

Fylgisbreytingar í aðdraganda borgarstjórnarkosninga vorið 2002

Guðmundur B. Arnkelsson dósent
við félagsvísindadeild Háskóla Íslands

1. tbl. 2. árg. 2006
2. Fræðigreinar



Útdráttur

Lýst er annmörkum þess að álykta um fylgi og fylgisþróun á grundvelli niðurstæðna einstakra fylgiskannana. Gerð er grein fyrir almennri umræðu í aðdraganda borgarstjórnarkosninga vorið 2002 og helstu annmarka þeirrar umræðu. Grein er gerð fyrir þeim eiginleikum sem sambæft mat á fylgisþróun sem nær yfir niðurstöður margra kannana þarf að hafa. Staðbundin aðfallsgreining (local regression) er notuð til að fá yfirlit yfir fylgisþróun síðustu fimm vikur fyrir kosningar, kosningaspá og mat á óvissu. Í ljós kom að fylgibreytingar voru mjög hægar, yfirburðir R-lista minnkuðu fram að 14. maí en jukust síðan fram að kejrdegi. Minnstur var fylgismunurinn tæp sex prósentustig og sigur R-lista því aldrei verulega í hættu. Með hlidsjón af mikilli óvissu í niðurstöðum einstakra kannana, sáust fá merki um umtalsverðan kerfisbundinn mun eftir könnunaraðilum. Niðurstöður rannsóknarinnar undirstrika mikilvægi þess að rannsaka fylgisþróun og hagnýta vitneskju um hana til að búa til líkön sem gera kleift að spá fyrir um niðurstöður kosninga og lágmarka óvissu í slíkri spá.

Fylgibreytingar í aðdraganda borgarstjórnarkosninga vorið 2002

Við lifum á tímum fylgiskannana þar sem umræða og fréttir vikunnar fyrir kosningar snúast að umtalsverðu leyti um niðurstöður kannana, fylgi og fylgibreytingar framboðslista. Miklir hagsmunir eru í húfi. Fyrir hinn almenna borgara er mikilvægt að fá sem réttastar upplýsingar til skilnings á þeim ferlum sem ráða aðdraganda og niðurstöðum kosninga. Hagsmunir frambjóðenda eru augljósir í því að fá sem gleggstar upplýsingar til leiðbeiningar í kosningabaráttunni. Fyrir könnunaraðila eru hagsmunir fjárhagslegir en einnig tengdir ímynd sem öflug og trausts fyrirtækis á sviði skoðanakannana.

Aðferðir könnunaraðila eru ólíkar að mörgu leyti. Úrtaksaðferðir eru ólíkar. Sumir draga úrtak úr þjóðskrá og hafa umtalsverða fyrirhöfn af því að leita uppi sem flesta þeirra sem völdust í könnunarhópinn. Aðrir fara eftir símaskrá og skilgreina jafnvel engan hóp fyrir fram heldur hringja þar til tiltekinn fjöldi hefur svarað. Orðalag spurninga og fjöldi eru ólíkur, tímasetning kannana og lengd könnunartímabila eru breytileg frá einum aðila til annars. Þegar niðurstöður birtast er eitthvað liðið frá því könnun var gerð og þessi birtingartöf er mismikil milli kannana og milli könnunaraðila. Eðlilega eru niðurstöður breytilegar milli kannana, að einhverju leyti vegna ólíkra aðferða, tímasetninga, lengdar könnunartímabila og birtingartafa auk þess sem fylgið sjálft breytist milli kannana.

Í kennslubókum má lesa hvernig best sé að gera kannanir og í fullkomnum heimi er gerlegt að fylgja þeim leiðbeiningum til hlítar. Sum ofangreindra atriða geta verið rétt í þeim skilningi að þau fylgi slíkri forskrift. Það er því hægt að gagnrýna einstakar kannanir eða jafnvel könnunaraðila fyrir að gera suma hluti ekki rétt meðan önnur atriði eru þess eðlis að þau hljóta alltaf að vera álitamál, jafnvel í fullkomnum heimi. Hér er því þó varpað fram að oft séu þessi atriði

stórlega ofmetin og í reynd sé hrein óvissa, breytileiki í niðurstöðum þrátt fyrir óbreytt fylgi og óbreyttar aðferðir, ráðandi fyrir niðurstöður skoðanakannana.

Það er hollt að velta fyrir sér hugtakinu kjörfylgi. Hvað myndi ég kjósa ef kosið væri til borgarstjórnar núna, *í mars 2006*? Mikilvæg og gagnleg spurning sem ég er vís til að svara samkvæmt bestu samvisku. En *ef* það væru kosningar núna í mars, væri kosningabaráttan og stjórnsmálaumræðan væntanlega í hámarki. Hvað ég myndi kjósa núna ef það væru *raunverulega* kosningar er hugsanlega ekki það sama og ég tel að ég myndi kjósa í kosningum sem eru aðeins hugsanlegar en ekki raunverulegar. Fylgi stjórnsmálahreyfingar á kjörtímabilinu er hugtak með ákveðna merkingu en ekki sömu merkingu og fylgið sem kemur upp úr kjörkössunum við lok þess. Því lengra frá kjördegi sem könnun er gerð, þeim mun minna á niðurstaðan sameiginlegt hugtakinu kjörfylgi sem verður aðeins metið á kjördag eða á tímamarki mjög nálægt kosningum.

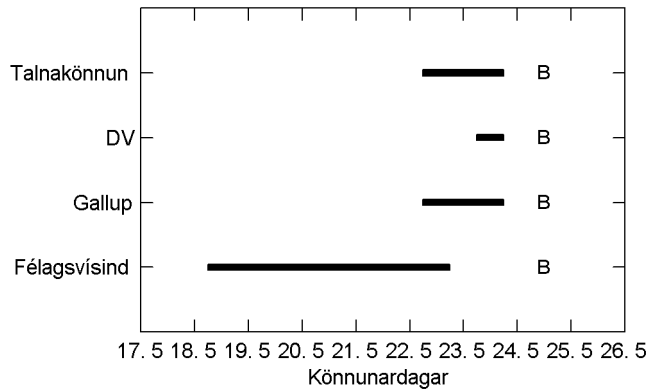
Hér verður fjallað um fylgibreytingar eins og þær birtast í skoðanakönnunum og þær ályktanir sem draga má um þær. Samkvæmt ofangreindu höfum við ekkert ytra viðmið um rétt fylgi óháð könnunum nema ef vera kynni að við ímyndum okkur risastóra könnun í tilviljunarúrtaki kosningabærra manna þar sem allir svara, spurningarnar eru rétt orðaðar og rétt aðferð notuð að öllu leyti. Þrátt fyrir dæmigert orðalag spurninga getur samanburðurinn ekki verið við kosningar. Hvernig rétt spurning hljóðar getur verið umdeilanlegt og sannarlega er breytilegt frá einum könnunaraðila til annars hvaða orðalag eða aðferð er notuð.

Síðustu kosningaspár vorið 2002

Til að sjá vandann í hnotskurn getum við litið á síðustu kosningakannanir vorið 2002. Þetta er ákjósanlegt dæmi, því viðmiðunin er skýr í niðurstöðum kosninganna þann 25. maí. Kannanirnar birtust allar föstudaginn 24. maí og því nær sem niðurstöðurnar eru því kjörfylgi sem fæst daginn eftir, því trúverðugri eru kannanirnar og þær aðferðir sem könnunaraðilarnir beita.

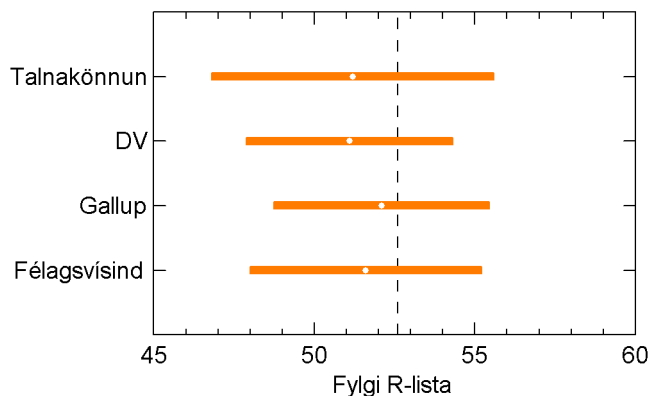
Mynd 1 sýnir fyrsta vandann sem við stöndum frammi fyrir. Þótt allar kannanirnar birtist á sama degi, eru þær greinilega misgamlar og taka yfir mislöng tímabil. DV er með yngstu upplýsingarnar þar sem könnunin var gerð daginn áður. Talnakönnun og Gallup gerðu sínar kannanir á tveimur dögum og upplýsingarnar því eins til tveggja daga gamlar. Félagsvísindastofnun gerði sína könnun á fimm dögum og þar sem birtingartöfin var á annan sólarhring voru upplýsingarnar eins til sex daga gamlar.

Stjórnsmál og stjórnsýsla vef tímarit (fræðigreinar)



Mynd 1. Tímasetning, tímalengd og birtingardagur síðustu kosningakannana.

Ef við gerum ráð fyrir því að fylgið sé að breytast þessa síðustu viku fyrir kosningar, gætu þessar tímasetningar skipt máli. Kannanirnar birtast allar sama daginn og því er þessi aldursmunur upplýsinganna ekki áberandi í umfjölluninni. Það er ýmis annar munur á þeim aðferðum sem notaðar eru. Sumar kannanir byggja á tilviljunarúrtaki úr þjóðskrá en aðrar byggja á símaskrá. DV spyr einnar spurningar en aðrir spyrja þá óákveðnu hvað sé líklegast að þeir muni kjósa. Misjafnt er hvaða dagsetningu spurningin miðast við: DV spyr um borgarstjórnarkosningar sem væru núna, Talnakönnun um kosningar sem væru í dag, Félagsvísindastofnun spyr um kosningar sem væru á morgun og Gallup spyr beinlínis hvað viðkomandi hafi hugsað sér að kjósa á kjördag.



