

## Model predikce výsledků voleb – alternativní přístup k odhadům volebních výsledků ve volebních obvodech

TOMÁŠ KOSTELECKÝ\*

Sociologický ústav AV ČR, Praha

### A Model for Predicting Election Results – An Alternative Approach to Estimating Election Results in Electoral Districts

**Abstract:** To provide reliable pre-election forecasts it is necessary to estimate the election results as precisely as possible in individual electoral districts. Such estimates are traditionally based on a series of pre-election surveys in the individual electoral districts. The article presents an alternative method of making pre-election estimates, which combines the results of national representative surveys and information on the spatial distribution of party support from the previous election, and the results of the model-based estimates are compared with traditional survey-based estimates. The article demonstrates the usefulness of the model-based approach and discusses the conditions that either support or impede the quality of model-based estimates.

*Sociologický časopis/Czech Sociological Review, 2005, Vol. 41, No. 1: 79–101*

### Úvod

Jedním z nejdůležitějších, nejvíce sledovaných a zároveň nejcitlivějších typů výzkumů veřejného mínění jsou výzkumy veřejného mínění zaměřené na sledování stranických preferencí obyvatel, které měří oblibu jednotlivých politických stran, hnutí a koalic v populaci. Výzkumy stranických preferencí jsou do jisté míry „zkušebními volbami“, které se na reprezentativním vzorku populace provádějí „nanečisto“, v období mezi skutečnými volbami. Souvislost mezi měřením stranických preferencí a volbami je patrná už ze znění otázek, které jsou respondentům pokládány. Jejich formulace, např. „Pokud byste šel(šla) volit, řekněte mi, prosím, kterou stranu byste volil(a)?“ nebo „Představte si, že by příští týden byly volby do Poslanecké sněmovny. Kterou stranu byste volil(a)?“, mají respondentům navozovat situaci skutečných voleb. V mezivolebním období jsou takto získané údaje o popularitě stran nezastupitelné, mimo jiné proto, že představují jeden z mála věrohodných způsobů, jak mohou voliči reagovat na činnost politických elit a jak mohou naopak politici získávat informace o reakcích veřejnosti na své konání. V ideálním případě poskytují výzkumy stranických preferencí věrohodné informace o rozložení politických sil a stabilitě či změnách politických nálad ve společnosti, zvláště pokud jsou výzkumy prováděny stej-

---

\* Veškerou korespondenci posílejte na adresu: RNDr. Tomáš Kostecký, CSc., Sociologický ústav AV ČR, Jilská 1, 110 00 Praha 1, e-mail: tomas.kostecky@soc.cas.cz.

nou metodou v pravidelných intervalech a dlouhých časových řadách. Vzniká tak určitá zpětná vazba mezi politiky a veřejností, která umožňuje politikům korigovat svoje konání již v období mezi volbami a celé společnosti pak dává možnost vyhnout se velkým politickým otřesům spojeným s náhlými změnami mocenských poměrů a politického klimatu, ke kterým by mohlo docházet v případě, že jedinou informací o vztahu mezi voliči a politickými stranami by byly výhradně volby, které se zpravidla opakují v několikaletém intervalu.<sup>1</sup>

Simulovat situaci skutečných voleb výzkumem stranických preferencí se ovšem nemůže nikdy zcela beze zbytku podařit, protože se výzkum veřejného mínění v několika podstatných ohledech od skutečných voleb liší. Respondenti vědí, že nejde o skutečné volby, a že tudíž jejich „volba“, tj. odpověď na otázku ve výzkumu, nemůže mít žádné bezprostřední politické důsledky. To jim dává svobodu být méně odpovědní, např. dát svou preferenci „z protestu“ některé z extremistických stran, nebo preferovat některou opoziční stranu, aniž by ji hodlali skutečně volit, a vyslat tak signál vládě o nespokojenosti s její momentální politikou... Jako v každém výzkumu veřejného mínění se někteří oslovení odmítnou zúčastnit nebo už oslovení respondenti odmítnou odpovídat na tento typ otázek. Výsledný soubor respondentů, kteří na otázky po stranických preferencích odpověděli, nemusí být totožný s lidmi, kteří se voleb skutečně zúčastní. Další odlišnost mezi výzkumem veřejného mínění a volbami spočívá v tom, že respondentům (včetně těch, kteří jsou určitě rozhodnutí se voleb zúčastnit) je zpravidla dovoleno odpovědět „nevím“, zatímco u účastníků skutečných voleb tato možnost v podstatě neexistuje, pomíneme-li jen málo pravděpodobnou možnost, že by se někdo zúčastnil voleb jen proto, aby do urny na znamení své nerozhodnutosti úmyslně vhodil prázdnou úřední obálku.

## 1. Stranické preference, volební preference, volební prognózy

V období těsně předcházejícímu skutečným volbám přestává měření stranických preferencí plnit svoji informativní úlohu „nastaveného zrcadla“ politickým náladám ve společnosti. S blížícími se volbami už nestačí pouhé *měření stranických preferencí* a začíná růst význam výzkumu veřejného mínění jako nástroje, který potenciálně poskytuje možnost *odhadovat na základě zjištěných stranických preferencí skutečné výsledky voleb*. Skutečnost, že měření stranických preferencí není totéž jako odhad výsledků voleb vytvořený na základě zjištěných stranických preferencí, není veřejnosti příliš známa. Agentury, které se v České republice předvolebními výzkumy systematicky zabývají, si tuto skutečnost uvědomují, a snaží se proto při zveřejňování svých výsledků odlišovat *stranické preference* a *volební preference* resp. *volební prognózy*. Zatímco termín *stranické preference* se všeobecně vžil a je víceméně jednotně chápán jako „procento respondentů, kteří preferují jednotlivé politické strany“, jak je

---

<sup>1</sup> Pro tento okamžik pomíneme možnost, že samo provádění výzkumů veřejného mínění se za určitých okolností může stát nástrojem cílené manipulace s veřejným míněním a nikoliv nástrojem poskytujícím pravdivý obraz o náladách veřejnosti.

zachytí výzkumy veřejného mínění, dva zbylé uvedené termíny už evidentně tak úspěšné nebyly. Termínem *volební preference* je označován výsledek odpovědi na otázku „Kterou stranu byste volil?“ pouze těch respondentů, kteří uvedli, že by šli k volbám (tedy ne všech respondentů, jako je tomu u stranických preferencí). V obdobném smyslu je používán termín *volební prognóza*, která je udávána v % platných hlasů. Jedná se o matematický přepočít, jehož cílem je dát představu o reálném výsledku voleb, tedy bez nerozhodnutých voličů. Bere se v úvahu i míra připravenosti příznivců stran zúčastnit se voleb.

## 2. Předvolební odhady volebních výsledků

Ačkoliv jsou volební preference či volební prognózy blíže výsledkům voleb než běžně publikované stranické preference, zůstávají v jistém smyslu jen na půli cesty ke skutečným odhadům volebních výsledků. Zpřesňují totiž pouze naši představu o pravděpodobném procentu hlasů odevzdaných ve volbách kandidujícím stranám, zatímco výsledkem voleb, o něž primárně v předvolebních odhadech jde, je složení budoucího parlamentu, tedy *rozdělení mandátů mezi kandidující strany*. Ačkoliv je dobře udělaná volební prognóza předpokladem pro provedení spolehlivého odhadu rozdělení mandátů v parlamentu, je to podmínka nutná, nikoliv však postačující. Rozdělení křesel v parlamentu mezi kandidující strany závisí kromě samotného podílu hlasů, které dostanou kandidující strany od voličů, na volebním systému, který je při daných volbách použit, neboť různé volební systémy mohou vést i při stejných výsledcích hlasování k radikálně odlišným výsledkům ve složení parlamentu [Lebeda 2000; Kostecký 2000a]. Různé volební systémy ovšem nejenom přepočítávají odevzdané hlasy na parlamentní mandáty odlišným způsobem, ale jsou i různě citlivé na kvalitu a formu zjišťovaných stranických preferencí, které slouží jako vstupní data při odhadech povolebního rozdělení mandátů v parlamentu. Relativně nejjednodušší situace nastane v případě, že je v zemi používán systém poměrného zastoupení, v němž je celá země jediným volebním obvodem, jako je tomu například v Izraeli [Kostecký, 2000b]. V tomto případě je možné jako vstupní data pro výpočet rozdělení mandátů v parlamentu použít výsledky výzkumu stranických preferencí, provedené na vzorku populace reprezentativním pro celou populaci (tj. ony stranické preference, které jsou běžně publikovány v mezivolebním období), protože i mandáty se v tomto případě přidělují na celostátní úrovni. Volební prognóza odvozená od těchto stranických preferencí se jednoduše použije namísto skutečných (před volbami samozřejmě dosud neznámých) čísel udávajících podíl hlasů pro jednotlivé strany, na takto získaná čísla se aplikuje procedura daná příslušným volebním zákonem a výsledkem je relativně spolehlivý odhad rozdělení mandátů v budoucím parlamentu. Komplikace ovšem spočívá v tom, že většina zemí, včetně České republiky, používá volební systémy, v nichž je země rozdělena na menší či větší počet volebních obvodů. Mandáty jsou v tomto případě přidělovány na úrovni jednotlivých obvodů, zpravidla v počtu přímo úměrném počtu obyvatel (eventuálně počtu voličů nebo počtu odevzdaných hlasů) v jednotlivých volebních obvodech. Pro

tože jsou mandáty přidělovány v jednotlivých volebních obvodech, je nutné i při vytváření odhadu rozdělení mandátů postupovat tak, jak se postupuje při skutečných volbách. Tedy: nejprve odhadneme výsledky hlasování v jednotlivých volebních obvodech, poté v každém volebním obvodě zvlášť rozdělíme pomocí procedury dané volebním zákonem mandáty mezi strany a nakonec počty mandátů získané stranami v jednotlivých volebních obvodech sečteme, a získáme tak konečný výsledek. Teoreticky je tato operace jednoduchá, neboť neznamená v podstatě nic jiného než opakování stejné procedury v každém z volebních obvodů, kterých je v českém volebním systému do Poslanecké sněmovny 14,<sup>2</sup> při volbách do třetiny Senátu 27.<sup>3</sup>

