhodnut(a)
- Y) nemohu si vzpomenout na název strany
- Z) odmítá uvést / nevolí

IF V02=Y.

V03. Kterou stranu, nebo hnutí byste tedy volil(a)?

(uzavřená otázka – tazatel předkládá seznam stran)

IF (V01=1 or 2) and (V02=X).

V04. Uvedl(a) jste, že kdyby se volby konaly nyní, tak byste se jich asi zúčastnil(a), ale nejste rozhodnut(a), koho byste volil(a). O jakých stranách nebo hnutích byste při rozhodování, koho volit, uvažoval(a)? (1 - 3 odpovědi).

(otevřená otázka – pokud si respondent nemůže vzpomenout na stranu, tazatel předkládá seznam stran – tj. systém obdobný v02)

V20. Kterou stranu, nebo hnutí jste volil(a) při posledních volbách do Poslanecké sněmovny, které se konaly v červnu roku 2006?¹⁵¹

(uzavřená otázka – abecední seznam stran překračujících zisk 1 % v posledních volbách do PSP ČR + jiné strany)

Y nevím, už si nepamatuji

Z voleb jsem se nezúčastnil(a)

AA voleb jsem se nemohl(a) zúčastnit, byl(a) jsem mladší 18 let

Zdroj: informace společnosti MEDIAN

¹⁴⁹ V těsně předvolebním období roku 2010 nahrazeno zněním: „V01. Zúčastníte se voleb do Poslanecké sněmovny (28. až 29. května 2010)?“

¹⁵⁰ V těsně předvolebním období roku 2010 nahrazeno zněním: „V02. Kterou stranu nebo hnutí budete asi volit v květnových volbách do Poslanecké sněmovny?“

¹⁵¹ V povolebním období nahrazeno zněním na volby v květnu 2010.

Příloha 2. Tabulky kategorizace chybovosti a odchýlení + rekód aktivit

TAB 1. Kategorizace predikční kvality podle hodnotících kritérií

KVALITA PREDIKCE	Suma čtverců odchylek	Prům.abs. odchylka (p.b.)	Max.abs. odchylka (p.b.)	Prům. rel. odchylka (%)	Max. rel. odchylka (%)	Odchylka v rozdílu mezi 1. a 2. stranou (p.b.)	Počet chyb v kvalifikaci stran
1 - výborná	0-9,99	0-0,49	0-0,99	0-4,99	0-19,99	0-0,99	0 chyb
2	10-39,99	0,5-1,49	1-2,99	5-14,99	20-39,99	1-1,99	-
3	30-69,99	1,5-2,49	2-4,99	15-24,99	40-59,99	2-2,99	1 chyba
4	50-99,99	2,5-3,49	5-6,99	25-34,99	60-79,99	3-3,99	-
5	70-129,99	3,5-4,49	7-8,99	35-44,99	80-99,99	4-4,99	2 chyby
6	90-159,99	4,5-5,49	9-10,99	45-54,99	100-119,99	5-5,99	-
7 - zcela chybná	160 a více	5,5 a více	11 a více	55 % a více	120 % a více	6 p.b. a více	více než 2 chyby

TAB 2. Intervaly slabého vychýlení oproti volebním výsledkům 2006

	Volby 2006	Interval slabého vychýlení (výzkumy od 12.5.)		Interval slabého vychýlení (výzkumy do 12.5.)	
		dolní hranice	horní hranice	dolní hranice	horní hranice
ODS	35,4	31,6	39,3	28,5	43,2
ČSSD	32,3	28,7	36,1	25,8	39,8
KSČM	12,8	10,3	15,7	9,3	17,3
KDU-ČSL	7,2	5,3	9,5	4,8	10,5
SZ	6,3	4,5	8,5	4,1	9,3
Ostatní strany	6,0	4,3	8,1	3,8	9,0

TAB 3. Intervaly slabého vychýlení oproti volebním výsledkům 2010

	Volby 2010	Interval slabého vychýlení (výzkumy od 7.5. 2010)		Interval slabého vychýlení (výzkumy do 7.5. 2010)	
		dolní hranice	horní hranice	dolní hranice	horní hranice
ČSSD	22,1	18,9	25,6	17,0	28,1
ODS	20,2	17,1	23,6	15,4	25,9
TOP 09	16,7	13,9	19,9	12,5	21,8
KSČM	11,3	8,9	14,0	8,0	15,5
VV	10,9	8,6	13,6	7,7	15,0
KDU-ČSL	4,4	2,9	6,3	2,6	6,9
SPOZ	4,3	2,8	6,2	2,6	6,8
Suverenita	3,7	2,4	5,5	2,1	6,0
SZ	2,4	1,3	3,9	1,2	4,3
Ostatní	4,0	2,6	5,8	2,3	6,4

TAB 4 Domácí a venkovní volnočasové aktivity¹⁵²

Č.	Aktivita	Typ aktivity	Č.	Aktivita	Typ aktivity
1	Kondiční cvičení	neutrální	15	Fotografování	venkovní
2	Týmové sporty, hry (kopaná, basketbal,...)	venkovní	16	Kultura, umění – aktivní, tvůrčí činnost	neutrální
3	Individuální sporty ve volné přírodě	venkovní	17	Kultura, umění – návštěva kina, kulturních akcí	venkovní
4	Návštěva sportovních utkání	venkovní	18	Zahrádkaření	neutrální
5	Ruční práce (pletení, šití, apod.)	domácí	19	Péče o domácí zvířata	domácí
6	Kuťlství, práce na domě, bytě	domácí	20	Věnování se partnerovi, milování	domácí
7	Práce na autě a jiné technické zájmy	neutrální	21	Vycházky, hry, sportování s dětmi, vnoučaty	venkovní
8	Chataření, chalupaření	neutrální	22	Posezení s přáteli	neutrální
9	Cestování, výlety po ČR	venkovní	23	Návštěva velkých zábavních, nákupních center	venkovní
10	Cestování mimo ČR	venkovní	24	Návštěva diskoték, plesů, tanečních zábav	venkovní
11	Sledování videa, DVD	domácí	25	Návštěva společenských akcí, večírků	venkovní
12	Poslech CD, MC	neutrální	26	Návštěva restaurací, kaváren, barů, hospod	venkovní
13	Hraní her na počítači	domácí	27	Lenošení, nicnedělání	domácí
14	Četba knih	neutrální			