Mynd 2. Fylgi R-lista samkvæmt síðustu kosningakönnunum.

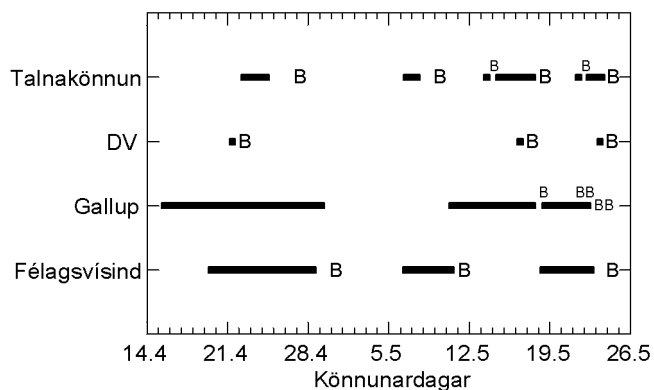
Miðað við allan þennan breytileika, hversu nálægt kosninganiðurstöðum komast kannanirnar? Mynd 2 birtir niðurstöður kannananna fyrir R-lista með 95% öryggisbilum.¹ Kosningaúrslitin, 52,6% fyrir R-lista, eru sýnd með brotinni línu til samanburðar (Hagstofa Íslands, 2005). Þar sem þýðisgildið – kosningaúrslitin – er þekkt, er túlkun öryggisbilanna tiltölulega einföld; lítil frávik frá kosningar-

¹ Síðasta könnun Fréttablaðsins (2002, 25. maí) gaf ekki upp úrtaksstærð og er því ekki með í þessum samanburði.

úrslitunum eru líklegri en mikil. Þrátt fyrir ólík könnunartímabil, mislanga birtingartöf, ólíkar úrtaksaðferðir og mismunandi spurningar, verður tæpast gert upp á milli niðurstaðnanna. Allir könnunaraðilar vanmeta fylgi listans en miðað við óvissuna í niðurstöðunum – lengd öryggisbilanna – verður tæpast gert upp á milli könnunaraðila. Stærsta frávikið er hjá DV sem vanmetur fylgið um eitt prósentustig en í samanburði við sex og hálfis prósentustigs breitt öryggisbil er það mjög viðunandi niðurstaða. Til samanburðar vanmetur Gallup niðurstöðuna um hálf prósentustig en þar sem öryggisbilið er svipað og hjá DV er óvissan mikil í niðurstöðunni og því hugsanlega heppileg tilviljun sem skilaði svo litlu fráviki frá kosningaúrslitum.

Almennt má segja að allar kannanir hafi farið mjög nálægt kosningaúrslitum en óvissan sé almennt mjög mikil. Á kosningadegi höfum við ekki ástæðu til að treysta einni könnun umfram aðrar. Að úrslitum fengnum má álykta að frávikin séu að stærstum hluta vegna úrtakadreifingar (*sampling distribution*), þ.e. önnur könnun sama könnunaraðila hefði gefið aðra niðurstöðu með, eftir atvikum, umtalsvert meira eða minna fráviki frá kosningaúrslitum.

Sambæfing ólíkra kannana í einni spá

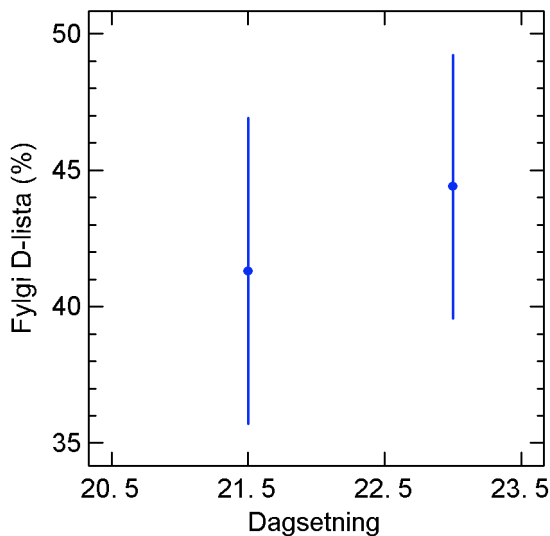


Mynd 3. Könnunartímabil og birtingardagur skoðanakannana.

Mynd 3 sýnir að könnunartímabil skoðanakannana eru mjög mislöng. Sumar kannanir standa yfir í aðeins einn dag og gefa því vitneskju um stöðu mála á þeim tímapunkti. Aðrar kannanir standa yfir í allt að hálfan mánuð og gefa því eins konar meðaltal fyrir það tímabil. Í umfjöllun er þetta gjarnan lagt að jöfnu, þrátt fyrir að upplýsingarnar geti verið allt að vikugamlar eða jafnvel að hluta til rúmlega hálf mánaðar gamlar sé miðað við upphaf tímabilsins.

Jafnvel þótt litið sé til sambærilegra kannana skapar óvissan í niðurstöðunum mikinn vanda. Mynd 4 sýnir niðurstöður tveggja sambærilegra kannana með öryggisbilum. Í könnun sem var gerð 21. maí sögðust 41,3% þátttakenda myndu kjósa D-lista ef „kosið væri í dag.“ Í könnun sem gerð var næstu tvo daga á eftir kom hins vegar í ljós að 44,4% þátttakenda sögðust myndu kjósa D-

lista. Hvað segja þessar niðurstöður einar og sér um raunverulegar fylgibreytingar D-lista yfir þetta tímabil?

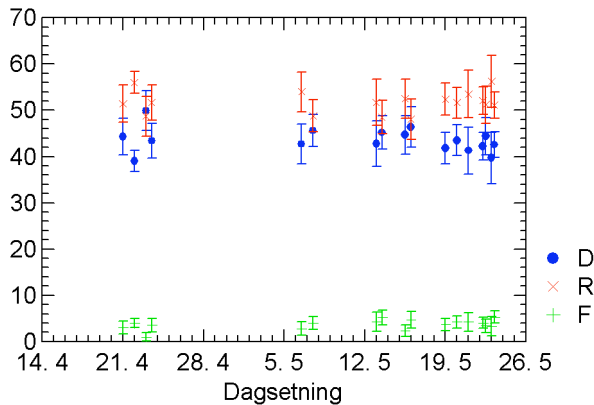


Mynd 4. Fylgi D-lista samkvæmt tveimur síðustu könnunum Talnakönnunar.

Við getum litið á niðurstöðurnar sjálfar og dregið þá ályktun að fylgið hafi aukist um rúm þrjú prósentustig á um einum og hálfum sólarhring.² Við getum framkvæmt marktektarpróf og komist að því að okkur sé ekki stætt á því að fullyrða að það sé raunveruleg breyting á fylgi D-listans yfir þetta tímabil. Að síðustu getum við reiknað 95% öryggisbil fyrir fylgibreytingu D-listans á tímabilinu og dregið þá ályktun að sennileg fylgibreyting sé frá tæplega 4 prósentustiga minnkun í fylgi upp í um 10 prósentustiga aukningu á aðeins einum til tveimur sólarhringum. Þótt niðurstöður kannananna sýni vaxandi fylgi og marktektarpróf segi að ekki sé hægt að fullyrða að fylgið sé ólíkt á þessum tveimur tímupunktum, þá getur minnkandi fylgi engu að síður samræmst þessum niðurstöðum.

Við getum alhæft þessar niðurstöður á svipuð tilvik þar sem bornar eru saman einstakar kannanir. Óvissan í skoðanakönnunum er af þeirri stærðargráðu að einfaldur samanburður niðurstaðna, jafnvel þegar kannanir eru sambærilegar að öllu leyti, gerir okkur ekki kleift að fullyrða hvort fylgi hafi aukist, sé óbreytt eða jafnvel hafi minnkað.

² Fyrri könnunin var framkvæmd á einum degi en sú síðari á tveimur og því álitamál hvar innan dagsins eigi að staðsetja kannanirnar.



Mynd 5. Niðurstöður kannana með 95% öryggisbilum staðsettar á miðjum könnunartímabilum.

Ef lítið er til marga kannana samtímist breytist myndin lítillega. Mynd 5 sýnir niðurstöður kannana með 95% öryggisbilum, staðsettar á miðjum könnunartímabilunum. Grófa myndin er nokkuð ljós. F-listi er með langminnst fylgi eða að jafnaði um þrjú til fjögur prósent. D-listi er greinilega að jafnaði um átta prósentustigum fyrir neðan R-listann. Að þessu leytnu gefur það skýrari mynd að skoða margar kannanir samtímis. Ef við hins vegar viljum lesa nákvæmari upplýsingar um fylgi á ákveðnum tímum eða fylgisþróun yfir tíma, verður niðurstaðan óljósari. Það hversu breið öryggisbilin eru gerir erfitt að meta fylgi listanna á ákveðnum tímamarkum. Það er einnig erfitt að meta hvort fylgismunur D- og R-lista sé að aukast eða minnka á ákveðnum tímabilum. Svo virðist sem fylgismunurinn gæti verið í lágmarki á tímabilinu 12. til 18. maí en óvissan er það mikil að það er erfitt að meta hversu miklu munar á fylkingunum á þessum tíma eða nákvæmlega hvenær lágmarki var náð.

Ályktanir sem dregnar eru nú þegar

Óvissa í niðurstöðum einstakra kannana ætti að leiða til varkárni í samanburði einstakra kannana. Það virðist þó ekki alltaf fyrir hendi heldur virðast könnunaraðilar, fréttaskýrendur og þátttakendur kosninga á stundum tilbúnir að bera saman einstakar kannanir eða draga ályktanir um fylgisbreytingar.