### *2.1 Tradiční metody odhadu volebních výsledků – výběrová šetření*

Tradičně jsou předvolební odhady volebních výsledků založeny na výzkumech stranických preferencí, které jsou prováděny tak, aby jejich výsledky byly reprezentativní pro populaci jednotlivých volebních obvodů. Zásadní praktická komplikace však spočívá v relativně velkém počtu volebních obvodů, které je nutno takovým výzkumem pokrýt. Při běžných výzkumech stranických preferencí, které jsou reprezentativní pro populaci celé České republiky, se pracuje se vzorkem cca 1000 respondentů. Při této velikosti výběrového vzorku je nutné počítat s výběrovou chybou cca  $\pm 3$  %. V předvolební době se zpravidla dělá výzkum na větším vzorku respondentů – 2000 až 3000 dotázaných – což ovšem sníží velikost výběrové chyby jen na zhruba  $\pm 2$  %. Cena takového výzkumu se podle typu výběru pohybuje minimálně v řádu statisiců korun. Aby bylo možno získat stranické preference pro všechny volební obvody s výběrovou chybou maximálně  $\pm 3$  %, muselo by být v každém volebním obvodě dotázáno přes 1000 respondentů. Celkový počet respondentů ve všech obvodech by tak dosáhl zhruba 14 000 v případě voleb do Poslanecké sněmovny, respektive kolem 27 000 u voleb do třetiny Senátu. Celková částka vynaložená na tak velké šetření by byla o řád vyšší, než stojí běžný výzkum. Společnosti, které se snaží o provedení věrohodného odhadu výsledku voleb, stojí před dilematem. Snaha o provedení série relativně přesných výzkumů na velkých vzorcích v jednotlivých volebních obvodech je extrémně drahá, naráží se při ní navíc na technické a organizační problémy, jako je nedostatek vlastních tazatelů, nutnost spojování více tazatelských sí-

---

<sup>2</sup> Od roku 1990 do roku 2002 bylo ve volbách do Poslanecké sněmovny používáno 8 volebních obvodů, které se kryly s tehdy existujícími administrativními kraji. Reforma volebního zákona schválena parlamentem v roce 2000 předpokládala vytvoření 35 speciálních volebních obvodů pro volby do Poslanecké sněmovny [Kostelecký, 2000a]. Poté, co Ústavní soud vyhověl stížnosti prezidenta a velkou část tohoto volebního zákona (včetně ustanovení o počtu volebních obvodů) zrušil coby odporující platné ústavě, bylo přijato kompromisní řešení, které stanovilo počet obvodů na 14. Tyto volební obvody se kryjí s novými kraji zavedenými administrativní reformou v roce 2000.

<sup>3</sup> Volby do Senátu se konají v 81 volebních obvodech speciálně pro tento účel vytvořených. První volby do Senátu se v roce 1996 konaly ve všech 81 volebních obvodech. Od roku 1998 se každé dva roky konají volby v jedné třetině volebních obvodů.

tí pro tento speciální výzkum, obtížnost efektivní kontroly tazatelů a s tím související riziko poklesu kvality celého výzkumu. Zmenšování počtu respondentů pod 1000 v jednom volebním obvodu s sebou zase nese riziko zvětšování výběrové chyby, a tudíž malé spolehlivosti zjištěných výsledků. Výsledky uskutečněných předvolebních výzkumů, které byly provedeny v jednotlivých volebních obvodech pro různé typy voleb v období 1998 až 2002, ukazují, že v praxi je zpravidla hledán kompromis mezi přesností výzkumu a jeho financovatelností. Před volbami do Poslanecké sněmovny v květnu roku 2002 byly společnostmi STEM a TNSF provedeny nezávisle na sobě předvolební výzkumy ve všech 14 volebních krajích České republiky. Lebeda, který srovnával předvolební výzkumy volebních preferencí z roku 2002 z produkce tří etablovaných společností dlouhodobě se zabývajících výzkumy stranických preferencí v České republice – STEM, TNSF a CVVM, uvádí, že agentura STEM se dotazovala v průměru 600 respondentů v každém kraji, agentura TNSF v průměru kolem 285 respondentů.<sup>4</sup> Nízký počet respondentů v jednotlivých krajích byl zřejmě jednou z významných příčin, proč byly krajské volební prognózy podstatně méně kvalitní než klasické průzkumy reprezentativní pro celou Českou republiku [Lebeda 2003].<sup>5</sup> Před prvními volbami do zastupitelstev krajů v době po pádu komunismu, v listopadu 2000, se dvě agentury – SC&C a STEM – spojily do konsorcia, aby mohly provést velký předvolební výzkum, při kterém bylo osloveno celkem 11 917 respondentů, v každém kraji zhruba 800 až 1000 občanů [FSC&C, STEM 2000b]. Ve stejném roce byly před volbami do třetiny Senátu, při kterých je na rozdíl od většiny ostatních voleb v České republice používán dvoukolový většinový systém s jednomandátovými obvody, provedeny týž konsorciem výzkumy v jednotlivých volebních obvodech. Před prvním kolem senátních voleb byly uskutečněny dvě po sobě se opakující série výzkumů jen v 9 vybraných mediálně neatraktivnějších volebních obvodech. Jejich výsledky byly postupně uveřejňovány v deníku Mladá fronta Dnes ve dnech 18. 10. 2000 až 7. 11. 2000. Udávané počty respondentů v jednotlivých obvodech se pohybovaly mezi 435 a 519. Před druhým kolem voleb byl uskutečněn výzkum ve všech 26 obvodech, ve kterých ještě nebylo o vítězi rozhodnuto.<sup>6</sup> Celkem bylo v tomto výzkumu dotazováno 10 058 respondentů [FSC&C, STEM 2000a], což znamená, že bylo v jednotlivých volebních obvodech získáno v průměru kolem 390 dotazníků.

<sup>4</sup> Agentura CVVM krajské výzkumy neprováděla a omezila se na výzkumy reprezentativní pro populaci celé České republiky.

<sup>5</sup> Další možné příčiny relativně nízké kvality krajských volebních prognóz vidí Lebeda v přetíženosti tazatelských sítí extrémně velkým množstvím práce koncentrované do malého časového úseku, obtížné kontrolovatelnosti práce tazatelů a ve skutečnosti, že tazatelé dostávali v těchto předvolebních výzkumech za jeden uskutečněný rozhovor menší odměnu, než je obvyklé.

<sup>6</sup> V jednom z 27 volebních obvodů získal kandidát v prvním kole více než 50 % hlasů, takže se druhé kolo již nekonalo.

## 2.2 Alternativní metody odhadu volebních výsledků – modely

Snaha vyhnout se obtížnostem spojeným s prováděním volebních odhadů tradiční metodou – sérií výběrových šetření v jednotlivých volebních obvodech – nás přivedla na nápad pokusit se k vytváření odhadů volebních výsledků v jednotlivých volebních obvodech použít zcela jiný přístup. Tento přístup vychází z myšlenky, že za určitých okolností je možné odhadnout volební výsledky v jednotlivých volebních obvodech za použití určitého matematického modelu, který bude výsledky za jednotlivé volební obvody predikovat na základě znalosti výsledků aktuálního výzkumu stranických preferencí na celostátním vzorku populace (o cca 1000–2000 respondentech) a analýzy výsledků předchozích voleb, která nám dodá informace o distribuci stranických preferencí mezi jednotlivé volební obvody. Pokud by byla prostorová distribuce volební podpory politickým stranám naprosto neměnná v čase, tedy prostorový obraz jejich volební podpory zachycený ve volebních mapách by zůstal stejný,<sup>7</sup> metoda by nutně musela být velmi přesná, neboť by stranické preference, resp. volební prognózy pro jednotlivé politické strany zjištěné na úrovni celé republiky, pouze „přeočítala“ podle vzorců známých z minulých voleb do úrovně jednotlivých volebních obvodů. Ačkoliv je předpoklad naprosté neměnnosti prostorové distribuce voličské podpory stranám v praxi samozřejmě nerealistický, empirické práce analyzující volební mapy v České republice [Jehlička, Sýkora 1991; Kostecký 2001, 2003; Sokol 2003; Jehlička, Kostecký 2003] shodně potvrzují, že charakter regionální diferenciace volebních výsledků politických stran má neobvykle silnou tendenci přetrvávat v čase.