¹⁵² Kondiční cvičení (1), aktivní umění/tvůrčí činnost (15), posezení s přáteli (22), poslech hudby (12), četbu knih (14) může respondent provozovat doma i mimo domov; práce na autě (7) může probíhat v těsné blízkosti domu i mimo; u chataření a chalupaření a zahrádkaření (8,9) není podmínkou, že daný objekt má či má poštovní adresu a jestli respondent slyší příchod tazatele – tj. jestli touto činností respondent snižuje šanci na zastižení.

Příloha 3: Srovnání reprezentativity jednotlivých (CAPI,CATI) a spojených sběrů (CAPI+CATI)

TAB 1. Základní sociodemografie – struktura plných vzorků vzhledem k populaci (12-79 let)

POZN: Srovnávány jsou data před vážením, s nímž oba sběry počítají. Hodnoty tak nelze brát jako hodnocení finální reprezentativity produkované sběry dat, ale spíše jako komparativní hodnocení typu vychýlení, k němuž bez vážení směřují.

	Teor.čet.	CAPI				CATI				CATI+CAPI
		04/10 (n=1016)	05/10 (n=980)	07/10 (n=1077)	04+05+07/10 (n=3073)	04/10 (n=1358)	05/10 (n=1343)	07/10 (n=1040)	04+05+07/10 (n=3841)	04+05+07/10 (n=6941)
OBLAST	(ČSU 2010a)									
Praha + Stř. Čechy	23,7%	-0,8	-3,6	-2,8	-2,4	-1,8	1,0	0,8	-0,1	-1,1
Zbytek čech	36,7%	-0,8	0,7	-0,8	-0,3	2,2	1,8	2,6	2,2	1,1
Morava	39,5%	1,6	2,9	3,6	2,7	-0,4	-2,8	-3,4	-2,1	0,0
Chi-sq (sig)		0,58	0,02	0,02	0,00	0,15	0,10	0,06	0,01	0,06
Chi-sq (hodnota)		11,40	25,14	24,77	33,83	18,25	19,68	21,89	28,03	21,78
Suma abs. Odch.		3,2	7,2	7,3	5,4	4,4	5,7	6,8	4,4	2,2
VZDĚLÁNÍ	(ČSU 2010c)									
ZŠ či žádné	18,0%	-1,1	-1,3	0,2	-0,7	-4,4	-1,4	6,0	-0,3	-0,5
SŠ bez maturity	34,7%	4,7	3,9	6,4	5,0	0,7	8,8	3,6	4,4	4,7
SŠ s maturitou	34,0%	2,5	1,4	-1,6	0,7	0,4	-2,8	-7,6	-3,1	-1,4
VŠ	13,2%	-6,0	-3,9	-5,0	-5,0	3,3	-4,6	-2,0	-1,0	-2,8
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Chi-sq (hodnota)		37,0	17,3	33,8	81,9	26,4	55,9	50,0	35,2	89,2
Suma abs. Odch.		14,3	10,5	13,2	11,5	8,8	17,6	19,3	8,8	9,4
POHLAVÍ	(ČSU 2010e)									
MUZ	49,5%	-5,3	-7,2	-4,9	-5,8	0,5	-1,0	-0,7	-0,4	-2,8
ZENA	50,5%	5,3	7,2	4,9	5,8	-0,5	1,0	0,7	0,4	2,8
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,72	0,48	0,62	0,63	0,00
Chi-sq (hodnota)		11,5	20,1	10,5	41,1	0,1	0,5	0,3	0,2	21,4
Suma abs. Odch.		10,6	14,3	9,9	11,6	1,0	1,9	1,5	0,8	5,6
VMB	(ČSU 2010g)									
do 999 ob.	16,6%	1,7	-4,5	0,5	-0,7	3,1	3,6	2,0	2,9	1,3
1 000 - 4 999 ob.	20,7%	-3,8	-0,6	-4,1	-2,9	-2,9	-2,3	-0,1	-1,8	-2,3
5 000 - 19 999 ob.	18,6%	0,8	0,3	0,3	0,5	1,7	1,6	0,3	1,3	0,9
20 000 - 99 999 ob.	22,1%	1,9	3,5	3,4	2,9	-0,4	-2,3	-1,9	-1,5	0,5
100 000 ob. a více	22,0%	-0,6	1,3	-0,1	0,2	-1,6	-0,6	-0,4	-0,9	-0,4
Chi-sq (sig)		0,02	0,00	0,01	0,00	0,00	0,00	0,32	0,00	0,28
Chi-sq (hodnota)		11,2	18,6	14,5	25,6	16,9	19,1	4,7	34,6	5,1
Suma abs. Odch.		8,8	10,3	8,3	7,2	9,6	10,4	4,6	8,4	5,4
VĚK	(ČSU 2010a)									
12 - 17 let	7,1%	3,0	2,8	3,4	3,1	-3,7	-1,0	0,0	-1,6	0,5
18 - 24 let	10,8%	-4,4	-3,2	-4,1	-3,9	-1,0	0,4	0,7	0,0	-1,7
25 - 34 let	18,9%	-5,1	-6,3	-6,2	-5,9	-1,3	-0,9	-2,2	-1,4	-3,4
35 - 44 let	17,3%	0,0	-0,3	0,1	-0,1	-0,6	1,1	0,7	0,4	0,2
45 - 54 let	15,5%	0,5	0,9	1,2	0,9	0,8	-0,4	0,7	0,3	0,6
55 - 64 let	16,6%	1,1	3,0	1,5	1,8	2,7	2,4	-0,6	1,6	1,7
65 - 79 let	13,7%	5,1	3,0	4,1	4,1	3,1	-1,7	0,6	0,7	2,2
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,08	0,61	0,00	0,21
Chi-sq (hodnota)		64,7	52,5	72,0	184,7	45,3	11,3	4,5	26,9	8,5
Suma abs. Odch.		19,1	19,4	20,6	19,7	13,3	8,0	5,5	6,1	10,3