Könnunaraðilar hafa augljósa hagsmuni af umfjöllun um kannanir. Einkennandi er að oftast er aðeins fjallað um eigin niðurstöður og þær bornar saman innbyrðis þrátt fyrir að aðrir aðilar hafi birt kannanir á sama tímabili. Á þessu eru þó heiðarlegar undantekningar, svo sem pistill Benedikts Jóhannssonar (Könnun í Reykjavík: Bilið að lokast, 2002) sem bar saman niðurstöður kannanna frá Talnakönnun, Gallup og DV.

Samanburður einstakra kannana er algengur þótt ekki sé alltaf ljóst hvort átt sé við að raunverulegt fylgi hafi breyst eða staðið í stað, hvort eingöngu sé vísað sé til *niðurstaðna* kannana eða hvort yfirhöfuð sé gerður greinarmunur á þessu

tvennu. Dæmi um slíkan einfaldan samanburð niðurstaðna er t.d. Hnífjafnt með D- og R-lista (2002), Gallup (2002b) og Átta prósentustig skilja milli R- og D-lista (2002). Miðað við óvissu kannana ætti þetta að vera undantekning fremur en reglan í umfjöllun um kjörfylgi samkvæmt skoðanakönnunum.

Dæmi um hið gagnstæða er frétt Ríkisútvarpsins þar sem því er haldið fram daginn fyrir kosningar að enginn viti hvort fylgi fylkinganna tveggja sé á upp- eða niðurleið (Gallup: R-listinn með 52,1% fylgi, 2002). Ríkisútvarpið byggði þetta á raðkönnunum Gallup. Þetta er skynsamleg niðurstaða ef horft er til einstakra kannana en í reynd var sennilega að draga sundur með D- og R-lista á þessum tímapunkti eins og sést þegar tekið er tillit til allra kannana sem gerðar voru síðustu vikurnar fyrir kosningar.

Frambjóðendur tjá sig lítið í rituðu máli um niðurstöður kannana. Björn Bjarnason er þar nokkur undantekning (Björn Bjarnason, 2002). Þar telur hann hafa verið 10–17 prósentustiga mun á R- og D-lista fram að laugardeginum 18. maí en þann dag hafi ekki verið neinn „tölfræðilegur munur“ á listunum tveimur og því fylgissveifla yfir til Sjálfstæðisflokksins. Það má skilja Björn þannig að það hafi orðið miklar fylgissveiflur þarna rétt fyrir hvítasunnuna, fylgið hafi streymt til flokksins en farið til baka til R- og F-listans fram að kosningum 25. maí.

Fréttaskýrendur og blaðamenn eru eðlilega reiðubúnir til að fjalla um fylgið og fylgisbreytingar frá degi til dags. Fréttablaðið (Sveiflukenndri baráttu lýkur í dag, 2002) lýsir miklum fylgissveiflum og gefur yfirlit yfir helstu skoðanakannanir (þó ekki DV). Hér eru kannanir allt frá janúar 2002 skoðaðar og því hugsanlega rétt metið að miklar fylgisbreytingar hafi orðið. Það er annað mál hvort um sveiflur sé að ræða.

Jónas Kristjánsson (2002) telur miklar og marktækar sveiflur hafa orðið á fylgi framboða á síðustu tveimur vikum kosningabaráttunnar. Hann líkir kjósendum við þeytisþjöld en bendir á að einungis 10–20 prósent kjósenda skipti um skoðun á síðustu stigum baráttunnar. Í sama tölublaði gefur Fréttablaðið (Mikil hreyfing á fylginu, 2002) yfirlit yfir fylgisbreytingar síðustu fjögurra vikna kosningabaráttunnar. Þar er stiklað á helstu könnunum og bent á samræmi og ósamræmi. Heildarniðurstaðan virðist vera að miklar fylgisbreytingar hafi orðið, væntanlega á grundvelli einstakra kannana. Óljóst er hvenær munur R- og D-lista er talinn hafa náð lágmarki en hvítasunnuhelgin, 19. og 20. maí, greinilega talin vera vendipunktur.

Á þriðjudegi eftir hvítasunnu telur Ásgeir Friðgeirsson (2002) Sjálfstæðisflokkinn í mikilli sókn. Hann horfir til greiningar Benedikts Jóhannssonar (Könnun í Reykjavík: Bilið að lokast, 2002) og telur þær kannanir sem voru næst hvítasunnuhelginni sýna minnstan mun milli fylkinga. Á þessum tímapunkti hefur Ásgeir auðvitað ekki aðgang að könnunum Gallup og Talnakönnunar sem báðar sýna vaxandi yfirburði R-lista yfir D-lista.

Aðferðir sem er beitt til að álykta um fylgisbreytingar

Það eru ýmsar aðferðir notaðar til að draga ályktanir um fylgisbreytingar. Algengasta aðferðin virðist vera að bera saman einstakar kannanir ýmist óformlega eða með tölfræðiprófum. Dæmi um slíkan einfaldan samanburð niðurstaða er t.d. Hnífjafnt með D- og R-lista (2002), Gallup (2002b) og Átta prósentustig skilja milli R- og D-lista (2002).

Slíkur samanburður hefur augljósa annmarka. Skoðanakannanir spyrja lítinn hluta kjósenda í því skyni að álykta um fylgi framboða meðal allra. Niðurstöðurnar eru því háðar mikilli óvissu, því meiri sem fjöldi svarenda er minni. Þetta þýðir að ólíkar niðurstöður geta hæglega verið samfara litlum sem engum fylgisbreytingum og sömu eða svipaðar niðurstöður geta fengist þrátt fyrir verulegar fylgissveiflur.

Það þarf ekki að vera neitt rangt við það þegar könnunaraðili tilkynnir að færri eða fleiri styðji tiltekið framboð heldur í síðustu könnun sömu aðila, svo lengi sem vísað er eingöngu í þá sem svara, þ.e. niðurstöður kannananna sem slíkra fremur en sjálf fylgið.

Hér má oft sjá dálítinn línudans í frásögnum af niðurstöðum kananna. Við mat á þeim þarf að hafa annars vegar í huga hvaða skilning almennur lesandi leggur í orðin og hins vegar hversu skýrt hugsunin sjálf er orðuð. Fullyrðing Talnakönnunar að það sé „nokkru minni munur ... en var í síðustu könnun...“ (R-listinn með 8 gegn 7, 2002) er t.d. alveg skýr og það fer ekki milli mála að vísað er í niðurstöðurnar sjálfar. Samkvæmt Morgunblaðinu 24. maí hefur hins vegar „munur á fylgi Reykjavíkurlistans og Sjálfstæðisflokksins ... aukist á nýjan leik“ þar sem fylgismunur var áður um 8%, fór niður í rúm 3% en var kominn aftur upp í 8% samkvæmt könnunum Félagsvísindastofnunar (Átta prósentustig skilja á milli R- og D-lista, 2002). Svo gæti vírst sem niðurstöður þriggja kannana nægi til að álykta um fylgisbreytingar yfir fimm vikna tímabil þrátt fyrir að hver og ein niðurstaða sé óviss. Á sama hátt má skilja frétt DV (Hnífjafnt með D- og R-lista, 2002) þar sem fullyrt er að „D-listi fengi 46,4 prósentu fylgi ef kosið væri nú en R-listi 48 prósent.“

Ýmsar ástæður geta verið fyrir slíkum túlkunum. Að einhverju leyti er þetta spurning um orðalag. Vandí blaðamannsins er einnig verulegur sem fær kosningafrétt sem verkefni án þess að hafa þá talnaþekkingu sem þarf til. Það sem þó hugsanlega skiptir sköpum er sú tilhneiging fjölmiðla og könnunaraðila að kynna sínar eigin kannanir og ræða um þær eins og engar aðrar upplýsingar liggi fyrir. Það getur verið stutt frá þeirri hugsun að aðrar kannanir skipti ekki máli yfir í það að þessi tiltekna könnun upplýsi sannleikann til fulls.

Stundum er leitast við að samhæfa óformlega niðurstöður ólíkra kannana. Dæmi um þetta er greining Benedikts Jóhannssonar (Könnun í Reykjavík: Bilið að lokast, 2002) á fylgisbreytingum í vikunni fyrir hvítasunnuhelgi. Tvær samantektir Fréttablaðsins eru einnig dæmi um slíkar óformlegar samantektir (Mikil hreyfing á fylginu, 2002; Sveiflukenndri baráttu lýkur í dag, 2002). Slíkur samanburður er yfirleitt gerður án þess að beita neinum tölfræðilegum nálgunum

heldur leitast við að leggja mat á þróunina og leita óformlegra skýringa á frávikum. Þannig er í samantekt Fréttablaðsins á fylgisbreytingum (Mikil hreyfing á fylginu, 2002) minnst á könnun Talnakönnunar sem sýndi D-lista með meirihluta yfir R-lista seinni hluta aprílmánaðar og lagt mat á niðurstöðu hennar: „Í ljósi annarra kannana var þessi niðurstaða dregin í efa.“

Niðurstöður kannana geta verið mótsagnakenndar. Síðari hluti aprílmánaðar er dæmi um þetta en þá voru gerðar fjórar kannanir. Þessar kannanir spönnuðu allt frá einum degi upp í tvær vikur en þyngdarpunktur þeirra allra var á tímabilinu 21. til 23. apríl. Fjöldi svarenda var á bilinu 340 til 790 manns eftir könnunum.

Í yfirliti yfir niðurstöður kannana kemst Fréttablaðið að skynsamlegri niðurstöðu: „Niðurstöður þeirra voru býsna ólíkar“ (Mikil hreyfing á fylginu, 2002). Samkvæmt könnunum DV og Félagsvísindastofnunar var fylgi R-listans 7–8 prósentustig umfram D-listann, Gallup taldi yfirburðina vera um 17 prósentustig en Talnakönnun taldi D-lista hafa um tveggja prósentustiga yfirburði.

Þrátt fyrir þessar ólíku niðurstöður geta allar kannanirnar verið að lýsa sama raunveruleikanum og engin þeirra þarf að vera beinlínis röng. Könnun Gallup sker sig dálítið úr. Niðurstaðan í könnun Talnakönnunar er hins vegar ekkert sérkennileg, jafnvel þó að ólíkt hinum sé niðurstaða hennar D-listanum í vil.