Přispívá k tomu skutečnost, že velká část voličů svoje stranické preference nemění, míra voličské loajality má navíc tendenci se zvyšovat [Plecitá 2003]. Zároveň je známo, že se obyvatelé České republiky relativně málo stěhují, přičemž intenzita migrace se v posledních desetiletích ještě dále snižuje [Čermák 1999], takže změny ve složení voličů v jednotlivých volebních obvodech způsobené migrací jsou zanedbatelné. Pokud se obliba nějaké strany zvyšuje či snižuje, celková změna ve volební podpoře strany má tendenci probíhat „proporcionálně“ (vzájemné poměry podílu získaných hlasů stranou v jednotlivých volebních obvodech jsou neobvykle stabilní veličinou), proto jsou volební mapy zachycující výsledky stran v jednotlivých parlamentních volbách u většiny politických stran značně podobné. Lze ukázat [Kostecký 2001], že při snaze „předpovědět“ podíly hlasů pro strany v jednotlivých okresech regresními modely jsou nejlepším prediktorem volební výsledky téže strany ve volbách předešlých. Je sice pravda, že se volební mapy vždy do jisté míry změní (každé nové

---

<sup>7</sup> Upozorňujeme, že je třeba rozlišovat mezi neměnností volebních preferencí jako takových a neměnností prostorové distribuce volební podpory. Neměnnost prostorové distribuce volební podpory nevyklučuje změny v popularitě jednotlivých stran, ale říká, že např. procento hlasů získaných ODS v Praze by bylo vždy o polovinu vyšší než průměrné procento hlasů, které tato strana získala v celé České republice, podíl hlasů pro ODS v Ostravě by byl vždy o třetinu nižší než celorepublikový průměr atd. Neměnné jsou vzájemné poměry mezi velikostí volební podpory v jednotlivých volebních obvodech, nikoliv volební zisky stran.

hlasování je reakcí na jinou situaci), počet dalších faktorů, které modifikují predikční sílu hlavního faktoru (volební výsledky v minulých volbách) je však za normální situace relativně omezený. Nejdůležitějšími doplňkovými faktory ovlivňujícími, jak budou vypadat nové volební mapy, jsou: regionálně se měnící účast ve volbách, reakce voličů na regionálně odlišný ekonomický vývoj [Kostecký 2001] a „přesun“ populárních osobností z jednoho volebního obvodu do druhého [Kostecký 1993].

### **3. Srovnání tradičních metod výběrových šetření a alternativních „modelových“ metod odhadu volebních výsledků**

#### *3.1 Teoretická východiska*

Teoreticky je možné dělat odhady volebních výsledků oběma způsoby – jak za použití tradičních metod výběrových šetření, tak s využitím matematických modelů. V tomto smyslu jsou obě metody v zásadě rovnocenné a jde o to posoudit, která z nich poskytuje přesnější a spolehlivější výsledky a za jakých podmínek. Z hlediska praktického však metody zcela rovnocenné nejsou, protože zatímco metoda výběrových šetření pro svou aplikaci žádné modely nepotřebuje, „modelem“ vytvářené odhady volebních výsledků existenci výběrových šetření na celostátní úrovni přepokládají a jejich výstupy používají jako jeden z klíčových vstupů do modelu. Při porovnávání obou metod je třeba mít neustále na paměti, že otázka nestojí: „buď výběrová šetření nebo modely“, ale spíše: „buď série výběrových šetření v jednotlivých volebních obvodech nebo standardní reprezentativní šetření na celonárodním vzorku + modely“. Ať tak či onak, předvolební výzkumy stranických preferencí prováděné metodou výběrových šetření na vzorku reprezentujícím národní populaci jsou naprosto nezastupitelné, protože jenom ony měří momentální popularitu politických stran a její změny v čase. Z toho důvodu také nemůže „modelová“ metoda výběrová šetření beze zbytku nahradit, může ovšem teoreticky nahradit rozsáhlé a drahé série výběrových šetření prováděné paralelně v jednotlivých volebních obvodech.

#### *3.2 Silné a slabé stránky tradičních a alternativních metod, formulace hypotéz*

Obě porovnávané metody odhadů volebních výsledků mají některé silné a některé slabé stránky, jejich úspěšnost je závislá na splnění určitých podmínek. Největší komparativní výhodou metody odhadů volebních výsledků založené na provedení výběrových šetření v jednotlivých volebních obvodech je její schopnost zachytit nečekané změny v politických preferencích obyvatel, zvláště ty, které jsou unikátní v prostoru a v čase, a odhalit regionálně specifický vývoj politických preferencí. Změní-li například některý z populárních politiků volební obvod, v němž kandiduje, nebo postaví-li jiná politická strana do voleb známou, oblíbenou a respektovanou osobnost, která se předchozích voleb nezúčastnila, dá se předpokládat, že tím vyvolané regionálně specifické změny stranických preferencí zachytí výběrová šetření

prováděná v jednotlivých volebních obvodech lépe než model, který má tendenci tyto regionálně specifické změny ignorovat a naopak předpokládá neměnnost regionálního obrazu volební podpory. Některé regionálně specifické přesuny volebních preferencí je sice možné do modelových odhadů zakomponovat, ale ne všechny, a vždy jen do určité míry. Je například možné odhadnout, jaký vliv na mapu volební podpory může mít přesunutí stranického leadera z jednoho volebního obvodu do jiného, protože k těmto věcem již v minulosti docházelo a existuje tak určitý precedens,<sup>8</sup> z něhož lze při modelování vycházet. Obdobně lze do modelu zahrnout předpokládanou změnu stranických preferencí ve volebním kraji, v němž v období od posledních voleb například podstatně vzrostla nezaměstnanost, víme-li z minulých zkušeností, že míra nezaměstnanosti, resp. její růst, je faktorem, který podstatně ovlivňuje stranické preference v regionech [Kostelecký 2003] a dokážeme odhadnout „směr“ jejího vlivu na stranické preference, a do jisté míry kvantifikovat i intenzitu těchto změn. Metody odhadů volebních výsledků založené na výběrových šetřeních ve volebních obvodech jsou samozřejmě také vhodnější v okamžiku, kdy dochází k závažným institucionálním změnám ve spektru politických stran – například k dělení nebo slučování stávajících politických stran, zakládání zcela nových stran (zvláště pokud se v nich angažují vlivní a známí lidé, kteří dosud stáli mimo politiku). Modelová řešení mohou být sice za určitých okolností použita i v tomto případě, zvláště pokud existuje dostatek na individuální úrovni získaných informací o tom, jak se v takové situaci pravděpodobně zachovají bývalí voliči rozpadnutších se nebo sloučených stran. Podobně jsou předvolební odhady postavené na výběrových šetřeních ve volebních obvodech ve výhodě v případě, že je jedná o první volby do nově vzniklých institucí. Za této situace samozřejmě modelová řešení narážejí na problém, že nemohou pracovat s informacemi o tom, „jak se voliči chovali minule“, a musí chování voličů do jisté míry odhadovat na základě analogií s jiným typem v minulosti proběhlých voleb, což může výsledek zkreslit. Obecně lze říci, že s použitím modelových odhadů předvolebních výsledků může být problém v období politické a institucionální nestability, zatímco v dobách „normální politiky“ se dá očekávat, že modelová řešení mohou být použita s úspěchem.

Podle toho, co bylo řečeno výše, by se nedalo očekávat, že by mohla být modelová řešení k užítku pro předvolební odhady výsledků voleb v zemi, jakou je Česká republika v období velkých politických změn spojených s procesem post-komunistické transformace. Je však na místě také připomenout komparativní výhody modelových odhadů volebních výsledků, z nichž přinejmenším některé jsou také dány specifickými podmínkami v post-komunistické transformující se zemi. Všechny vý-

---

<sup>8</sup> V letech 1990 až 1996 kandidoval tehdejší předseda ODS Václav Klaus ve volebním obvodu na severní Moravě, zatímco v roce 1998 a 2002 kandidoval v Praze. Jeho přesun ze severomoravské kandidátky na pražskou byl následován relativním poklesem volební podpory ODS na severní Moravě. Podobně poznamenal volební mapu ČSSD přesun tehdejšího předsedy Miloše Zemana z jihočeské kandidátky na severomoravskou, ke které došlo mezi volbami 1992 a 1996. O vlivu populárních osobností na volební podporu ve volebních krajích, kde kandidují, podrobněji pojednává Kostelecký [1993].



zkumy stranických preferencí se v posledních letech potýkají se dvěma metodologickými problémy – vysokým podílem lidí, kteří se odmítají těchto výzkumů zúčastnit, a také neochotou oslovených odpovídat pravdivě na otázky po jejich stranických preferencích. Stranické preference respondentů výzkumu a voličské preference skutečných voličů se mohou značně lišit, zamlčení skutečných stranických preferencí či dokonce uvedení jiných než skutečných preferencí poskytuje výzkumníkům zavádějící informace. Ačkoliv mohou výzkumné agentury vycházet ze svých zkušeností z minulých voleb a při vytváření volebních prognóz obě tyto chyby do jisté míry „korigovat“, oba zmíněné problémy přesto značně komplikují přesnost odhadů volebních výsledků. S problémem zmlčených, eventuálně nepravdivých stranických preferencí zjištěných dotazováním, se sice musí potýkat obě metody předvolebních odhadů volebních výsledků, protože i modelové odhady používají jako jeden ze vstupů výsledky výběrového šetření na národním reprezentativním vzorku. Výhoda modelových řešení ovšem spočívá v tom, že, na rozdíl od šetření v jednotlivých volebních obvodech, pracuje s informacemi o regionální distribuci volebních preferencí ze skutečných voleb, které nejsou ani zatíženy výběrovou chybou, protože byly získány od „úplného vzorku“ účastníků voleb, ani u nich nepřichází v úvahu komplikace se zmlčenými nebo nepravdivými stranickými preferencemi.