TAB 2. Základní sociodemografie – struktura vzorku voličů a předpokládané populace voličů¹⁵³

	Teor. Čet.	CAPI			CATI			CAPI + CATI
VZDĚLÁNÍ:	(SC&C 2010b)	04/10* (n=602)	05/10* (n=498)	04+05/10* (n=1100)	04/10* (n=851)	05/10* (n=713)	04+05/10* (n=1564)	04+05/10* (n=2664)
ZŠ či žádné	6%	0,6	-0,4	0,2	3,2	4,5	3,8	2,3
SŠ bez maturity	26%	14,5	13,8	14,2	6,8	17,2	11,5	12,6
SŠ s maturitou	45%	-2,0	-3,0	-2,5	-6,8	-8,4	-7,5	-5,4
VŠ	23%	-13,2	-10,3	-11,9	-3,1	-13,3	-7,8	-9,5
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Chi-sq (hodnota)		125,71	87,39	188,66	65,84	206,66	213,62	348,69
Suma abs. Odch.		30,4	27,5	28,7	19,9	43,4	30,6	29,8
VĚK:	(SC&C 2010b)							
18 - 29 let	21%	-10,5	-7,3	-9,1	-4,1	-4,3	-4,2	-6,2
30 - 44 let	29%	-1,3	-4,7	-2,8	-4,3	2,0	-1,4	-2,0
45 - 59 let	26%	2,9	5,9	4,3	0,4	4,0	2,1	3,0
60 a více let	24%	8,9	6,1	7,6	8,0	-1,7	3,6	5,2
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,01	0,00	0,00
Chi-sq (hodnota)		62,88	38,72	90,75	42,66	18,23	32,11	102,58
Suma abs. Odch.		23,6	24,1	23,8	16,8	12,0	11,3	16,4
POHLAVÍ:	(SC&C 2010b)							
Muž	48%	-0,8	-6,4	-3,4	4,8	0,5	2,8	0,3
Žena	52%	0,8	6,4	3,4	-4,8	-0,5	-2,8	-0,3
Chi-sq (sig)		0,69	0,00	0,03	0,01	0,78	0,03	0,78
Chi-sq (hodnota)		0,16	8,26	4,99	7,73	0,08	5,02	0,08
Suma abs. Odch.		1,6	12,9	6,7	9,5	1,1	5,7	0,5
OBLAST	(ČSU 2010b)							
Praha + Stř.Čechy	24,2%	-9,8	-10,5	-10,1	0,4	-2,5	-1,0	-5,0
Zbytek čech	36,0%	2,2	-7,0	-2,2	12,9	12,0	12,5	6,0
Morava	39,8%	4,6	-3,7	0,6	6,0	-7,6	-0,5	0,0
Chi-sq (sig)		0,02	0,22	0,01	0,18	0,00	0,00	0,00
Chi-sq (hodnota)		25,16	16,53	29,21	17,48	30,07	31,87	35,53
Suma abs. Odch.		16,6	21,2	12,9	19,4	22,1	13,9	11,1

POZN: S teoretickými četnostmi jsou srovnávány subvzorky respondentů, kteří avizovali účast u voleb.

** Data z května 2010 jsou použita do 28. 5. 2010 14:00*

¹⁵³ Ve třech ze čtyř kategorií vycházíme z dat exit-pollu (SC&C 2007), které mohou být rovněž zatíženy chybou.

TAB 3. Specifické sociodemografie a další faktory – struktura vzorků vzhledem k popsané populaci