Sams konar sveiflur verða á niðurstöðum einstakra kannana allar fimm síðustu vikur kosningabaráttunnar. Benda má á kannanir Félagsvísindastofnunar og Talnakönnunar sem hafa þyngdarpunkta á 5. og 6. maí og kannanir sem voru gerðar vikuna 12. til 18. maí. Slíkt misræmi í niðurstöðum gerir erfitt um vik að samhæfa niðurstöður óformlega; ósamræmi niðurstaðnanna gæti því verið það sem lá til grundvallar hinni almennu umræðu um fylgissveiflur í aðdraganda kosninganna.

Algeng leið er að vísa til öryggisbila eða marktektarprófa. Þetta hefur á sér yfirbragð vísinda þar sem byggt er á mælingum og formlegum ákvarðanareglum tölfræðinnar. Þannig vísar Ásgeir Friðgeirsson (2002) til þess að munur fylkinga sé innan vikmarka og því sé hann ekki mældur með viðunandi vissu. Hann dregur þá ályktun af þessu og fleiru að hratt dragi saman með listunum. Aðrir benda á að kannanir hafi „náð saman á skekkjumörkunum“ á tímabilinu 15.–30. apríl og bil framboða því minnkað mikið (Mikil hreyfing á fylginu, 2002). Hér virðist ranglega ályktað að skörun öryggisbila feli í sér að skortur sé á tölfræðilegri marktekt (Schenker & Gentleman, 2001).

Öryggisbil gefa til kynna talnabil sem sennilega inniheldur „rétt“ fylgi viðkomandi framboðs. Það er mikilvægt að gera sér grein fyrir að öryggisbil eru lýsandi og vísa aðeins til viðkomandi könnunar. Bilið segir því að ef aðeins niðurstöður viðkomandi könnunar liggja fyrir en aðrar ekki, þá sé raunverulegt fylgi sennilega einhvers staðar innan öryggisbilsins.

Oftast liggja fyrir niðurstöður nokkurra kannana sem gerðar hafa verið á svipuðum tíma. Öryggisbil einnar könnunar er því ekki mjög lýsandi fyrir þá óvissu sem er raunverulega fyrir hendi um fylgið þar sem það horfir fram hjá upplýsingum annarra kannana. Hver könnun um sig hefur öryggisbil sem metur

nákvæmni viðkomandi könnunar en þau hver um sig eru villandi sem mælikvarði á óvissuna þegar allar kannanirnar koma saman.

Stundum er vísað beint í tölfræðilega marktækt eða skort á henni. Bent er á að munur á fylgi fylkinga sé marktækur (Átta prósentustig skilja á milli R- og D-lista, 2002), hafi verið marktækur en sé það ekki lengur (Kannanir: Enn minnkar bilið í borginni, 2002) eða kannanir náí „saman á skekkjumörkunum“ og því hafi bil milli framboða minnkað (Mikil hreyfing á fylginu, 2002). Björn Bjarnason (2002) bendir á að ekki sé lengur „tölfræðilegur munur“ á könnunum en áður hafi hann verið töluverður. Einstakar kannanir eru bornar saman og munurinn talinn ómarktækur (Gallup, 2002c; Könnun í Reykjavík – Er bilið að lokast, 2002; R-listinn með 8 gegn 7, 2002) og jafnvel gengið svo langt að bera saman síðasta mann inn hjá annarri fylkingunni saman við næsta mann inn hjá hinni (Gallup, 2002c, 2002d).

Slíkar aðferðir hafa alla galla marktæktarprófunar. Marktækt milli tveggja kannana er túlkuð eins og prófuð sé ein tiltekin tilgáta með einu prófi þegar í reynd er valið eitthvert eitt par kannana af þeim könnunum sem mögulegt væri að bera saman án þess að leiðrétta niðurstöðuna á nokkurn máta (Klockars & Sax, 1986). Einnig virðist skortur á marktækt vera túlkaður sem vitnisburður um að engin breyting hafi orðið á fylgi milli tímabila og þannig litið fram hjá því að núlltilgátum má hafna en erfiðleikum bundið að staðfesta þær (Schmidt, 1996).

Það má velta fyrir sér rökvisi tilgátuprófunar hvort sem horft er til þess að fylgi breytist yfir tíma og því óhugsandi að núlltilgátan um enga breytingu geti verið rétt eða til fylgismunar í kosningum sem hlýtur alltaf að verða einhver (Cohen, 1994).³ Þótt núlltilgátan sé aldrei rétt strangt til tekið þegar niðurstöður kannana eru bornar saman, gæti þó mismunurinn verið mjög óverulegur (Seaman & Serlin, 1998). Tölfræðipróf getur því verið ómarktækt sökum þess að fylgismunurinn sé óverulegur eða afköst prófsins of lítil jafnvel fyrir umtalsverðan mun. Skortur á marktækt getur hins vegar ekki verið sökum þess að fylgismunur sé enginn. Við getum heldur ekki ályktað að marktækt feli í sér verulegan og ómarktækt próf óverulegan fylgismun; prófun núlltilgátu og mat á sennilegu fráviki frá henni eru tvö ólík úrlausnatriði sem fara ekki saman í einni og sömu tölfræðiaðgerðinni.

³ Atkvæði féllu jöfn í Bolungarvík í sveitarstjórnarkosningum vorið 2002. Í jafnstóru sveitarfélagi og Reykjavíkurborg er sú niðurstaða nánast óhugsandi þótt hægt sé að ímynda sér að atkvæði falli það jafnt að til vandræða horfi við talningu. Í forsetakosningunum í Bandaríkjunum haustið 2000 féllu atkvæði hnífjafnt í Flórídafylki. Kjarni vandans þar var þó ekki sá að frambjóðendur hefðu nákvæmlega sama atkvæðafjölda, heldur að svo litlu munaði að ýmis framkvæmdaratriði við talningu urðu til þess að vekja deilur um niðurstöðu kosninganna.

Samhæft mat á fylgi framboðslista

Ljóst er að þær aðferðir sem almennt eru notaðar til að fjalla um niðurstöður skoðanakannana eru ófullnægjandi. Nokkrar leiðir má þó fara til að mynda samhæft mat á fylgi og fylgisþróun.

Helgi Tómasson (2005) lýsir Bayes-nálgun við að álykta um fylgi út frá núverandi upplýsingum þar sem tekið er tillit til eldri niðurstaðna. Hann reiknar út það vægi sem eldri kannanir fá í þeim tilgangi að fá sem nákvæmast mat á núverandi fylgi byggt á samantekt yfir niðurstöður bæði nýjustu og eldri skoðanakannana. Með þessu er tekið tillit til þess að niðurstöður einstakra kannana eru óvissar en óvissuna megi minnka með því að hafa hliðsjón af öðrum eldri könnunum.

Helgi virðist miða alfarið við spá fram á við í tíma, þ.e. tekið er tillit til eldri upplýsinga þegar staðan er metin eins og hún er í dag eða í framtíðinni. Þetta er í hæsta máta eðlilegt þegar við viljum meta fylgi fylkinga frá degi til dags og viljum fá sem nákvæmast mat á hverjum tíma. Ef við viljum hins vegar meta fylgið eins og það var áður, horfir slík uppfærsla eldri niðurstaðna fram hjá hluta upplýsinganna. Ef ég vil vita hvert fylgi D-lista var t.d. 13. maí 2002, gefa skoðanakannanir þann dag og skoðanakannanir þann 7. og 8. maí mér umtalsverðar upplýsingar. En þar sem ég fer aftur í tímann, hef ég einnig aðgang að nýrri skoðanakönnunum sem ekki lágu fyrir þann 13. maí, þ.e. skoðanakannanir á tímabilinu 14. til 23. maí. Þessar upplýsingar sem koma á eftir í tíma eru jafngagnlegar og upplýsingar sem koma fyrir í tíma en það er aðeins þær síðarnefndu sem tekið er tillit til við uppfærslu eldri upplýsinga með nýjum eins og gert er í hefðbundinni Bayes-greiningu.

Í þessu felst einnig að horft er fram hjá leitni yfir tíma, fylgisþróuninni. Síðustu skoðanakannanir vorið 2002 voru gerðar 23. maí, tveimur dögum fyrir kosningar. Til að meta sennilegt fylgi á kjördag þurfum við að framlengja mat okkar á fylgi fylkinga um tvo daga. Búast má við að fylgi sumra fylkinga sé að aukast, annarra að minnka og hugsanlega standi fylgi sumra því nær í stað. Fyrir fram virðist það trúlegt að slík leitni feli í sér mikilvægar viðbótarupplýsingar sem við viljum nýta okkur við mat á fylginu á kjördag.

Taka má tillit til beggja þessara atriða í Bayes-líkani með því að gera það ögn flóknara. Nákvæmlega hvaða aðferð er notuð við að samhæfa niðurstöður skoðanakannana er þó hugsanlega ekki mikilvægasta álitaeefnið, svo lengi sem aðferðin hefur þá eiginleika sem slíkt mat þyrfti að hafa. Aðferðin þarf að taka tillit til stærðar kannana, meta niðurstöður einnar könnunar með hliðsjón af öðrum sem gerðar eru á svipuðum tíma þar sem nálægar kannanir hafa meiri áhrif en fjarlægjar og taka tillit til þess hvenær könnun er framkvæmd óháð því hvenær hún birtist.

Óvissa kannana fer að mestu eftir stærð þeirra þótt hlutfall þeirra sem fylgja viðkomandi lista hafi einnig áhrif. Stærstu kannanirnar fyrir borgarstjórnarkosningarnar 2002 voru með um 900 þátttakendur og þær minnstu með rúmlega 300. Stærstu kannanirnar eru því um 75% nákvæmari en þær minnstu. Litlar

kannanir eru algengastar eða á bilinu 300 til 600 þátttakendur en þar munar samt um 40% á nákvæmni kannana. Almennt séð er erfitt að taka tillit til óvissu með óformlegum hætti og því þarf það að gerast kerfisbundið með stærðfræðilegum aðferðum.