Velmi podstatnou výhodou modelového řešení je také jeho nízká cena, která může být řádově nižší než série šetření v jednotlivých volebních obvodech. Kromě nižší ceny mají modelová řešení ještě jednu nezanedbatelnou výhodu. Zatímco šetření v jednotlivých volebních obvodech je z praktických důvodů (nestandardně velké množství dotazníků, delší doba terénní fáze sběru dat) prakticky nemožné provést v čase těsně před volbami, modely pracují jen s výsledky standardně velkého celostátního šetření, které je snazší uskutečnit v posledním možném termínu umožňujícím publikaci výsledků ještě před volbami, kdy jsou stranické preference relativně nejbližší preferencím v době konání voleb a podíl nerozhodnutých voličů je nejmenší. Poté, co jsou k dispozici výsledky reprezentativního celostátního šetření, je aplikace modelu vedoucí k odhadům výsledků voleb v jednotlivých volebních obvodech již otázkou pouhých desítek minut.

Na první pohled se zdá, že při volbě mezi metodou výběrových šetření v jednotlivých volebních obvodech a modelovou metodou se jedná o rozhodnutí mezi drahým, ale kvalitním a vyzkoušeným řešením, a levným, ale nekvalitním řešením. Toto tvrzení je ovšem založeno na apriorním předpokladu, že odhady volebních výsledků ve volebních obvodech získaných metodou série výběrových šetření v jednotlivých volebních obvodech jsou bližší skutečným výsledkům voleb než modelově vytvořené odhady. Protože tento předpoklad nebyl nikdy v praxi prověřen, rozhodli jsme se provést praktický test kvality odhadů volebních výsledků provedený oběma typy metod porovnáním se skutečnými výsledky voleb. Před provedením tohoto testování jsme si formulovali několik základních hypotéz.

Hypotéza 1: Kvalita modelových odhadů volebních výsledků v jednotlivých volebních obvodech bude obecně srovnatelná s odhady provedenými na základě série výběrových šetření v obvodech.

Hypotéza 2: Čím více budou volby rozhodováním o podpoře konkrétních individuálních kandidátů, tím budou modelová řešení méně použitelná pro odhady výsledků voleb. Jinými slovy: čím více jde o stranické preference a čím méně o konkrétní osoby na kandidátkách, tím je lepší shoda mezi modelovými odhady a skutečnými výsledky voleb.

Hypotéza 3: Modelová řešení budou méně kvalitní v případě, že budou použita pro odhady výsledků voleb do nově vzniklých institucí, a naopak kvalitnější v případě, že budou použita na odhady volebních výsledků voleb do institucí, které již nějakou dobu fungují v relativně neměnné podobě.

### 3.3 Popis použité modelové metody odhadů volebních výsledků

Existuje celá řada možností, jak modelově odhadovat volební výsledky v jednotlivých volebních obvodech. Všechny modely jsou v zásadě postaveny na stejném principu: známé stranické preference (eventuálně volební preference či volební prognózy) zjištěné výběrovým šetřením na národním vzorku populace se „pře počítají“ na úroveň jednotlivých volebních obvodů na základě informací o distribuci hlasů pro politické strany v jednotlivých obvodech získaných z výsledků některých předcházejících voleb. Prvním krokem při modelování je vytvoření neúplné kontingenční tabulky, u níž známe pouze řádkové a sloupcové součty, ale nikoliv hodnoty ve „vnitřních políčkách“ této tabulky (viz tabulku 1).<sup>9</sup>

Proměnná  $P_a$  značí počet hlasů, který pravděpodobně dostane strana A v blízkých se volbách.  $P_b$ ,  $P_c$  ... analogicky značí počet hlasů, které pravděpodobně dostanou strany B, C... Tyto údaje je možné získat z předvolebního průzkumu stranických preferencí provedeného na vzorku reprezentujícím celou populaci. Proměnné

**Tabulka 1. Výchozí situace modelového řešení – neúplná kontingenční tabulka předpokládaných volebních výsledků v blízkých se volbách**

	Hlasy pro stranu A	Hlasy pro stranu B	Hlasy pro stranu C	Hlasy celkem
Volební obvod 1				$P_1$
Volební obvod 2				$P_2$
Volební obvod 3				$P_3$
Volební obvod 4				$P_4$
Volební obvod 5				$P_5$
Celkem	$P_a$	$P_b$	$P_c$	$P$

<sup>9</sup> Tabulka popisuje situaci, kdy tři politické strany usilují o voliče v pěti volebních obvodech. V praxi je samozřejmě počet stran v tabulce závislý na počtu stran účastnících se voleb (eventuálně počtu relevantních stran, jejichž volební zisk predikujeme) a skutečném počtu volebních obvodů (resp. počtu obvodů, v nichž výsledky predikujeme).

**Tabulka 2. Cílová situace modelového řešení – úplná kontingenční tabulka předpokládaných volebních výsledků v blízkících se volbách**

	Hlasy pro stranu A	Hlasy pro stranu B	Hlasy pro stranu C	Hlasy celkem
Volební obvod 1	$P_{a1}$	$P_{b1}$	$P_{c1}$	$P_1$
Volební obvod 2	$P_{a2}$	$P_{b2}$	$P_{c2}$	$P_2$
Volební obvod 3	$P_{a3}$	$P_{b3}$	$P_{c3}$	$P_3$
Volební obvod 4	$P_{a4}$	$P_{b4}$	$P_{c4}$	$P_4$
Volební obvod 5	$P_{a5}$	$P_{b5}$	$P_{c5}$	$P_5$
Celkem	$P_a$	$P_b$	$P_c$	$P$

$P_1, P_2, P_3, \dots$  představují předpokládané počty hlasů, které budou ve volbách odevzdány ve volebních obvodech 1, 2, 3, ... Tyto hodnoty by bylo možné teoreticky získat z celonárodního předvolebního průzkumu, neboť jde v podstatě o odhad volební účasti v jednotlivých volebních obvodech, vzhledem k nedostatečnému počtu respondentů pro takový odhad je však lépe počet zúčastněných v jednotlivých volebních obvodech odhadnout na základě aktuálně očekávané míry volební účasti a skutečné účasti voličů v předcházejících volbách. Hodnota  $P$  značí očekávaný celkový počet voličů v nadcházejících volbách. Místo absolutních hodnot je možné též použít hodnoty relativní, tedy nikoliv absolutní počty hlasů, ale očekávané podíly hlasů pro jednotlivé politické strany, resp. očekávané podíly hlasů odevzdané v jednotlivých volebních obvodech. Pro další postup modelování není důležité, zda pracujeme s absolutními nebo relativními čísly.

Cílem modelování je odhadnout procenta hlasů, které získají jednotlivé strany v jednotlivých volebních obvodech. Technicky jde o to najít takovou úplnou kontingenční tabulku (viz tabulku 2), v níž budou hodnoty  $p_{a1}, p_{a2}, p_{a3}, \dots, p_{c5}$  takové, že budou jejich řádkové a sloupcové součty odpovídat řádkovým a sloupcovým součtům z tabulky 1.

Až do tohoto okamžiku je výše popsán příklad klasickou ukázkou problému ekologické inference, tedy odvozování individuálních dat z dat agregátních, který je řešen s většími či menšími úspěchy po mnoho let [viz např. Berglund, Thomsen 1990; King 1997]. Hlavním problémem všech metod ekologické inference je skutečnost, že teoreticky existuje nekonečně velké množství takových hodnot  $p_{xy}$ , které splňují podmínku, aby jejich sloupcové a řádkové součty odpovídaly řádkovým a sloupcovým součtům z tabulky 1.

V našem případě ovšem máme k dispozici ještě jeden zdroj informací, který míru neurčitosti podstatně omezí, totiž úplnou kontingenční tabulku se skutečnými volebními výsledky stran v minulých volbách (tabulka 3).

Hledáme tedy takové hodnoty  $p_{a1}, p_{a2}, p_{a3}, \dots, p_{c5}$ , pro něž platí, že jejich řádkové a sloupcové součty odpovídají řádkovým a sloupcovým součtům z tabulky 1, a zároveň jsou vztahy mezi proměnnými  $p_{a1}, p_{a2}, p_{a3}, \dots, p_{c5}$  co nejpodobnější vztahům me-

**Tabulka 3. Další vstupní parametr modelového řešení – úplná kontingenční tabulka skutečných volebních výsledků jednotlivých stran v jednotlivých volebních obvodech v některých z minulých voleb**

	Hlasy pro stranu A	Hlasy pro stranu B	Hlasy pro stranu C	Hlasy celkem
Volební obvod 1	$q_{a1}$	$q_{b1}$	$q_{c1}$	$Q_1$
Volební obvod 2	$q_{a2}$	$q_{b2}$	$q_{c2}$	$Q_2$
Volební obvod 3	$q_{a3}$	$q_{b3}$	$q_{c3}$	$Q_3$
Volební obvod 4	$q_{a4}$	$q_{b4}$	$q_{c4}$	$Q_4$
Volební obvod 5	$q_{a5}$	$q_{b5}$	$q_{c5}$	$Q_5$
Celkem	$Q_a$	$Q_b$	$Q_c$	$Q$

zi proměnnými  $q_{a1}$ ,  $q_{a2}$ ,  $q_{a3}$  ...  $q_{c5}$ , které můžeme odvodit z tabulky 3. Při konkrétním řešení tohoto matematického úkolu jsme použili program LOCCONTINGENCY,<sup>10</sup> který byl pro podobné typy úloh vytvořen v Sociologickém ústavu Akademie věd České republiky.