	Teor.čet.	CAPI				CATI				CATI+CAPI
PRACOVNI STATUS¹⁵⁴	(ČSU 2011f)	04/10	05/10	07/10	04+05+07/10	04/10	05/10	07/10	04+05+07/10	04+05+07/10
zaměstnanec	46,5%	1,4	-0,7	-1,8	-0,4	-1,8	-2,8	-4,1	-2,8	-1,8
podnikatel, (spolu) vlastník firmy bez zam.	8,0%	-3,4	-3,4	-1,9	-2,8	-1,0	1,1	-0,5	-0,1	-1,3
podnikatel (spolu) vlastník firmy se zam.	2,1%	-1,1	-0,8	-0,4	-0,7	0,1	0,1	-0,4	-0,1	-0,3
nezaměstnaný	4,4%	-0,1	0,1	0,1	0,0	-0,1	-0,4	-0,5	-0,3	-0,1
důchodce (nepracující)	23,7%	6,0	6,4	5,3	5,9	4,0	-0,8	1,8	1,6	3,5
v domácnosti, mateřská dovolená	4,1%	-0,3	0,2	-0,7	-0,3	0,5	1,5	0,6	0,9	0,4
student, žák, učeň	10,4%	-3,0	-1,8	-0,3	-1,7	-2,9	0,3	1,4	-0,5	-1,0
jiné	0,7%	0,4	0,0	-0,3	0,0	1,2	0,9	1,7	1,3	0,7
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,04	0,14	0,00	0,00
Chi-sq (hodnota)		43,44	37,63	30,32	75,74	39,87	22,86	18,53	29,84	67,61
Suma abs. Odch.		15,7	13,4	10,8	11,9	11,4	7,9	11,0	7,6	9,2
PRÍJEM NA ČLENA¹⁵⁵	(ČSU 2011d)									
do 4000	3,1%	-0,9	1,0	-0,3	-0,1	0,9	0,4	-0,3	0,4	0,2
4001 - 6000	6,0%	0,8	1,3	0,7	0,9	4,6	6,2	10,0	6,7	4,0
6001 - 8000	14,4%	-3,9	-3,2	-1,2	-2,7	-0,4	4,1	1,5	1,7	-0,3
8001 - 10000	21,0%	7,2	4,9	5,1	5,7	3,3	1,9	3,1	2,7	4,1
10001 - 15000	36,4%	0,0	-4,4	-3,1	-2,5	-6,2	-8,7	-10,3	-8,2	-5,6
15001 - 20000	11,7%	-2,8	-1,3	-1,9	-2,0	-3,9	-3,6	-3,7	-3,7	-2,9
20001 - 30000	5,5%	-0,3	1,7	1,1	0,9	0,9	-0,2	-0,2	0,2	0,5
30001 - 50000	1,5%	0,4	0,4	0,1	0,4	1,0	0,1	0,2	0,5	0,4
vice nez 50 000	0,5%	-0,4	-0,4	-0,5	-0,5	-0,1	-0,3	-0,4	-0,2	-0,4
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Chi-sq (hodnota)		55,76	45,73	32,62	98,74	103,79	145,84	226,78	395,55	346,78
Suma abs. odch.		16,8	18,7	14,1	15,7	21,3	25,5	29,7	24,3	18,5
POČET ČLENU¹⁵⁶	(ČSU 2011d)									
1 až 2	34,9%	3,8	2,6	5,0	3,8	11,4	8,8	6,6	9,1	6,7
3 až 4	53,6%	1,9	2,9	0,9	1,9	-9,9	-8,6	-7,4	-8,7	-4,0
více než 4	11,5%	-5,7	-5,4	-5,8	-5,7	-1,5	-0,1	0,8	-0,4	-2,7
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Chi-sq (hodnota)		59,95	53,70	67,32	135,74	111,26	77,62	51,08	184,38	193,64
Suma abs. odch.		11,4	10,8	11,7	11,3	22,8	17,6	14,7	18,2	13,5
INTENZITA UŽÍVÁNÍ INTERNETU¹⁵⁷	(Eurostat 2010)									
silní uživatelé (denně)	38 %	-9,5	-8,9	-9,3	-9,2	16,4	13,6	13,3	14,5	4,0
střední uživatelé	31 %	-3,4	-2,8	-3,4	-3,2	-12,0	-9,3	-10,2	-10,5	-7,3
neuživatelé	31 %	12,9	11,5	12,7	12,4	-4,4	-4,3	-3,2	-4,0	3,3
Chi-sq (sig)		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Chi-sq (hodnota)		128,80	107,25	130,72	309,22	219,16	182,70	137,39	471,30	235,61
Suma abs. odch.		25,9	23,2	25,4	24,9	32,8	27,2	26,7	29,1	14,5

¹⁵⁴ Srovnána je část vzorku 15-79 let. U omezení populace 15+ zkoumané ČSU (2010f) je použit předpoklad, že lidé nad 79 jsou důchodci.

¹⁵⁵ Srovnáván je plný vzorek 12-79 s údaji pro celou populaci (ČSU 2010d).

¹⁵⁶ Srovnáván je plný vzorek 12-79 s údaji pro celou populaci (ČSU 2010d).

¹⁵⁷ Srovnána je část vzorku 16-74 let odpovídající populaci zkoumané Eurostatem (2010a).

Příloha 4: Kvalita predikce při zařazení nerozhodnutých voličů

		DUBEN 2010 - CAPI+CATI – vážení Alternative v8									
	Vysledky voleb	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	Rovnoměrně
ČSSD	22,10	26,5	26,3	26,1	25,9	25,9	25,7	25,5	25,3	25,2	25,1
ODS	20,22	22,4	22,3	22,2	22,1	22,1	22,0	22,0	21,9	21,8	21,7
TOP09	16,71	11,0	13,1	13,1	12,9	13,0	12,9	12,8	12,7	12,5	12,4
KSČM	11,27	13,9	11,0	11,1	11,3	11,2	11,4	11,5	11,8	12,0	12,1
VV	10,88	8,2	9,7	9,7	9,7	9,8	9,8	9,8	9,8	9,8	9,8
KDU-ČSL	4,39	5,4	4,3	4,5	4,6	4,6	4,7	4,9	5,0	5,1	5,2
SPOZ	4,33	4,8	4,6	4,7	4,7	4,7	4,7	4,8	4,8	4,8	4,9
Suverenita	3,67	1,4	1,3	1,3	1,3	1,3	1,3	1,3	1,4	1,5	1,5
SZ	2,44	2,4	3,0	3,1	3,1	3,1	3,1	3,1	3,2	3,2	3,2
Ostatní	3,99	4,0	4,4	4,4	4,4	4,3	4,3	4,2	4,2	4,2	4,1
Hodnocení chybovosti											
Suma čtverců odchylek		76,8	43,0	41,0	40,6	39,1	38,9	37,2	37,4	38,0	37,8
Prům. abs. odchylka		2,13	1,52	1,48	1,47	1,45	1,46	1,45	1,48	1,51	1,51
Maximální abs. odchylka		5,68	4,19	3,99	3,84	3,77	3,85	3,88	4,06	4,25	4,32
Prům. relativní odch.		20,8%	17,4%	17,3%	17,3%	17,3%	17,6%	17,8%	18,2%	18,7%	18,9%
Max. relativní odch.		61,2%	65,4%	65,2%	63,8%	64,9%	63,5%	63,3%	61,8%	60,3%	59,7%
Rozdíl mezi lídry		2,2	2,1	2,0	1,9	1,9	1,8	1,7	1,6	1,5	1,4
Chyby v kvalifikaci stran		0	0	0	0	0	0	0	0	0	0
Kategorie kvality predikce											
Součet čtverců odchylek		4	3	3	3	2	2	2	2	2	2
Průměrná abs. odchylka		3	3	2	2	2	2	2	2	3	3
Maximální abs. odchylka		4	3	3	3	3	3	3	3	3	3
Prům. relativní odch.		3	3	3	3	3	3	3	3	3	3
Max. relativní odch.		4	4	4	4	4	4	4	4	4	3
Rozdíl mezi lídry		3	3	3	2	2	2	2	2	2	2
Chyby v kvalifikaci stran		1	1	1	1	1	1	1	1	1	1
Cellkový index kvality		3,1	2,9	2,7	2,6	2,4	2,4	2,4	2,4	2,6	2,4