Eins og áður var bent á er erfitt að draga ályktanir um fylgi fylkinga á grundvelli stakra kannana, til þess eru þær einfaldlega of ónákvæmar. Eðlilegasta nálgunin er að horfa til niðurstaðna annarra kannana. Rétt er að líta bæði til fyrri kannana – hvað hafa kannanir fram að þessu sýnt – en einnig til ókominna kannana. Almennt séð eru niðurstöður sem eru í samræmi við aðrar kannanir nálægar í tíma, traustari en niðurstöður sem eru í ósamræmi. Því meir sem ein könnun vísar frá niðurstöðum annarra nálægra kannana, því sennilegra er að niðurstaðan sé tilkomin vegna tilfallandi óvissu í niðurstöðunum. Traustasta niðurstaðan fæst því með því að leiðrétta slíkar niðurstöður „inn á við,“ þ.e. í átt að niðurstöðum annarra kannana.

Ef margar kannanir eru gerðar á sama tíma ættu þær að veða jafnt í hlutfalli við nákvæmni sína við mat á fylgi framboða á þeim tíma. Fyrri kannanir og þær sem síðar koma ættu að veða minna og því minna eftir því sem þær eru lengra frá í tíma. Fjarlægð í tíma ætti hins vegar að hafa sömu áhrif hvort sem hún er til baka eða áfram í tíma. Kannanir sem eru mjög fjarlægjar í tíma ættu hins vegar engin áhrif að hafa.

Með þessum hætti er hægt að nýta saman sem flestar kannanir í einu samhæfðu fylgismati á tilteknum tímamarki. Eftir því sem lengra líður frá viðkomandi tímamarki, hvort sem það er til fortíðar eða framtíðar, því meiri líkur eru á því að niðurstöður kannana séu ósambærilegar vegna fylgisbreytinga.

Hér má staldra við. Svo gæti virst að með því að taka tillit til kannana sem eru nálægt í tíma sé verið að jafna út skammtímasveiflur í fylgi. Með því að veða kannanir síðustu og næstu daga inn í samhæft fylgismat er dregið úr þeim breytileika sem er í niðurstöðum kannana og þannig jafna út allar upplýsingar um skammtímasveiflur.

Slíkum andmælum má svara með því að benda á þá óvissu sem áður hefur verið lýst varðandi niðurstöður einstakrar könnunar. Sveiflur milli kannana eru því að mestu leyti tilkomnar vegna tilfallandi áhrifa óvissu en ekki vegna skammtímabreytinga á fylgi framboða. Til að geta fylgt eftir breytingum á fylgi þarf því að jafna út þennan breytileika milli kannana. Ef við viljum geta séð breytingar í fylgi frá einum degi til annars, getum við ekki stuðst við niðurstöður núverandi skoðanakannana. Við samhæfingu nokkurra kannana saman í einu fylgismati er því ekki verið að tapa upplýsingum um skammtímasveiflur heldur öðlast upplýsingar um raunverulegar breytingar fylgis frá einum tíma til annars.

Aðferð staðbundinnar aðfallsgreiningar

Hér er kynnt ein þeirra aðferða sem hafa þessa æskilegu eiginleika en hún er ýmist nefnd jöfnun (*smoothing*; Jacoby, 1997), úrtaksbundin aðfallsgreining (*non-*

parametric regression; Fox, 2000) eða staðbundin aðfallsgreining (*local regression*; Cleveland, Grosse & Shyu, 1992). Síðasta nafngiftin er nákvæmust þar sem hún lýsir aðferðinni sem slíkri en ekki notkunarviði, þ.e. hagnýtingu hennar og fjölmargra annarra aðferða við að jafna út villu í tengslum tveggja eða fleiri breyta. Staðbundin aðfallsgreining hefur þann kost umfram t.d. aðferðir sem byggja á reglu Bayes að vera stærðfræðilega einfaldari og vera til staðar í misumfangsmiklum útfærslum í algengum tölfræðiforritum svo sem SPSS (e.d.), Systat (e.d.) og R (The R Project for Statistical Computing, e.d.).⁴

Staðbundin aðfallsgreining byggist á þeirri hugmynd að þótt fylgisbreytingar yfir tíma séu flóknari en svo að línuleg aðfallsgreining nái að lýsa þeim, séu breytingar yfir stuttan tíma mun einfaldari. Ef litið er til stutts tíma, t.d. breytingu frá einum degi til annars, sé fylgi framboðslista ýmis að aukast, minnka eða standa í stað og sé því lýsanlegt með línulegri eða annars stigs aðfallsgreiningu. Ef horft er til nægilega stutts tímabils sé því hægt að gera líkan af leitni fylgisins yfir tíma nægjanlega vel til að geta metið fylgið á ákveðnum tímamarki. Með því að endurtaka þetta fyrir nægilega marga tímamarkta fæst mat á flókinni ólínulegri fylgisþróun jafnvel þótt á hverjum tímamarki fyrir sig sé gert ráð fyrir því að leitni fylgi einfaldri línulegri eða annars stigs aðfallslínu.

Staðbundna aðfallsgreiningu má nota til einfaldrar jöfnunar (*smoothing*) fylgnirita þar sem mögulegum ólínulegum tengslum tveggja breyta er lýst. Hana má einnig nota í flókni aðfallsgreiningum þar sem skilyrðisrit (*conditioning plots*; *coplots*), sneiðing (*slicing*) samfelldra breyta og athugun og jöfnun leifar er beitt til að uppgötva flókin og hugsanlega ólínuleg tengsl breytanna (Cleveland, 1993).

Staðbundin aðfallsgreining byggist því á röð aðfallsgreininga sem hver um sig er ófullnægjandi ef litið er til alls tímabilsins en lýsir þróuninni á fullnægjandi hátt þegar litið er til tímamarktsins sem er til athugunar. Þannig má til dæmis hugsa sér að fylgi R-lista hafi verið vaxandi 19. maí 2002 og þeim vexti megi lýsa sem ákveðinni fylgisaukningu á hverja tímaeiningu, án þess að í því felist að sama aukning eigi sér stað daginn eftir eða að fylgið sé nauðsynlega vaxandi á þeim degi.

Við getum hugsað okkur mat á fylgisþróun þannig að línuleg aðfallsgreining sé gerð fyrir hvern dag hvort sem könnun hafi verið gerð fyrir þann dag eða ekki. Niðurstaða greiningarinnar gefur okkur mat á fylginu þann daginn og með því að endurtaka þetta fyrir hvern dag tímabilsins og tengja línu milli niðurstaðnanna fæst yfirlit yfir þróun fylgisins. Matið fyrir hvern dag fyrir sig byggist á því að framkvæma aðfallsgreiningu þar sem niðurstöður nálægra kannana vega þungt en vægi þeirra minnkar eftir því sem þær eru fjarlægari í tíma. Með þeim hætti er tekið tillit til allra upplýsinga sem liggja fyrir en þess gætt að taka mest mark á nálægum könnunum og minna mark á þeim sem eru fjarlægari í tíma.

⁴ SPSS býður upp á einfalda jöfnum fylgnirita með loess-aðferð Cleavelands. Systat er með SMOOTH, sem er tiltölulega þróað forrit til staðbundinnar aðfallsgreiningar. R hefur mikinn fjölda af ólíkum forritum fyrir staðbundna aðfallsgreiningu, meðal annars loess í stat-safninu.

Lykilatriðið í staðbundinni aðfallsgreiningu er að ákvarða hve margar nálægar kannanir eigi að taka með þegar fylgið er matið á hverjum tímapunkti. Ef matið miðast við margar kannanir, fæst nákvæmt mat á fylginu á þeim tímapunkti þar sem óvissa einstakra kannana jafnast því betur út sem fleiri þeirra eru teknar saman.⁵ Á móti kemur að hættan á skekkju (*bias*) eykst þar sem kannanir sem eru fjarlæggar í tíma hafa áhrif á matið; matið gæti verið kerfisbundið of hátt eða of lágt. Ef við lítum á hinn bóginn til aðeins fárra nálægra kannana minnkar skekkjan en ónákvæmnin eykst. Þetta er klassískt viðfangsefni tölfræðinnar, að vega nákvæmni á móti skekkju. Í staðbundinni aðfallsgreiningu birtist það í því að við viljum að matið á hverjum tímapunkti byggist á sem fæstum könnunum en þó nægilega mörgum til að fá nauðsynlegan stöðugleika í matið.

Ýmsar leiðir eru til að velja fjölda kannana en algengast er að skoða leifina, þ.e. frávik frá línunni sem metur fylgisþróunina samkvæmt niðurstöðum staðbundnu aðfallsgreiningarinnar. Þá er byrjað á því að nota tiltölulega fáar kannanir í matinu og athugað hvort hægt sé að sjá einhverja leitni yfir tíma í leifinni, síðan er fjöldi kannana aukinn smátt og smátt þar til öll leitni er horfin og leifin birtist sem tilviljunarbundinn breytileiki í kringum fylgisþróunina (Cleveland, 1993).

Auðvelt er að taka tillit til mismikillar óvissu vegna ólíks fjölda í skoðana-könnunum. Stærðin er ráðandi þáttur í óvissu kannana og því má einfaldlega vega hverja könnun með stærð hennar. Þar sem nákvæmnin eykst með kvaðratrótinni af fjöldanum er eðlilegast að miða við að hver könnun sé vegin í hlutfalli við kvaðratrót af þátttakendafjölda.