### 3.4 Výsledky empirických testů

První modelový odhad volebních výsledků v jednotlivých volebních obvodech, který jsme dělali, predikoval výsledky voleb do Poslanecké sněmovny v roce 2002. Soustředili jsme se v něm na odhadnutí procenta hlasů získaných čtyřmi největšími volebními stranami – ČSSD, ODS, KSČM a KOALICE – v jednotlivých volebních krajích. Svůj zájem jsme úmyslně omezili jen na tyto čtyři politické subjekty, protože to byly jediné, které měly podle výsledků výzkumů stranických preferencí reálnou šanci získat více než 5 % hlasů, a tak překročit zákonem danou hranici potřebnou pro vstup do Poslanecké sněmovny. Strany s menší voličskou podporou bylo možno zanedbat, protože jim odevzdané hlasy se při rozdělení mandátů nijak neuplatňují. Ačkoliv jsou skutečným výsledkem voleb počty rozdělených křesel mezi strany, omezili jsme se v našem testu na určení procenta hlasů pro strany v jednotlivých obvodech. Porovnávat procenta hlasů a nikoliv počty rozdělených mandátů v krajích je v naší situaci výhodnější, protože umožňují měřit přesněji rozdíly mezi odhady a skutečnými výsledky voleb i mezi oběma typy odhadů navzájem.<sup>11</sup> Zdrojem infor-

<sup>10</sup> Podrobný popis matematického algoritmu, na kterém je práce programu LOCCONTINGENCY založena, je podán jeho autorem v publikaci Vajda [2002] a Vajda, van der Meulen [2003]. Popis vlastností programu a možnosti jeho užití v analýzách politického chování je podán v publikaci Kostecký, Čermák [2003].

<sup>11</sup> Samotná distribuce křesel mezi strany podle podílu získaných hlasů v jednotlivých krajích je v tomto ohledu jen pouhou aplikací matematické procedury přesně určené volebním zákonem. Z matematického hlediska dochází při přepočtu hlasů v kraji na mandáty k jistým „chybám ze zaokrouhlování“, které mohou za určitých podmínek odhad počtu mandátů o ně-

**Tabulka 4. Volby do Poslanecké sněmovny Parlamentu v roce 2002 – podíl hlasů pro čtyři nejúspěšnější volební strany v jednotlivých volebních krajích odhadnutý metodou LOCCONTINGENCY**

	ČSSD	ODS	KSČM	KOALICE
Praha	18,43	42,88	10,75	20,44
Středočeský	25,73	30,17	17,66	14,98
Jihočeský	24,28	29,06	17,27	18,78
Plzeňský	24,89	28,02	18,15	17,09
Karlovarský	27,45	27,64	17,62	13,14
Ústecký	27,83	24,97	21,48	10,90
Liberecký	25,34	30,14	13,98	17,15
Královéhradecký	22,54	30,10	14,51	21,05
Pardubický	24,65	26,14	15,83	21,89
Vysočina	25,08	21,23	19,97	22,70
Jihomoravský	24,15	23,70	19,09	22,33
Olomoucký	29,00	22,63	18,69	17,66
Moravskoslezský	32,78	22,66	17,78	15,11
Zlínský	25,27	24,41	15,67	24,46

Zdroj: Vlastní výpočty

mací o stranických preferencích na úrovni celé České republiky bylo šetření agentury STEM, které bylo provedeno ve dnech 31. 5. až 1. 6. roku 2002. Odhad rozdělení hlasů mezi jednotlivé volební kraje v roce 2002 byl odvozen od skutečného regionálního rozložení<sup>12</sup> odevzdaných hlasů ve volbách v roce 1998 (publikováno na volebním serveru Českého statistického úřadu [www.volby.cz](http://www.volby.cz)). Jako vstupní parametr modelového odhadu byly použity skutečné výsledky voleb do Poslanecké sněmovny z roku 1998. Určitou komplikaci způsobila skutečnost, že jedním z významných politických subjektů před volbami v roce 2002 byla koalice dvou stran – KDU-ČSL a US-DEU – které se ucházely o přízeň voličů pod názvem KOALICE. KOALICE zahrnovala ve skutečnosti strany, které se v roce 1998 zúčastnily voleb jako tři samostatné subjekty.<sup>13</sup> Volební výsledky z roku 1998 jsme proto upravili tak, že jsme sečetli hla-

---

co přiblížit nebo naopak poněkud vzdálit počtu křesel distribuovaných ve skutečných volbách. Vždy přitom ovšem platí, že pokud je odhad procenta hlasů pro strany v jednotlivých krajích zjištěný metodou X skutečným procentům získaným stranami v krajích bližší než odhad procenta hlasů získaný metodou Y, bude i odhad počtu mandátů vycházející z výsledků metody X stejný nebo bližší než odhad počtu mandátů vycházející z výsledků metody Y.

<sup>12</sup> Volební kraje v roce 1998 a v roce 2002 nekorespondovaly. Volební účast v roce 1998 byla proto nejprve přepočtena na územní jednotky totožné s kraji použitými v roce 2002.

<sup>13</sup> Zmíněnými třemi stranami byly Unie svobody (US), Demokratická unie (DEU) a Křesťanská a demokratická unie-Československá strana lidová (KDU-ČSL). V období mezi volbami

**Tabulka 5. Volby do Poslanecké sněmovny Parlamentu v roce 2002 – podíl hlasů pro čtyři nejúspěšnější volební strany v jednotlivých volebních krajích odhadnutý na základě výběrových šetření v jednotlivých volebních krajích agenturou STEM (stranické preference přepočtené bez odpovědí „nevím“)**

	ČSSD	ODS	KSČM	KOALICE
Praha	25,90	33,30	10,50	18,10
Středočeský	23,60	30,90	14,30	16,30
Jihočeský	22,60	33,30	14,30	17,90
Plzeňský	24,00	28,80	16,50	15,10
Karlovarský	26,50	26,30	18,10	11,20
Ústecký	28,60	22,80	21,20	10,50
Liberecký	27,20	32,10	10,30	15,30
Královéhradecký	23,40	26,90	13,70	18,50
Pardubický	24,90	27,10	14,30	21,40
Vysočina	24,30	24,70	16,20	20,80
Jihomoravský	23,70	20,10	17,20	21,90
Olomoucký	22,90	24,20	19,00	19,60
Moravskoslezský	30,90	24,60	16,40	14,70
Zlínský	27,30	26,50	14,40	22,60

Zdroj: Lebeda [2003]

sy získané všemi třemi zmíněnými politickými stranami a použili je jako ukazatel předikující potenciální zisk KOALICE v jednotlivých volebních obvodech v roce 2002. Výsledky modelového výpočtu jsou uvedeny v tabulce 4. Tabulka 5 ukazuje, jak odhadovala výsledky jednotlivých politických stran ve volebních krajích agentura STEM tradiční metodou výběrového šetření v jednotlivých volebních krajích. Stranické preference publikované agenturou STEM jsme ještě poněkud vylepšili – aby se více přiblížily možným výsledkům voleb, nebyly v celkovém součtu hlasů brány v potaz odpovědi „nevím“. Výsledná procenta podpory pro jednotlivé politické strany se tak spíše podobala volebním preferencím [Lebeda 2003]. Pro srovnání jsou v tabulce 6 uvedeny skutečné výsledky voleb.

Existuje celá škála ukazatelů, které je možno použít k určení toho, do jaké míry se předvolební odhady blíží skutečným volebním výsledkům. Různé ukazatele se vyznačují různou mírou citlivosti na počet porovnávaných čísel v tabulce, absolutní hodnoty porovnávaných čísel, na extrémní hodnoty v tabulce nebo hodnoty blízké 0. Ačkoliv různé ukazatele měří odlišným způsobem „vzdálenost“ mezi modelovými a skutečnými výsledky, tedy míru shody resp. neshody mezi odhady a skutečnými

---

1998 a 2002 se několikrát změnila forma spolupráce zmíněných stran. US a DEU se sloučily do jediné strany, takže v době voleb šlo formálně o koalici dvou stran.

**Tabulka 6. Volby do Poslanecké sněmovny Parlamentu v roce 2002 – podíl hlasů pro čtyři nejúspěšnější volební strany v jednotlivých volebních krajích – skutečné volební výsledky**

	ČSSD	ODS	KSČM	KOALICE
Praha	25,85	33,84	11,11	18,47
Středočeský	31,53	26,84	18,67	11,91
Jihočeský	30,33	25,99	18,25	12,25
Plzeňský	30,34	25,51	19,61	11,76
Karlovarský	29,32	24,70	21,98	8,85
Ústecký	29,19	24,22	25,09	7,82
Liberecký	27,06	27,41	17,18	13,25
Královéhradecký	27,48	26,59	16,06	14,46
Pardubický	29,46	22,90	17,39	16,80
Vysočina	31,97	19,27	19,70	16,92
Jihomoravský	29,91	20,95	19,76	17,35
Olomoucký	31,92	20,35	21,02	13,95
Moravskoslezský	36,14	19,38	21,07	11,83
Zlínský	29,07	21,48	16,64	19,46