Pozn: Pro kategorizaci kvality predikce jsou použita kritéria z Přílohy 2. Vizualizace kvality predikce je provedena v MS Excel, automatickým podmíněným formátováním color scales (rozložení barev mezi minimum a maximum v daném řádku).

DUBEN-KVĚTEN 2010 - CAPI+CATI - vážení Alternative v8											
	volby	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	rovnomerně
ČSSD	22,10	22,5	23,2	22,9	22,7	22,6	22,4	22,1	21,9	21,7	21,5
ODS	20,22	19,5	20,5	20,4	20,2	20,3	20,1	20,0	19,8	19,6	19,6
TOP09	16,71	13,0	15,1	15,2	14,9	15,2	15,0	15,0	14,8	14,5	14,4
KSČM	11,27	15,9	12,3	12,5	12,7	12,7	12,9	13,1	13,3	13,5	13,6
VV	10,88	9,1	9,7	9,7	9,8	9,8	9,9	10,0	10,1	10,1	10,2
KDU-ČSL	4,39	5,9	5,0	5,1	5,2	5,2	5,3	5,4	5,5	5,6	5,7
SPOZ	4,33	3,6	3,8	3,8	3,8	3,8	3,9	3,9	3,9	3,9	4,0
Suverenita	3,67	1,9	1,6	1,7	1,8	1,7	1,8	1,8	1,9	2,0	2,1
SZ	2,44	3,3	3,6	3,6	3,6	3,6	3,6	3,7	3,7	3,7	3,7
Ostatní	3,99	5,4	5,1	5,1	5,1	5,1	5,1	5,1	5,1	5,2	6,2
Hodnocení chybovosti											
Suma čtverců odchylek	47,7	13,8	13,3	14,0	13,0	14,0	14,3	16,1	18,3	22,8	
Prům. abs. odchylka	1,75	1,07	1,04	1,05	1,01	1,04	1,03	1,10	1,19	1,33	
Maximální abs. odchylka	4,63	2,03	1,99	1,91	1,95	1,87	1,83	2,02	2,25	2,38	
Prům. relativní odch.	25,5%	19,4%	19,3%	19,6%	19,3%	19,6%	19,6%	20,2%	20,8%	23,6%	
Max. relativní odch.	48,2%	55,4%	54,3%	52,0%	53,2%	50,9%	49,7%	51,0%	52,4%	54,4%	
Rozdíl mezi lídry	1,1	0,8	0,6	0,6	0,5	0,4	0,3	0,2	0,2	0,1	
Chyby v kvalifikaci stran	0	0	0	0	0	0	0	0	0	0	
Kategorie kvality predikce											
Suma čtverců odchylek	3	2	2	2	2	2	2	2	2	2	
Průměrná abs. odchylka	3	2	2	2	2	2	2	2	2	2	
Maximální abs. odchylka	3	2	2	2	2	2	2	2	2	2	
Prům. relativní odch.	4	3	3	3	3	3	3	3	3	3	
Max. relativní odch.	3	3	3	3	3	3	3	3	3	3	
Rozdíl mezi lídry	2	1	1	1	1	1	1	1	1	1	
Chyby v kvalifikaci stran	1	1	1	1	1	1	1	1	1	1	
Celkový index kvality	2,7	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	

		KVĚTEN 2010 - CAPI+CATI – vážení Alternative V8									
	Vysledky voleb	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	rovno merně
ČSSD	22,10	20,7	20,6	20,4	20,3	20,2	20,1	19,9	19,8	19,7	19,6
ODS	20,22	15,3	17,9	17,6	17,3	17,4	17,1	16,9	16,5	16,2	16,0
TOP09	16,71	16,0	17,3	17,5	17,4	17,7	17,6	17,8	17,6	17,5	17,5
KSČM	11,27	14,3	11,7	11,8	11,9	11,9	12,1	12,2	12,3	12,4	12,5
VV	10,88	10,9	11,8	11,7	11,7	11,7	11,7	11,7	11,7	11,7	11,7
KDU-ČSL	4,39	5,6	4,8	4,9	4,9	5,0	5,0	5,1	5,2	5,3	5,3
SPOZ	4,33	6,1	5,2	5,2	5,4	5,3	5,4	5,5	5,6	5,8	5,8
Suverenita	3,67	2,5	2,1	2,2	2,3	2,2	2,3	2,4	2,5	2,7	2,7
SZ	2,44	4,0	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3	4,3
Ostatní	3,99	4,5	4,3	4,3	4,4	4,3	4,4	4,4	4,4	4,4	4,5
Hodnocení chybovosti											
Suma čtverců odchylek	44,3	16,0	18,1	19,9	20,5	22,6	25,6	28,4	31,6	34,0	
Prům. abs. odchylka	1,63	1,08	1,16	1,20	1,23	1,28	1,36	1,42	1,48	1,52	
Maximální abs. odchylka	4,90	2,36	2,59	2,88	2,83	3,12	3,37	3,68	4,00	4,18	
Prům. relativní odch.	23,9%	19,0%	19,5%	20,0%	20,1%	20,5%	21,1%	21,5%	22,0%	22,4%	
Max. relativní odch.	64,7%	76,3%	76,3%	75,8%	76,3%	75,8%	75,8%	75,3%	74,7%	74,5%	
Rozdíl mezi lídry	3,5	0,9	0,9	1,1	0,9	1,1	1,2	1,4	1,6	1,7	
Chyby v kvalifikaci stran	0	0	0	0	0	0	0	0	0	0	
Kategorie kvality predikce											
Součet čtverců odchylek	3	2	2	2	2	2	2	2	2	2	
Průměrná abs. odchylka	3	2	2	2	2	2	2	2	2	3	
Maximální abs. odchylka	3	2	2	2	2	3	3	3	3	3	
Prům. relativní odch.	3	3	3	3	3	3	3	3	3	3	
Max. relativní odch.	4	4	4	4	4	4	4	4	4	4	
Rozdíl mezi lídry	4	1	1	2	1	2	2	2	2	2	
Chyby v kvalifikaci stran	1	1	1	1	1	1	1	1	1	1	
Celkový index kvality	3,0	2,1	2,1	2,3	2,1	2,4	2,4	2,4	2,4	2,6	