Eins og aðrar tölfræðiaðgerðir byggir staðbundin aðfallsgreining á ákveðnum forsendum. Línuleg leitni fylgisþróunar þegar litið er til nægilega stutts tíma er í reynd hagnýt einföldun fremur en eiginleg forsenda. Ef fylgisþróunin er ólínuleg yfir tímabil sem nær yfir daga og vikur, verður skekkjan mjög lítil þótt gert sé ráð fyrir því að þróunin sé línuleg á hverjum tímapunkti fyrir sig. Þetta er því hagkvæm nálgun sem gerir kleift að lýsa leitninni yfir lengra tímabil. Við þurfum að gera ráð fyrir því að við val á fjölda kannana sem stuðst er við á hverjum tímapunkti sé nákvæmni vegin rétt á móti skekkju, þ.e. nægjanlega nákvæmt mat fái á leitninni yfir tíma án þess að skekkja hana svo neinu nemi. Ef meta á nákvæmni niðurstöðunnar, t.d. með öryggisbilum í kringum matsniðurstöðurnar, þarf að gera ráð fyrir því að frávikin frá leitnarlínunni dreifist jafnt sama hvar á tímabilinu það er skoðað. Að síðustu þurfum við að gera ráð fyrir því að allar kannanir séu jafngildar að teknu tilliti til stærðar þeirra, þ.e. meti fylgi framboðanna óháð því hver könnunaraðilinn er og hvenær könnunin er framkvæmd.

Rétt er að staldra við síðustu forsenduna. Er Gallup að gera jafngóðar kannanir og DV, sem gerir jafngóðar kannanir og Heimur.is sem sjálfar eru jafngóðar og kannanir Fréttablaðsins? Slíkur jöfnuður fæst auðvitað ekki staðist.

⁵ Hér er hugtakið „nákvæmt“ notað í tölfræðilegri merkingu, þ.e. breytileiki yfir mörg úrtök úr sama þýði.

Hér er hins vegar tekin sú afstaða að niðurstaða stakrar könnunar sé svo ónákvæm að það sé ekki sjálfgefið að breytileiki í aðferðum eða aðrir eiginleikar sem tengjast könnunaraðilum vegi þungt. Þegar litið er til dæmigerðra öryggisbila sem eru sex til níu prósentustig á breidd, sést að það er óvissan sem er ráðandi fremur en kerfisbundin skekkja hjá einstökum könnunaraðilum. Ef könnunaraðili er óvandvirkur þannig að breytileiki í niðurstöðum sé verulega meiri en stærð kannana gefur til kynna, ætti slíkt að jafnast út þegar litið er til marga kannana. Ef hins vegar niðurstaðan er kerfisbundið röng, tilhneiging til ofmats eða vanmats á ákveðnum framboðslistum, þarf sú tilhneiging að vera mjög mikil til að skipta máli fyrir ónákvæmar niðurstöður kosningakannana. Ef einhver könnunaraðili t.d. ofmetur eitthvert framboð að jafnaði um þrjú til fjögur prósentustig, gæti það lítillaga skekkt niðurstöður á tímamarkum þar sem kannanir frá þeim aðila eru ráðandi. Slík kerfisbundin skekkja ætti þó að birtast sem óvenjumikil frávik frá fylgisþróuninni, leifinni, fyrir þennan könnunaraðila. Það er því mögulegt að meta frammistöðu einstakra könnunaraðila og sjá hvort frávik frá metinni fylgisþróun eru meiri en búast mætti við miðað við óvissu vegna stærðar kannananna.

Í þessari rannsókn er leitast við að meta þróun fylgis fyrir R- og D-lista í aðdraganda borgarstjórnarkosninga vorið 2002. Leitast er við að lýsa fylgisþróuninni síðustu fimm vikur fyrir kosningar, hversu hraðar breytingar urðu á fylgi og hvort meirihluti R-lista hafi á einhverjum tímamarki verið í hættu. Einnig verður athugað hvernig spá tvo daga fram í tímann frá síðustu könnunum fram að kjördegi gagnast. Að síðustu verður athugað hvort kerfisbundinn munur er á niðurstöðum eftir könnunaraðilum.

Aðferð

Safnað var skoðanakönnunum sem birtar voru opinberlega á tímabilinu 22. apríl til 25. maí 2002. Samtals voru þetta 19 kannanir sem birtust í Fréttablaðinu, DV, Morgunblaðinu, vefsvæði Ríkisútvarpsins og vefsvæðinu Heimur.is. Fimm ólíkir könnunaraðilar unnu kannanirnar: DV, Félagsvísindastofnun, Fréttablaðið, Gallup og Talnakönnun.

Könnun Fréttablaðsins sem birtist 25. maí gaf engar upplýsingar um fjölda svarenda. Þar sem könnunin hafði umtalsverð áhrif á niðurstöður útreikninga þótti ekki stætt á því að sleppa henni úr gagnasafninu. Hún var því tekin með en úrtaksstærðin áætluð 300 manns með hliðsjón af algengri stærð og svarhlutfalli kannana hjá Fréttablaðinu en þess um leið gætt að vegna óvissunnar væri vægi könnunarinnar í lægri kantinum. Kannanir Gallup sem birtust á tímabilinu 21. til 24. maí voru svonefndar raðkannanir, samtals fjórar kannanir þar sem hver könnun náði yfir tveggja til þriggja daga tímabil sem skaraðist um einn dag við bæði næstu könnun á undan og þá næstu á eftir. Með því að taka aðeins fyrstu könnunina og þá síðustu fengust tvær óháðar kannanir sem spönnuðu annars vegar 18. til 20. maí og hins vegar 22. til 23. maí og því töpuðust einungis upplýsingar um 21. maí frá því sex daga tímabili (18.–23. maí)

sem raðkannanirnar fjórar spönnuðu. Eftir stóðu þá 17 óháðar kannanir sem spönnuðu tímabilið 21. apríl til 23. maí 2002.

Fyrir hverja könnun var skráð hvenær könnunin birtist, könnunartímabilið, fjöldi þeirra sem tóku afstöðu og hlutfallslegur fjöldi sem valdi hvern framboðslista. Lítið samræmi var í upplýsingagjöf milli könnunaraðila og framsetning með ólíkum hætti. Þannig var hlutfallslegt fylgi framboðslista gefið upp með misjafnlega mörgum aukastöfum, jafnvel milli kannana hjá sama könnunaraðila. Fjöldi þeirra sem tók afstöðu var sjaldnast gefinn upp og þurfti því að meta hann út frá öðrum upplýsingum, þ.e. þátttakendafjölda og hlutfallslegum fjölda þeirra sem svöruðu. Þannig mátti meta allnákvæmlega fjölda þeirra sem tóku afstöðu nema í síðustu könnun Fréttablaðsins þar sem láðist að gefa upp fjölda þátttakenda.

Óvissan í hverri könnun var metin miðað við normalnálgun út frá fjölda þeirra sem tóku afstöðu og hlutfallslegum fjölda sem valdi hvert framboð. Hefðbundin lýsandi öryggisbil voru reiknuð miðað við 95% öryggi.

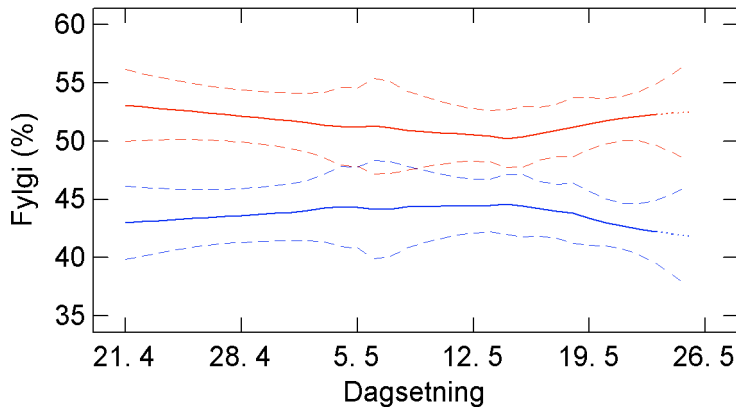
Staðbundin aðfallsgreining með loess-aðferð (Cleveland, 1993) var notuð til að meta fylgisþróun fyrir R- og D-lista. Með því að fylgja leiðbeiningum Clevelands um athuganir á leitni leifar var bandbreiddin 0,6 valin sem sú sem gæfi besta jafnvægi (*trade off*) nákvæmni og skekkju. Til að fá eðlilegra vægi kannana þar sem nákvæmari kannanir vægju meira en þær ónákvæmari, var niðurstaða hvernar könnunar vegin með kvaðratrót af fjölda þeirra sem afstöðu tóku. Þetta ætti einnig að stuðla að samleitni (*homoscedasticity*) villunnar.

Óvissan í metinni fylgisþróun var ákvörðuð með öryggisbili um matstölur fylgislínunnar (*point wise confidence interval*) sem reiknað var með aðferð Cleveland, Grosse og Shyu (1992). Þar sem tímabil síðustu kannana náðu aðeins til 23. maí var matslínan framlengd með loess-aðferð fram til kjördags 25. maí 2002.

Til að kanna breytileika skoðanakannana var frávikum frá matslínunni breytt í staðlaða leif með því að deila staðalvillu hvernar könnunar í frávik hennar frá matslínunni.

Niðurstöður

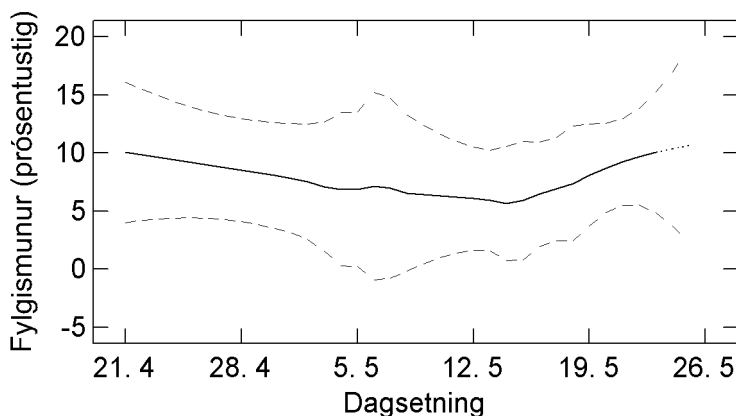
Mynd 6 sýnir fylgisþróun R- og D-lista samkvæmt staðbundinni aðfallsgreiningu. Brotalínur sýna 95% öryggisbil fyrir matsgildi línanna (*point wise confidence interval*).