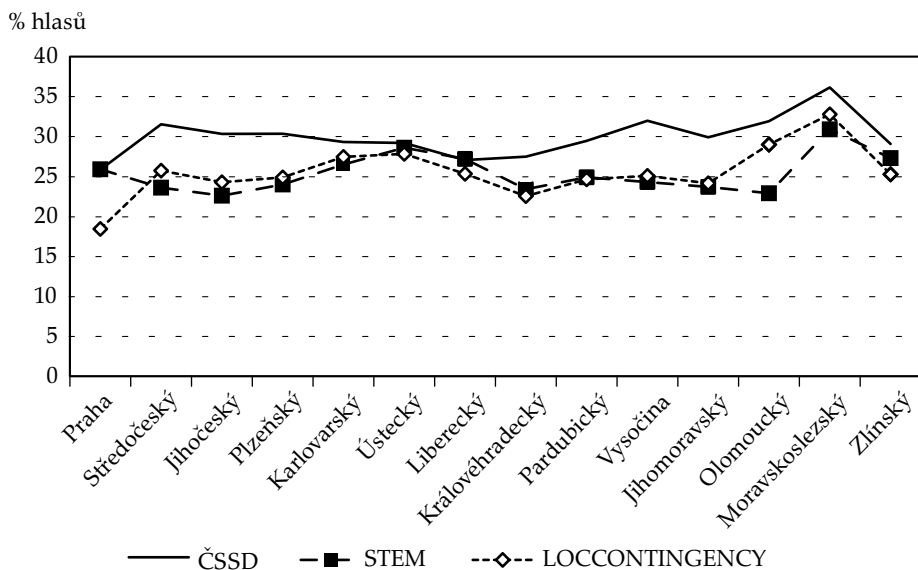
Zdroj: Český statistický úřad ([www.volby.cz](http://www.volby.cz))

výsledky, naprostá většina z nich se shoduje v určení toho základního – shodně určuje, který z odhadů je skutečným výsledkům bližší. Pro hodnocení shody odhadů a skutečných volebních výsledků použijeme v následujícím textu jednoduchého ukazatele  $K$ . U každé dvojice porovnávaných tabulek nejprve vypočítáme rozdíly (v absolutní hodnotě) mezi modelovým odhadem a skutečným výsledkem voleb pro každé políčko tabulky, zjištěné hodnoty sečteme a nakonec vydělíme počtem políček v tabulce. Získáme tak průměrnou odchylku mezi modelovým odhadem a skutečnými výsledky voleb připadající na jedno políčko tabulky. Čím větší je hodnota ukazatele  $K$ , tím větší je odchylka mezi odhadem a skutečnými volebními výsledky.<sup>14</sup>

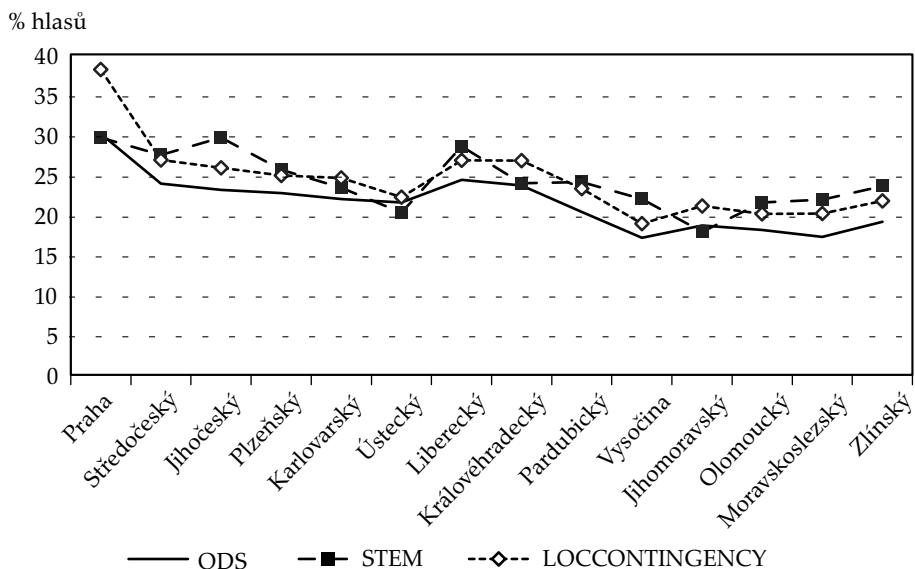
V našem konkrétním případě je u tradičního odhadu provedeného agenturou STEM hodnota tohoto ukazatele  $K = 3.73$  a v případě modelového odhadu provedeného pomocí programu LOCCONTINGENCY je hodnota ukazatele  $K = 3.48$ . Z porovnání je zřejmé, že postup používající modelové odhady poskytl v tomto případě předvolební odhady podobné kvality, jako tradiční série předvolebních výběrových šetření v jednotlivých krajích. Modelem předpovězené volební výsledky v jednotlivých volebních krajích byly dokonce v průměru o něco bližší skutečným výsledkům voleb.

<sup>14</sup> Výhodou námi zvoleného ukazatele je jeho jednoduchost a názornost. Nevýhodou je naopak nemožnost exaktně určit statistickou významnost zjištěného rozdílu. Pro tento účel by bylo třeba využít některého ze statistických testů dobré shody, např. Chí kvadrát.

**Graf 1. Volby do Poslanecké sněmovny Parlamentu ČR v roce 2002 podle volebních krajů. ČSSD – předvolební odhady STEM tradiční metodou výběrových šetření v krajích, modelové odhady metodou LOCCONTINGENCY a skutečné výsledky voleb**

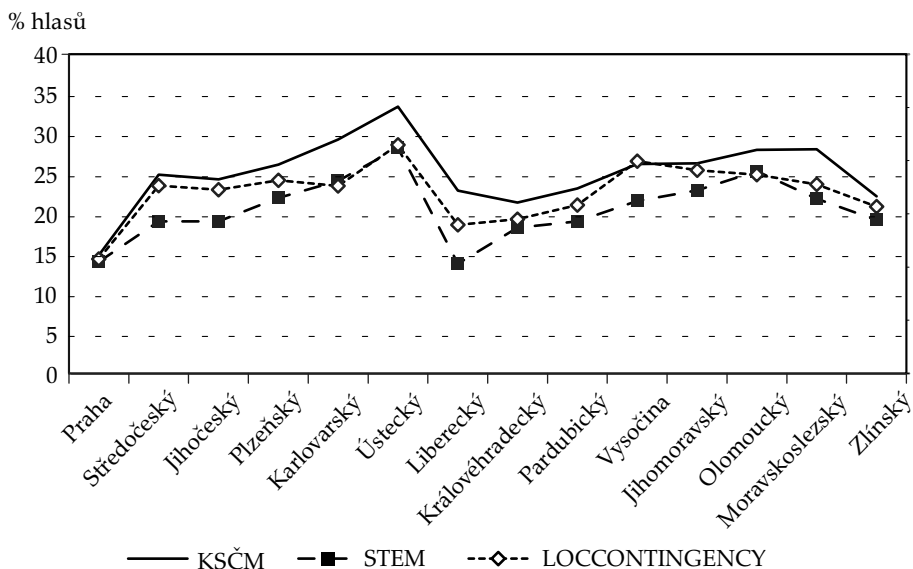


**Graf 2. Volby do Poslanecké sněmovny Parlamentu ČR v roce 2002 podle volebních krajů. ODS – předvolební odhady STEM tradiční metodou výběrových šetření v krajích, modelové odhady metodou LOCCONTINGENCY a skutečné výsledky voleb**

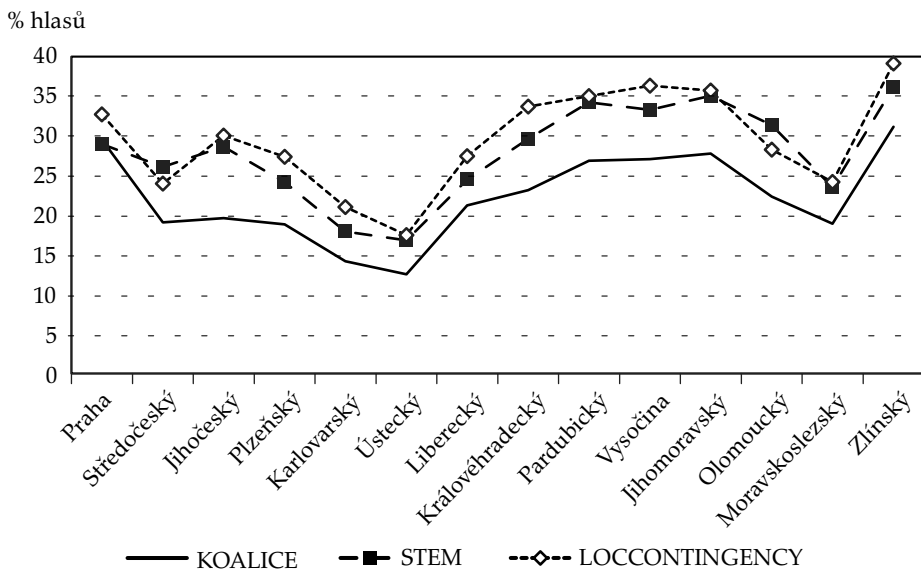




**Graf 3. Volby do Poslanecké sněmovny Parlamentu ČR v roce 2002 podle volebních krajů. KSČM – předvolební odhady STEM tradiční metodou výběrových šetření v krajích, modelové odhady metodou LOCCONTINGENCY a skutečné výsledky voleb**



**Graf 4. Volby do Poslanecké sněmovny Parlamentu ČR v roce 2002 podle volebních krajů. KOALICE – předvolební odhady STEM tradiční metodou výběrových šetření v krajích, modelové odhady metodou LOCCONTINGENCY a skutečné výsledky voleb**



Poněkud podrobnější pohled na srovnání modelových odhadů s tradičními odhady postavenými na výběrových šetřeních ve volebních krajích a skutečnými výsledky voleb nám, zvláště pro každou ze čtyř hlavních volebních stran, poskytnou grafy 1 až 4.