		DUBEN 2010 - CAPI- vážení Alternative V8									
	Vysledky voleb	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	rovno merně
ČSSD	22,10	27,7	26,5	26,5	26,5	26,5	26,5	26,5	26,5	26,5	26,5
ODS	20,22	21,5	20,0	20,1	20,1	20,1	20,1	20,2	20,2	20,3	20,3
TOP09	16,71	10,5	12,3	12,1	12,0	12,0	11,8	11,6	11,5	11,3	11,2
KSČM	11,27	12,1	10,7	10,7	10,9	10,8	10,9	11,0	11,2	11,3	11,4
VV	10,88	7,7	8,2	8,3	8,3	8,3	8,3	8,4	8,4	8,4	8,4
KDU-ČSL	4,39	6,1	5,6	5,7	5,8	5,9	5,9	6,1	6,1	6,2	6,2
SPOZ	4,33	6,3	6,1	6,1	6,2	6,1	6,2	6,2	6,3	6,3	6,3
Suverenita	3,67	2,7	2,7	2,7	2,8	2,7	2,8	2,8	2,9	3,0	3,0
SZ	2,44	3,2	4,3	4,3	4,2	4,2	4,1	4,1	4,0	3,9	3,9
Ostatní	3,99	2,2	3,6	3,5	3,4	3,4	3,2	3,1	3,0	2,9	2,8
Hodnocení chybovosti											
Suma čtverců odchylek	94,7	55,8	57,0	58,6	58,3	60,1	61,7	63,8	66,2	67,5	
Prům. abs. odchylka	2,44	1,84	1,87	1,87	1,89	1,90	1,92	1,93	1,95	1,98	
Maximální abs. odchylka	6,22	4,45	4,56	4,74	4,72	4,90	5,07	5,25	5,44	5,55	
Prům. relativní odch.	29,4%	25,7%	26,2%	26,1%	26,7%	26,6%	27,2%	27,1%	27,1%	27,4%	
Max. relativní odch.	45,4%	75,1%	74,2%	70,5%	73,3%	69,5%	68,6%	64,7%	60,8%	59,3%	
Rozdíl mezi lídry	4,4	4,6	4,6	4,5	4,5	4,5	4,4	4,4	4,4	4,3	
Chyby v kvalifikaci stran	0	0	0	0	0	0	0	0	0	0	
Kategorie kvality predikce											
Suma čtverců odchylek	4	3	3	3	3	3	3	3	3	3	
Průměrná abs. odchylka	3	3	3	3	3	3	3	3	3	3	
Maximální abs. odchylka	4	3	3	3	3	3	4	4	4	4	
Prům. relativní odch.	4	4	4	4	4	4	4	4	4	4	
Max. relativní odch.	3	4	4	4	4	4	4	4	4	3	
Rozdíl mezi lídry	5	5	5	5	5	5	5	5	5	5	
Chyby v kvalifikaci stran	1	1	1	1	1	1	1	1	1	1	
Celkový index kvality	3,4	3,3	3,3	3,3	3,3	3,3	3,4	3,4	3,4	3,3	

		DUBEN-KVĚTEN 2010 – CAPI – vážení Alternative V8									
	Výsledky voleb	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	rovnomerně
ČSSD	22,10	24,5	24,4	24,1	24,0	23,9	23,7	23,4	23,3	23,1	23,0
ODS	20,22	18,8	17,5	17,5	17,6	17,6	17,6	17,7	17,7	17,8	17,8
TOP09	16,71	13,4	14,9	14,7	14,5	14,6	14,4	14,2	14,0	13,8	13,7
KSČM	11,27	14,2	12,1	12,1	12,2	12,2	12,3	12,4	12,5	12,6	12,7
VV	10,88	6,9	7,6	7,8	7,8	7,9	7,9	8,1	8,1	8,2	8,2
KDU-ČSL	4,39	6,4	5,5	5,6	5,7	5,7	5,9	6,0	6,1	6,3	6,3
SPOZ	4,33	4,4	4,7	4,7	4,7	4,7	4,7	4,7	4,7	4,8	4,8
Suverenita	3,67	2,4	2,4	2,5	2,6	2,5	2,7	2,7	2,9	3,0	3,1
SZ	2,44	4,2	5,3	5,2	5,2	5,2	5,1	5,1	5,0	5,0	4,9
Ostatní	3,99	4,8	5,8	5,7	5,7	5,7	5,7	5,7	5,6	5,5	5,5
Hodnocení chybovosti											
Suma čtverců odchylek		52,4	41,6	39,7	39,3	38,0	37,9	36,7	37,0	37,6	37,7
Prům. abs. odchylka		1,99	1,82	1,80	1,80	1,78	1,77	1,75	1,75	1,75	1,74
Maximální abs. odchylka		3,93	3,27	3,12	3,09	2,97	2,94	2,79	2,76	2,92	3,03
Prům. relativní odch.		27,4%	30,0%	29,8%	29,5%	29,5%	29,2%	29,0%	28,6%	28,3%	28,1%
Max. relativní odch.		72,1%	117,4%	114,8%	112,8%	112,2%	110,1%	107,4%	105,3%	103,1%	101,6%
Rozdíl mezi lídry		3,8	5,0	4,7	4,5	4,4	4,2	3,9	3,7	3,5	3,3
Chyby v kvalifikaci stran		0	0	0	0	0	0	0	0	0	0
Kategorie kvality predikce											
Suma čtverců odchylek		3	3	2	2	2	2	2	2	2	2
Průměrná abs. odchylka		3	3	3	3	3	3	3	3	3	3
Maximální abs. odchylka		3	3	3	3	2	2	2	2	2	3
Prům. relativní odch.		4	4	4	4	4	4	4	4	4	4
Max. relativní odch.		4	6	6	6	6	6	6	6	6	6
Rozdíl mezi lídry		4	6	5	5	5	5	4	4	4	4
Chyby v kvalifikaci stran		1	1	1	1	1	1	1	1	1	1
Celkový index kvality		3,1	3,7	3,4	3,4	3,3	3,3	3,1	3,1	3,1	3,3