Mynd 6. Metin fylgisþróun fyrir R- (efri lína) og D-lista (neðri lína). Brotalína gefur til kynna 95% öryggisbil fyrir matsgildi línanna.

Samkvæmt myndinni var R-listi með um 53% fylgi í upphafi tímabilsins en fylgið minnkaði smám saman og var komið niður í rúm 50% 14. maí 2002. Fram að kjördag jókst fylgi R-lista aftur og var samkvæmt niðurstöðunum komið aftur upp í 52,5% á kjördag. Til samanburðar var raunverulegt kjörfylgi 52,6%. Athygli vekur hve hægar fylgisbreytingar voru því á tímabilinu 21. apríl til 14. maí minnkaði fylgið um 0,1 prósentustig á dag og hækkaði síðan um 0,2% fram á kjördag.

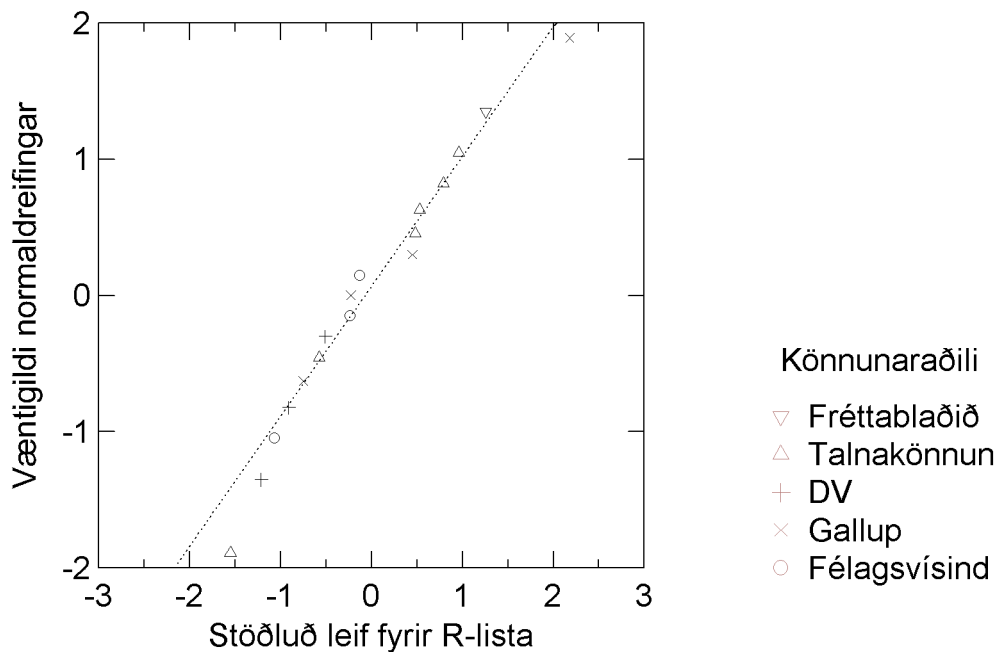
Fylgisþróun D-lista var nánast samhverfa þessa. Í upphafi tímabilsins var fylgið um 43%, jókst um tæpt 0,1 prósentustig á dag og var orðið 44,6% 14. maí. Síðan minnkaði það aftur um tæplega 0,3 prósentustig á dag og var metið 41,9% á kjördag. Til samanburðar fékk D-listi 40,2% á kjördag.



Mynd 7. Metinn fylgismunur á R- og D-lista með 95% öryggisbili matsgilda.

Mynd 7 sýnir þróun á fylgismun R- og D-lista yfir tímabilið. Í upphafi tímabilsins munaði 10,1 prósentustigi á listunum en munurinn minnkaði fram að 14. maí og var þá 5,6 prósentustig. Síðan jókst fylgismunurinn jafnt og þétt fram á kjördag og var þá metinn 10,6 prósentustig. Til samanburðar munaði 12,4 prósentustigum á kjörlistunum tveimur í borgarstjórnarkosningunum 25. maí 2002.

Aðferðin sem notuð var byggist á því að leggja allar kannanir að jöfnu að teknu tilliti til óvissu vegna úrtaksstærðar. Því er mikilvægt að skoða vel niðurstöður eftir könnunaraðilum. Ef það er svo að verulegur kerfisbundinn munur sé á niðurstöðum skoðanakannana eftir því hver gerir þær, veikir það niðurstöðurnar á Mynd 6 og Mynd 7 verulega. Hér verður að hafa hugfast að án efa er munur á niðurstöðum ólíkra könnunaraðila þar sem umtalsverður munur er á aðferðum svo sem úrtaksgerð, lengd tímabila og spurninga. Til þess að skipta máli þurfa þessi frávík þó að vera umtalsverð samanborið við þá miklu óvissu sem tengist einstökum könnunum.

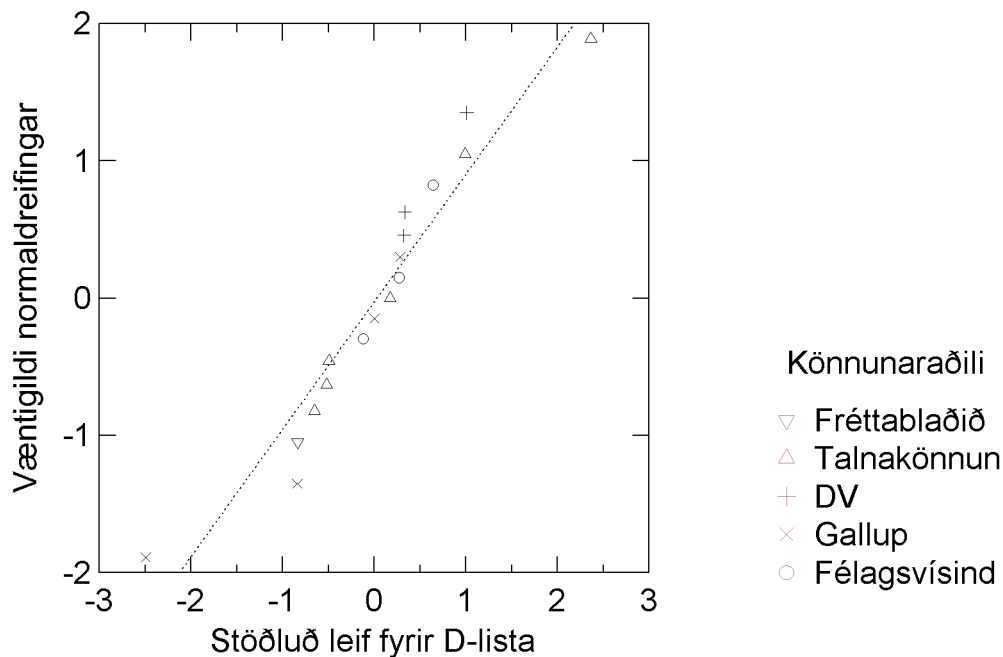


Mynd 8. Normalrit fyrir staðlaða leif R-lista.

Normalrit af staðlaðri leif úrvinnslunnar gerir kleift að fá glöggt yfirlit yfir helstu eiginleika í frávíkum einstakra kannana þar sem öll áhugaverð frávík birtast sem frávík frá beinni línu ritsins (Wang & Bushman, 1998). Mynd 8 sýnir staðlaða leif fyrir R-listann. Þar sem fylgishlutfallið er það nálægt 50% og þátttakendafjöldi það mikill, gerum við ráð fyrir því að leifin normaldreifist. Öll frávík frá beinu línunni benda til frávika frá normaldreifingu. Hver könnunaraðili er með sitt tákn á myndinni og ef umtalsverður munur er á niðurstöðum ættu þeir að raðast kerfisbundið ólíkt á beinu línuna. Þess sér ekki stað. Miðað við aðeins 17

kannanir er ekki hægt að krefjast þess að könnunaraðilar dreifist allir nánast jafnt á myndina en stórvægileg frávik gefa tilefni til efasemda. Slík frávik finnast einfaldlega ekki á myndinni.

Mynd 9 gefur sömu upplýsingar fyrir D-listann. Hér sjáum við að ein könnun Gallup hefur mælt fylgi D-lista óvenjulágt og önnur könnun frá Talmakönnun hefur mælt fylgið óvenjuhátt. Þetta eru önnur og þriðja könnun tímabilsins sem voru báðar í töluverðu ósamræmi við aðrar kannanir á sama tíma. Aðrar kannanir frá sömu aðilum virðast dreifast að mestu eðlilega um línuna. Dreifing annarra kannana er heldur minni en við var búist sem styður það að þetta séu þrátt fyrir allt tilfallandi frávik.



með könnunum almennt. Ef eitthvað í úrtaksgerð, spurningum eða úrvinnslu leiðir til áberandi frávíks frá öðrum könnunum, er það væntanlega grandskoðað. Sökum mikilla fjárhagslegra hagsmuna og mikilvægi álit á vettvangi markaðarins, er sennilegt að aðferðirnar séu í ýmsum smáatriðum afurð langrar reynslu sem hafi skilað ótvíræðum árangri. Í þessu felst það álit að ýmis skólabókarleg atriði sem í reynd sé óhugsandi að uppfylla skipti minna máli heldur en uppsöfnuð reynsla könnunaraðila af því hvaða frávik frá ströngustu forsendum könnunaraðferðanna séu afdrifarík fyrir niðurstöðurnar og hver þeirra séu í reynd lítilvæg hagkvæmnisatriði.