Z grafů je patrné, že jak krajské odhady STEM, tak modelové odhady dokázaly v zásadě predikovat volební zisky volebních stran v jednotlivých volebních krajích, přesnost odhadů se ovšem lišila pro jednotlivé strany i kraj od kraje. Obecně se dá říci, že největší problém bylo správně předpovědět volební výsledky ČSSD. Oba volební odhady zisk ČSSD podcenily, realitě přece jen o něco blíž byly odhady modelové. Hlavním důvodem pro to byla skutečnost, že stranické preference ČSSD v předvolebním období rostly – modelové odhady v tomto případě těžily z toho, že jako vstupní data pro výpočet mohla být využita data o stranických preferencích sebraná relativně blíže datu skutečných voleb, což si tradiční odhad provedený metodou šetření v jednotlivých krajích nemohl z organizačních a časových důvodů dovolit. Zatímco krajské výzkumy STEM byly uskutečňovány v průběhu měsíce května, model pracoval s výsledky celostátního výzkumu preferencí STEM ze dnů 31. 5. až 1. 6. Ze stejného důvodu zřejmě model předpověděl lépe i výsledky KSČM. Modelový odhad se od skutečných volebních výsledků nejvíce odchytil v případě KOALICE, jejíž volební zisky předvídal na vyšší úrovni než krajské výzkumy STEM. Je pravděpodobné, že použití součtu volebních zisků součástí KOALICE (KDU-ČSL, US a DEU) ve výsledcích voleb z roku 1998, jako jednoho ze vstupních parametrů modelu odhadujícího výsledky KOALICE ve volbách v roce 2002, necharakterizovalo skutečnou sílu KOALICE a bylo hlavní příčinou nadhodnocení jejího volebního zisku modelem.

Zajímavější než posuzování přesnosti odhadů podle jednotlivých stran, které je do jisté míry podmíněno načasováním sběru dat u tradičních výběrových šetření v krajích, je srovnání schopnosti obou typů metod odhalit regionální rozdíly ve volebních preferencích. Pohled na grafy 1 až 4 naznačuje, že modelové odhady by mohly být v tomto ohledu přesnější, protože průběh křivek popisujících modelové odhady v krajích je podobnější průběhu křivek charakterizujících skutečné volební výsledky. Tento předpoklad potvrzují i hodnoty Pearsonova korelačního koeficientu, které měří míru lineární asociace mezi odhady volebních výsledků v krajích tradiční metodou výběrových šetření v krajích, resp. metodou modelování s využitím programu LOCCONTINGENCY a skutečnými výsledky voleb v jednotlivých krajích (tabulka 7).

Obecně lze říci, že oba typy odhadů dobře vystihly regionální rozdíly ve volební podpoře jednotlivým stranám. Až na jedinou výjimku (krajské odhady STEM pro ČSSD) jsou korelační koeficienty statisticky významné na 99% hladině spolehlivosti. Údaje v tabulce také potvrzují, že modelové odhady jsou v tomto ohledu přesnější než tradiční metoda výběrových šetření v jednotlivých volebních krajích. Na závěr by bylo dobré ještě upozornit na jeden konkrétní případ, v němž se nejvíce odlišovaly modelové odhady od krajských výzkumů STEM, tedy výsledky ODS v Praze. Modelem předpovězený zisk ODS (42,88 %) byl o více než 9 procentních bodů vyšší, než

**Tabulka 7. Volby do Poslanecké sněmovny Parlamentu v roce 2002 – korelace mezi krajskými odhady STEM, modelovými odhady LOCCONTINGENCY v krajích a podíl hlasů pro čtyři nejúspěšnější volební strany v jednotlivých volebních krajích (Pearson r)**

	ČSSD	ODS	KSČM	KOALICE
odhady STEM	0,197	0,767**	0,892**	0,909**
odhady LOCCONTINGENCY	0,799**	0,970**	0,912**	0,940**

Zdroj: Vlastní výpočty

Pozn.: korelace označené \*\* jsou statisticky významné na 99% hladině spolehlivosti

odhad STEM učiněný na základě výběrového šetření v Praze (33,30 %). Odhad STEM byl v tomto případě skutečným výsledkům ODS v Praze (33,84 %) podstatně blíže než odhad modelu. Příčinou byly pravděpodobně hluboké rozpory uvnitř pražské ODS, které vyvrcholily vystoupením populárního primátora Prahy Jana Kasla ze strany těsně před volbami. Model, který odvozoval volební podporu ODS v Praze podle volebních výsledků v roce 1998, na tuto situaci, pochopitelně, nereagoval, zatímco lokální výzkum STEM uskutečněný v Praze tuto skutečnost dokázal zachytit.

Obdobným způsobem, jako jsme modelovali výsledky voleb do Poslanecké sněmovny Parlamentu v roce 2002 a porovnávali je s výsledky krajských průzkumů provedené před týmiž volbami agenturou STEM, jsme provedli ještě několik dalších testů modelu, jejichž výsledky jsme posléze porovnali s publikovanými výsledky tradičních předvolebních výzkumů volebních preferencí. Ve všech případech jsme pracovali pouze s výsledky čtyř největších stran – ČSSD, ODS, KSČM a (ČTYŘ)KOALICE. Je samozřejmě možné odhadovat modelem výsledky i více stran než čtyř a v některých případech (např. krajské volby) by to zřejmě bylo i vhodné. Kvůli vzájemné srovnatelnosti výsledků jednotlivých testů jsme ovšem používali stejný soubor volebních stran. Tabulka 8 dává přehled o tom, jaké volební výsledky byly modelovány, o vstupních parametrech modelů, a rovněž poskytuje informace o výběrových šetřeních, s nimiž byly modelové odhady porovnávány. (Pro úplnost jsou do tabulky zahrnuty i informace o výše popsaném modelovém odhadu výsledků voleb do Poslanecké sněmovny v roce 2002, které byly srovnávány s volebními odhady STEM).

Výběr se řídil především dostupností publikovaných regionálních odhadů prováděných tradičními metodami výběrových šetření. Tabulka 9 ukazuje základní výsledky testování kvality modelových řešení ve srovnání s tradičními metodami. U každého z testovaných případů uvádíme hodnotu ukazatele  $K$  = průměrnou odchylku mezi modelovým odhadem (resp. tradičním odhadem) a skutečnými výsledky voleb připadající na jedno políčko tabulky.

Výsledky testů sumarizované v tabulce nám umožní vyjádřit se k hypotézám, které jsme si před testováním modelu stanovili.

Tabulka 8. Přehled provedených testů – základní informace

Číslo testu	Rok	Typ voleb	Počet volebních obvodů	Vstupní parametry modelu	Porovnávané tradiční výběrové šetření
1a.	2000	Senát	9*	– Stranické preference STEM. – Výsledky voleb do Senátu 1996	Průzkum konsorcia SC&C a STEM ve vybraných senátních volebních obvodech
1b.	2000	Senát	9*	– Stranické preference STEM. – Výsledky voleb do Poslanecké sněmovny 1998	Průzkum konsorcia SC&C a STEM ve vybraných senátních volebních obvodech
2.	2000	Zastupitelstva krajů	13	– Stranické preference STEM. – Výsledky voleb do Poslanecké sněmovny 1998	Průzkum konsorcia SC&C a STEM v jednotlivých krajích
3a.	2002	Poslanecká sněmovna	14	– Stranické preference TNSF. – Výsledky voleb do Poslanecké sněmovny 1998	Krajské volební prognózy TNSF
3b.	2002	Poslanecká sněmovna	14	– Volební prognózy TNSF. – Výsledky voleb do Poslanecké sněmovny 1998	Krajské volební prognózy TNSF
4	2002	Poslanecká sněmovna	14	– Volební preference STEM. – Výsledky voleb do Poslanecké sněmovny 1998	Krajské volební preference STEM

Pozn.: Testy 1a a 1b (resp. 3a a 3b) se liší jen v jednom ze vstupních parametrů modelu, jde tedy jen o dvě různé varianty téhož testu.

\* Celkový počet volebních obvodů, v nichž se konaly volby do třetiny Senátu, byl 27. Modelovány byly jen výsledky v 9 vybraných obvodech, v nichž byly zároveň prováděny tradiční výzkumy formou výběrového šetření.

– *Hypotéza 1: Kvalita modelových odhadů volebních výsledků v jednotlivých volebních obvodech bude obecně srovnatelná s odhady provedenými na základě série výběrových šetření v obvodech.* Zdá se být nepochybné, že hypotézu 1 můžeme přijmout. Ve čtyřech ze šesti testů se modelové odhady od skutečných volebních výsledků odlišovaly v průměru méně, než odhady prováděné tradiční metodou. Jen ve zbývajících dvou případech byly modelové odhady horší než tradiční odhady, rozdíl v kvalitě obou odhadů však nebyl ani u těchto testů příliš velký.