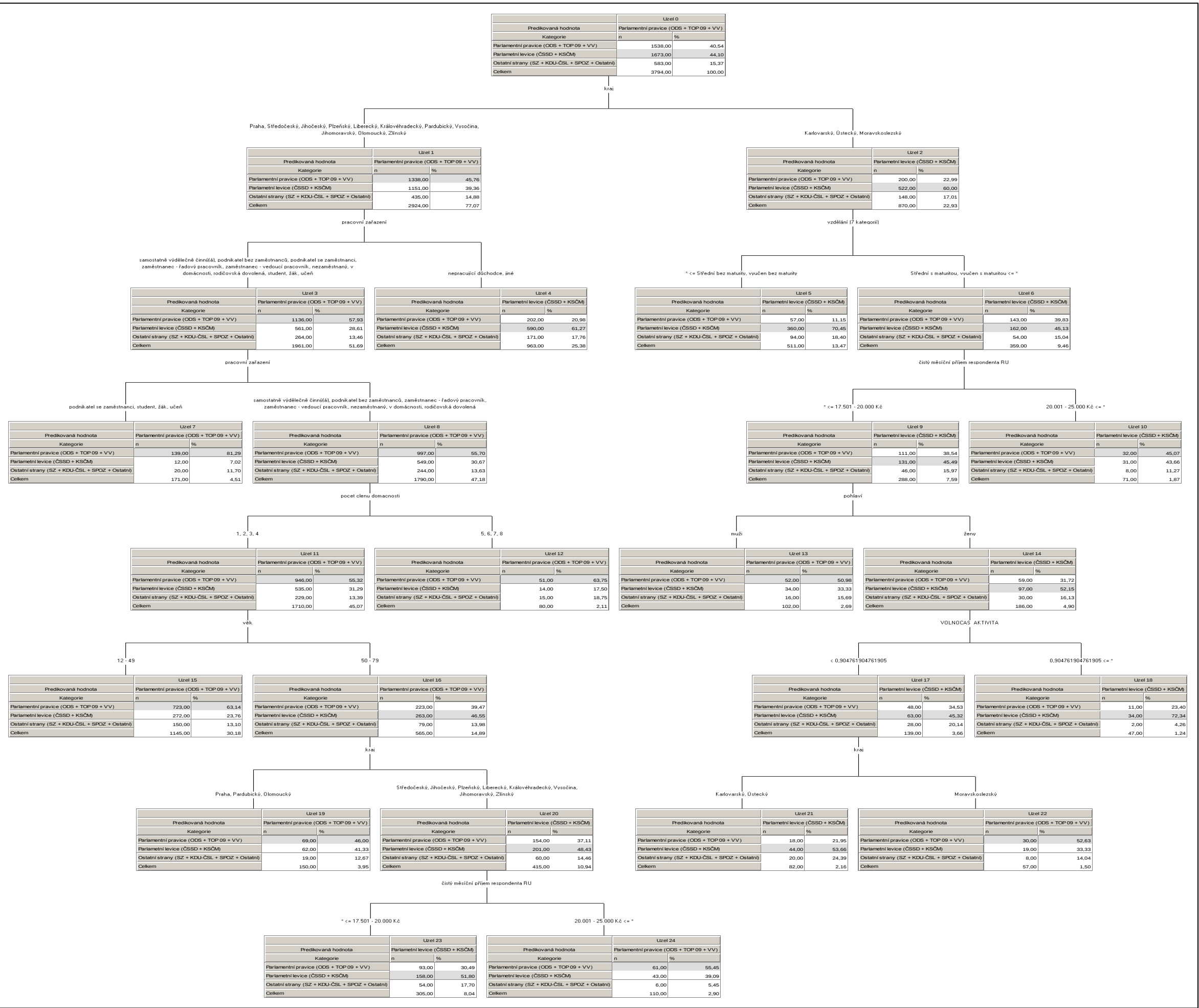
		KVĚTEN 2010 – CAPI – vážení Alternative V8									
	Vysledky voleb	bez	10-0-0	9-1-0	8-1-1	8-2-0	7-2-1	6-3-1	5-3-2	4-3-3	rovno merně
ČSSD	22,10	20,1	20,8	20,6	20,3	20,3	20,1	19,9	19,6	19,4	19,3
ODS	20,22	13,7	14,5	14,4	14,2	14,3	14,2	14,1	13,9	13,7	13,7
TOP09	16,71	19,0	19,2	19,0	18,9	18,8	18,7	18,5	18,4	18,3	18,2
KSČM	11,27	13,7	12,5	12,6	12,6	12,6	12,6	12,7	12,7	12,7	12,7
VV	10,88	11,1	11,6	11,6	11,6	11,7	11,7	11,7	11,7	11,7	11,7
KDU-ČSL	4,39	5,5	4,7	4,9	5,0	5,0	5,2	5,3	5,5	5,6	5,7
SPOZ	4,33	5,9	5,2	5,3	5,5	5,5	5,6	5,8	5,9	6,0	6,1
Suverenita	3,67	2,7	2,6	2,7	2,8	2,7	2,9	3,0	3,2	3,4	3,4
SZ	2,44	4,9	5,9	5,8	5,8	5,8	5,7	5,7	5,6	5,6	5,6
Ostatní	3,99	3,3	3,0	3,1	3,2	3,2	3,3	3,4	3,5	3,6	3,6
Hodnocení chybovosti											
Suma čtverců odchylek		68,5	58,0	58,4	60,3	59,2	61,4	62,7	65,6	69,0	70,8
Prům. abs. odchylka		2,03	1,83	1,85	1,88	1,88	1,90	1,93	1,95	1,98	2,00
Maximální abs. odchylka		6,48	5,76	5,82	5,99	5,88	6,05	6,12	6,30	6,48	6,55
Prům. relativní odch.		28,7%	29,0%	29,2%	29,0%	29,3%	29,1%	29,3%	29,1%	28,9%	28,9%
Max. relativní odch.		102,7%	141,0%	139,6%	137,1%	138,1%	135,6%	134,1%	131,5%	128,9%	127,7 %
Rozdíl mezi lídry		4,5	4,4	4,3	4,2	4,1	4,1	3,9	3,8	3,8	3,7
Chyby v kvalifikaci stran		0	0	0	0	0	0	0	0	0	0
Kategorie kvality predikce											
Suma čtverců odchylek		3	3	3	3	3	3	3	3	3	4
Průměrná abs. odchylka		3	3	3	3	3	3	3	3	3	3
Maximální abs. odchylka		4	4	4	4	4	4	4	4	4	4
Prům. relativní odch.		4	4	4	4	4	4	4	4	4	4
Max. relativní odch.		6	7	7	7	7	7	7	7	7	7
Rozdíl mezi lídry		5	5	5	5	5	5	4	4	4	4
Chyby v kvalifikaci stran		1	1	1	1	1	1	1	1	1	1
Celkový index kvality		3,7	3,9	3,9	3,9	3,9	3,9	3,7	3,7	3,7	3,9

Příloha 5: Popis všech testovaných vážení

Název vážení	Kategorie vážení (počet hodnot)
<i>Referenční vážení</i>	
Socio_1	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 2D - pohlaví X věk, pohlaví X kraj, region (8) X VMB
<i>Alternativní vážení bez minulého volebního chování</i>	
Alternative_1	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10), počet členů d. (4)
Alternative_2	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), počet členů d. (4)
Alternative_3	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - pracovní zařazení (8), příjem na člena d.(10), počet členů d. (4)
Alternative_4	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10)
Alternative_5	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D - pohlaví X věk, pohlaví X kraj
Alternative_6	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D - pohlaví X věk, pohlaví X kraj, region X VMB
Alternative_7	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D - věk x region, věk x pohlaví, věk x vzdělání
Alternative_8	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 2D - věk (4) x oblast (3), věk x pohlaví, věk x vzdělání , region X VMB
<i>Vážení s minulým volebním chováním</i>	
Sociovolby	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - chování ve volbách 2006 (9) 2D - pohlaví X věk, pohlaví X kraj, region X VMB
Alternative_V1	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10), počet členů d. (4), 1D - chování ve volbách 2006 (9)
Alternative_V6	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 1D - chování ve volbách 2006 (9) 2D - pohlaví X věk, pohlaví X kraj, region X VMB
Alternative_V8	1D - pohlaví (2), vzdělání (2), kraj (14), VMB (5), věk (7) 1D - internet (4), pracovní zařazení (8), příjem na člena d.(10) 1D - chování ve volbách 2006 (9) 2D - věk (4) x oblast (3), věk x pohlaví, věk x vzdělání , region X VMB

POZN: Znak 1D znamená vážení podle jednotlivé proměnné, znak 2D podle sdružených kategorií dvou proměnných. Čísla v závorkách uvádějí počet kategorií proměnné. Čísla v závorkách u 2D vážení jsou uvedena pouze v případě, že se počet kategorií s uproměnné liší od počtu jejích kategorií v 1D vážení.

Příloha 6. Řešení pomocí klasifikačních a regresních stromů – model s nejvyšší predikční schopností



Nezávislé proměnné (prediktory)	Typ (počet hodnot)
Věk	Kardinální
Kraj NUTS3	Nominální (7)
Vzdělání	Ordinální (7)
Pohlaví	Nominální (2)
VMB	Ordinální (5)
Pracovní zařazení	Nominální (9)
Čistý měsíční příjem domácnosti	Ordinální (16)*
Čistý měsíční příjem respondenta	Ordinální (16)*
Čistý měsíční příjem domácnosti na člena	Kardinální (nekategoriz.proměnná)
Počet členů domácnosti	Kardinální
Intenzita používání internetu	Ordinální (6)
Preference volnočasových aktivit	Kardinální (nekategoriz.proměnná)

*Možnosti „nevím“ a „odmítl odpovědět“ byl pro udržení ordinálního charakteru překódovány na misval.

Závislá prom. (predikovaná)	Hodnoty závislé proměnné
Typ volené strany	Parlamentní pravice (ODS, TOP09, VV) Parlamentní levice (ČSSD, KSČM) Ostatní strany

Typ tvorby modelu: globální

Typ omezení	Hodnota
Max. hloubka stromu	10 úrovní
Max. počet synů ve štěpení	6
Min. počet respondentů v synu	40
Min. počet respondentů v uzlu	80

Hodnocení klasifikace:

Správnost klasifikace %	Počet klasifikovaných	Počet správně rozklasifikovaných
61,73	3794	2342