– *Hypotéza 2: Čím více budou volby rozhodováním o podpoře konkrétních individuálních kandidátů, tím budou modelová řešení méně použitelná pro odhady výsledků voleb.* Výsledky testování ukazují, že i tato hypotéza by mohla být přijata. Nejvíce se modelové odhady přiblížily skutečným výsledkům u voleb do Poslanecké sněmovny, kde se voliči rozhodují značně stranicky. Naopak relativně nejhorší byly modelové odhady výsledků senátních voleb, které jsou díky použití většinového volebního systému v jednomandátových volebních obvodech typickým příkladem voleb, v nichž se voliči rozhodují více podle osobností kandidátů a méně podle politických stran, za něž se kandidáti ucházejí o jejich podporu. V tomto smyslu je zajímavé i srovnání

**Tabulka 9. Srovnání kvality modelových odhadů a odhadů provedených tradičními metodami výběrových šetření ve volebních obvodech (uvedena je hodnota  $K$  = průměrná odchylka mezi modelovým odhadem, resp. tradičním odhadem a skutečnými výsledky voleb připadající na jedno políčko kontingenční tabulky)**

Číslo testu	Rok	Typ voleb	Tradiční šetření	Kvalita modelového odhadu ( $K$ =)	Kvalita tradičního odhadu ( $K$ =)
1a.	2000	Senát	SC&C a STEM	7.67	6.49
1b.	2000	Senát	SC&C a STEM	6.70	6.49
2.	2000	Zastupitelstva krajů	SC&C a STEM	4.80	5.12
3a.	2002	Poslanecká sněmovna	TNSF	3.28	3.77
3b.	2002	Poslanecká sněmovna	TNSF	3.09	3.77
4	2002	Poslanecká sněmovna	STEM	3.48	3.73

Zdroj: vlastní výpočty

testů 1a a 1b, které se vzájemně lišily jen tím, jaké starší volební výsledky byly použity jako vstupní parametr modelu. Model 1a, v němž byly pro modelování volebních výsledků ve vybraných senátních obvodech v roce 2000 použity informace o skutečných výsledcích voleb v těchto senátních obvodech v roce 1996, předpověděl výsledky hůře než model 1b, v němž byly jako vstupní informace modelu použity volební výsledky voleb do Poslanecké sněmovny z roku 1998, přepočítané na příslušné senátní volební obvody. Je zřejmé, že výsledky senátních voleb v roce 2000 byly značně závislé na osobnostech individuálních kandidátů. Stejně tak tomu bylo i v roce 1996, proto nebyly výsledky senátních voleb z roku 1996 příliš vhodným prediktorem pro modelování. Lepší vstupní informaci pro modelování senátních voleb v roce 2000 paradoxně poskytly výsledky voleb do Poslanecké sněmovny, které byly daleko stráničtější, a sloužily tak jako informace o obecné míře podpory stran v senátních obvodech. Identifikace se stranou je stabilnější charakteristikou volebního obvodu než identifikace s konkrétním kandidátem.

– *Hypotéza 3: Modelová řešení budou méně kvalitní v případě, že budou použita pro odhady výsledků voleb do nově vzniklých institucí, a naopak kvalitnější v případě, že budou použita na odhady volebních výsledků voleb do institucí, které již nějakou dobu fungují v relativně neměnné podobě.* Z výsledků testování není úplně zřejmé, zda můžeme tuto hypotézu rovněž přijmout. Modelové volební odhady u voleb do Senátu a krajských zastupitelstev, které jsou relativně „mladšími“ institucemi, byly sice méně kvalitní než modelové odhady výsledků voleb do Poslanecké sněmovny, která už má v tomto smyslu určitou historii, není ale úplně jisté, zda jsou rozdíly způsobeny „novostí“ obou institucí nebo jinými faktory – např. použitým volebním systémem. Volby do Poslanecké sněmovny v roce 2002 také nemohou být považovány za vzor voleb ve stabilní situaci – oproti roku 1998 došlo ke slučování některých relevantních politických stran i k úpravě volebního systému.

#### **4. Závěr**

Provedené analýzy ukázaly, že modelové odhady volebních výsledků, které vycházejí z výsledků národních reprezentativních výzkumů stranických či volebních preferencí a informací o regionální distribuci volební podpory stranám získaným z výsledků některých voleb v minulosti, jsou životaschopnou alternativou tradičních odhadů vycházejících ze série šetření v jednotlivých volebních obvodech. Obrovskou výhodou modelových řešení je jejich řádově nižší cena. Relativně vysoká míra kontinuity volebního chování regionálních populací v čase spolu se skutečností, že informace o chování voličů získané ze skutečných volebních výsledků nejsou zatíženy výběrovou chybou, vysokým procentem nevyšetřených respondentů ani nepravdivými odpověďmi respondentů, kteří se účastní výběrových šetření, patří také k silným stránkám modelových odhadů. To ve značné míře kompenzuje hlavní metodologický nedostatek modelových řešení, totiž neschopnost zachytit regionálně specifické reakce na konkrétní politický vývoj, konkrétní kandidáty nebo události, což naopak představuje hlavní silnou stránku výběrových šetření. Relativně omezený počet volebních odhadů založených na provedení výběrových šetření v jednotlivých volebních obvodech neumožňoval detailně otestovat všechny hypotézy o vlastnostech modelových odhadů ve srovnání s tradičními odhady, a některé otázky tak zůstávají otevřené. Rozšíření možnosti testovat modelové odhady i na datech pocházejících ze zahraničí by bylo rovněž velmi cenné, neboť by nejenom dodalo více dat pro testování, ale zároveň poskytlo cenné informace o kulturním kontextu, v němž mohou být modelové odhady prováděny, a jeho vlivu na jejich kvalitu. Už doposud provedené testy ukazují, že existuje i prostor pro vylepšování kvality modelových odhadů, spočívající v zahrnutí odhadu vlivu některých faktorů (např. regionálně specifické změny v míře nezaměstnanosti) do modelových řešení. Je zřejmé, že zabývat se alternativními způsoby odhadování volebních výsledků ve volebních obvodech má smysl.

*TOMÁŠ KOSTELECKÝ je vědeckým pracovníkem Sociologického ústavu AV ČR, vedoucím oddělení Lokální a regionální problémy. Ve svém výzkumu se věnuje především studiu vlivu teritoriálních faktorů na lidské chování, politické geografii, regionálním aspektům vývoje společnosti a komparativní politice. V roce 2002 vydal v nakladatelství Woodrow Wilson Center Press monografii Political Parties After Communism. Developments in East-Central Europe.*

#### **Literatura**

- Berglund, S., S. R. Thomsen 1990. *Modern Political Ecological Analysis*. Copenhagen: Abo Academic Press.
- Čermák, Z. 1999. „Distinctive features of migration in the Czech Republic as part of the transformation of the Central European countries“. Pp. 209–222 in Hampf, M. et al.

- Geography of Societal Transformation in the Czech Republic*. Praha: Katedra sociální geografie a regionálního rozvoje, Přírodovědecká fakulta Univerzity Karlovy.
- FSC&C, STEM 2000a. „Jaké jsou šance kandidátů ve 2. kole voleb do Senátu“. *Mladá fronta Dnes*, 15. 11. 2000: 4.
- FSC&C, STEM 2000b. „Parlamentní strany opanují i krajské sněmy 2000“. *Mladá fronta Dnes*, 8. 11. 2000: 4.
- Jehlička, P., L. Sýkora 1991. „Stabilita regionální podpory tradičních politických stran v Českých zemích (1920–1990)“. *Sborník ČGS* 2: 81–95.
- Jehlička, P., T. Kostecký 2003: „Czech Greens in the 2002 General Election: A New Lease of Life?“ *Environmental Politics* 12 (2): 133–138.
- King, G. 1997. *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.
- Kostecký, T. 1993. *Analýza volebních výsledků jako součást politické geografie*. (Dizertace). Praha: Přírodovědecká fakulta Univerzity Karlovy.
- Kostecký, T. 2000a. „Navrhované změny volebního zákona vzešlé z dodatků ‚opoziční smlouvy‘ v roce 2000 a jejich možné důsledky“. *Sociologický časopis* 36 (3): 299–305.
- Kostecký, T. 2000b. „Volební systémy a politický konflikt v prostoru“. Pp. 242–262 in Jehlička, P., J. Tomeš, P. Daněk (eds.): *Stát, prostor, politika. Vybrané kapitoly politické geografie*. Praha: Katedra sociální geografie a regionálního rozvoje, Přírodovědecká fakulta Univerzity Karlovy.
- Kostecký, T. 2001. *Vzestup nebo pád politického regionalismu?* Working Papers 01:9. Praha: Sociologický ústav AV ČR.
- Kostecký, T. 2003. „Results of the 1992 and 1996 Parliamentary Elections in the Czech Republic from a Regional Perspective.“ Pp. 84–108 in Mansfeldová, Z. (ed.): *Czech Republic: The First Elections in the New Republic, 1992–1996*. Berlin: Edition Sigma.
- Kostecký, T., D. Čermák 2003. „Výběrová šetření a analýza agregátních dat – diskuse na téma použitelnosti různých přístupů v komparativních analýzách politického chování.“ *Sociologický časopis* 39 (4): 529–550.
- Lebeda, T. 2000. „Possibilities for Modifying the System of Proportional Representation Aimed at Stabilising the Executive in the ČR“. *Czech Sociological Review* 1: 49–68.
- Lebeda, T. 2003. „Průzkumy volebních preferencí před volbami do Sněmovny 2002. Srovnání výzkumů agentur CVVM, STEM a TNSF“. *Politologický časopis* 1: 22–37.
- Plecitá, K. 2003. *Voliči v letech 1992–2000. Studie vývoje stranické identifikace v České republice*. (Dizertace). Praha: Fakulta sociálních věd, Univerzita Karlova.
- Sokol, P. 2003. „Volební geografie při volbách do Poslanecké sněmovny“. Pp. 152–160 in Linek, L., L. Mrklas, A. Seidlová, P. Sokol (eds.): *Volby do Poslanecké sněmovny 2002*. Praha: Sociologický ústav AV ČR.
- Vajda, I. 2001. *Adaptation of Contingency Tables to Local Marginals*. Praha: Sociologický ústav AV ČR.
- Vajda, I., E. C. van der Meulen 2005. „On minimum divergence adaptation of discrete bivariate distributions to given marginals“. *IEEE Transactions on Information Theory* 51: 313–320.