Athygli vekur að það verður ekki séð að meirihluti R-lista hafi nokkurn tíma verið í hættu. Öryggisbilið fyrir fylgismuninn fer að vísu niður fyrir núll í kringum 6. til 7. maí, aðallega vegna ónákvæmni í matinu á þeim tímamarki. Fylgismunurinn er hins vegar í lágmarki þann 14. maí og þá eru neðri mörk bilsins um 0,7%. Það er því ákveðin óvissa um yfirburði R-lista á tímabilinu 5. til 15. maí eða þar um bil. Slík óvissa táknar ekki að það sé vafi um yfirburði R-listans á því tímabili heldur einungis að við gætum ekki fullyrt gegn öðrum vísbendingum að fylgi R-listans hafi verið meira en D-listans. Í heildina tekið benda niðurstöðurnar þvert á móti til þess að meirihluti R-lista hafi aldrei verið í verulegri hættu, fylgisýfirburðir hans hafi farið niður í tæp sex prósentustig í kringum 14. maí en síðan farið hækkandi fram á kjördag.

Fylgisþróunin er áberandi hæg. Þetta er í andstöðu við almenna umræðu á þessum tíma. Þótt erfitt sé að henda reiður á umræðu sem að miklu leyti er í töluðum orðum, lögðu margir áherslu á fylgissveiflur (Jónas Kristjánsson, 2002; Mikil hreyfing á fylginu, 2002), aðrir reyndu að tengja breytingar við einstaka viðburði svo sem umræðuþætti í sjónvarpi (Ásgeir Friðgeirsson, 2002; Könnun í Reykjavík: Bilið að lokast, 2002) eða jafnvel framsetningu fjölmiðla á hverjum tíma (Björn Bjarnason, 2002). Þessi stöðugleiki í fylgni er í samræmi við athugun Wlezien og Erikson (2002) á fylgisþróun í aðdraganda forsetakosninga í Bandaríkjunum. Niðurstaða þeirra er að fylgisbreytingar minnki mjög mikið þegar kosningabaráttan hefst og séu tiltölulega litlar síðustu tvo mánuði fyrir kosningar en þó mismiklar eftir kosningum.

Ef hægt er að gera ráð fyrir því að fylgi ráðist í senn af djúpstæðum lífskoðunum, mati á trúverðugleika einstakra frambjóðenda, mati á langtíma-hagsmunum kjóсандans og tryggð við ákveðin stjórnsmálaöfl, verður að teljast hæpið að einstakir atburðir eins og góð frammistaða í umræðuþætti eða einstök atriði í framsetningu og túlkun álitsgjafa skili sér í hröðum fylgissveiflum. Þetta virðist þó lítt kannað rannsóknarsvið hérlendis. Það væri t.d. mikilvægt að vita hvort umtalsverð fylgisaukning verði í kringum flokksþing eins og stundum heyrir, hversu mikil sveiflan er og hvenær hún hjaðnar. Það myndi gefa gagnlegar upplýsingar um sennileg áhrif minni atburða eins og umræðuþátta. Miðað við fylgisþróunina sem hér birtist má ætla að áhrif þeirra séu lítil eða að þau birtist hægt og blandist saman við almenn heildaráhrif kosningarbaráttunnar. Til samanburðar kemst Erikson (2001) að því að kappræður fyrir bandarísku forsetakosningarnar haustið 2000 hafi haft áhrif á fylgi frambjóðenda. Svo

virðist sem áhrifin séu allt að 3 prósentustig (mynd 5, bls. 41) en standi aðeins fáa daga. Slíkar breytingar á raunfylgi myndu ekki sjást í þeirri úrvinnslu á fylgi í aðdraganda borgarstjórnarkosninga sem hér er greint frá.

Eitt framlag þessarar rannsóknar er að samhæfa niðurstöður margra kannana í einu mati. Það gefur trúverðugri mynd af fylginu eins og það er á hverjum tímapunkti en gefur einnig yfirlit yfir fylgisþróunina sem ekki fæst með öðrum hætti. Mikilvægasta framlagið felst þó sennilega í þeirri minnkun á óvissu sem fæst með samhæfingu margra kannana. Þann 23. maí voru síðustu kannanir fyrir borgarstjórnarkosningar gerðar. Vikmörk (*margin of error*) fyrir fylgi D-lista miðað við 95% öryggi voru á bilinu 3,2 til 5,5 enda voru á bilinu 300 til rúmlega 900 þátttakendur í hverri könnun. Til samanburðar er matspunktur fylgislínunnar með vikmörkin 2,7 þann dag. Yfir tímabilið sem athugað var voru vikmörk matspunkta á bilinu 1,9 til 4,4 meðan einstakar kannanir voru með vikmörk á bilinu 3,2 til 5,5.

Annað mikilvægt framlag felst í því að líkanið gefur færi á að spá fyrir um niðurstöður kosninga. Strangt til tekið gefa einstakar kannanir ekki færi á því að meta fylgi á kjördag. Niðurstöður þeirra miðast eingöngu við könnunartímabilið, það er engin aðferð sem gerir okkur kleift að framlengja þær niðurstöður fram að kjördegi eða að meta óvissu matsins á kosninganiðurstöðunum. Við getum borið niðurstöður síðustu kannana við kosninganiðurstöður svipað og hér var gert með Mynd 2. Á kjördag eru þessar niðurstöður orðnar tveggja til þriggja daga gamlar og í reynd ógerningur að meta hversu samanburðarhæfar þær eru við kosninganiðurstöður og engin aðferð til að meta óvissu þeirra sem forspárniðurstöður fyrir kosningarnar. Til samanburðar gefa niðurstöður þessarar rannsóknar bæði forspá fram í tímann og þá óvissu sem fylgir spánni. Óvissan eykst réttilega þegar reiknað er fram úr gögnunum (*extrapolation*). Þannig eru 95% vikmörk fyrir fylgi D-lista 2,7 prósentustig þann 23. maí, hækka í 3,5 24. maí og eru komin í 4,4 prósentustig á kjördag.

Eðlilegt mat á óvissu er mikilvægt ef við viljum nota niðurstöður kannana til að ákvarða sennilegan sætafjölda fyrir hvern framboðslista. Slíkt mat er skynsamlegast ef litið er til metins fylgis á kjördag, þar sem fylgi milli kosninga skilar ekki sætum. Þar sem úthlutun sæta fer eftir ákveðnum reglum sem eðlilega skila einungis heilum sætum, er mikilvægt að niðurstöðurnar taki tillit til óvissu í fylgismatinu og séu á formi dreifingar á sennilegum sætafjölda fyrir hvern framboðslista (Ólafur Þ. Harðarson og Indriði H. Indriðason, 2005). Til að slíkt sé gerlegt þarf að lágmarka óvissuna í matinu, þ.e. nota eitthvert líkan sem gerir kleift að meta niðurstöður fram í tímann og sem gefur upp óvissu matsins. Sú aðferð sem notuð var í þessari rannsókn uppfyllir þetta.

Veikleiki staðbundinnar aðfallsgreiningar fyrir það verkefni sem hér var tekist á hendur er sá að litlar takmarkanir eru settar á fylgisþróun yfir tíma. Það leiðir meðal annars til þess að þegar þróunin er framlengd fram til kjördags eykst óvissan mikið með hverjum degi sem reiknað er fram yfir síðustu kannanir. Með hliðsjón af þeim hægu fylgisbreytingum sem hér hafa verið leiddar í ljós, væri æskilegt að þróa líkanið áfram þannig að skýrari takmarkanir

séu settar á þróun fylgis yfir tíma. Það myndi leiða til nákvæmari forspár og sennilega minni óvissu fyrir matspunktanna á fylgislínunni. Þannig er sennilega nægjanlegt að gera ráð fyrir því að fyrir borgarastjórnarkosningar vorið 2002 hafi fylgið þróast eftir vægri sveiglinu frá því seinni hluta apríl og fram til 14. maí, þar hafi komið brot í þróunina og hún snúist við og síðan þróast með svipuðum hætti fram að kjördegi. Ef hægt væri að gera ráð fyrir slíkum einföldum tengslum fylgis og tíma, með einum eða fleiri brotpunktum, væri stigið stórt skref í þá átt að geta gefið glöggar upplýsingar um fylgisþróun og sennilega niðurstöðu nokkrum dögum fyrir kjördag.

Umræða um skoðanakannanir, fylgisþróun og sennilegt kjörfylgi í aðdragannda borgarstjórnarkosninga vorið 2002 var ómarkviss, illa rökstudd og byggði á takmörkuðum forsendum. Hún var gjarnan á forsendum einstakra frambjóðenda eða könnunaraðila. Þegar hún snerist um einstakar prósentutölur fengnum úr tilteknum skoðanakönnunum var hún einna traustust og reis sjaldan hærra en það. Ef framlag þessarar rannsóknar verður að hnika umræðunni aðeins eina ögn í rétta átt, þá var vel af stað farið.

Heimildir

- Ásgeir Friðgeirsson (2002, 21. maí). *Björn saxar á forskot Ingibjargar Sólrúnar*. Vefsíða: www.strik.is/frettir/efni.ehtm?id=3003&cat=frettir. Sótt: 21. maí 2002.
- Átta prósentustig skilja milli R- og D-lista (2002). *Morgunblaðið*, bls. 6.
- Björn Bjarnason (2002). [Pistill án titils á heimasíðu Bjarnar.] Vefsíða: www.bjorn.is/safn.view.php?id=795&list=greinar. Sótt 27. maí 2002.
- Cleveland, W.S., Grosse, E., & Shyu, M.-J. (1992). A package of C and Fortran routines for fitting local regression models.
Vefskjal: <http://netlib.bell-labs.com/netlib/a/cloess.ps.gz>. Sótt: 16. maí 2006.
- Cleveland, W.S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997–1003.
- Erikson, R.S. (2001). The 2000 presidential election in historical perspective. *Political Science Quarterly*, 116, 29–52.
- Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots*. Sage university paper series on quantitative applications in the social sciences, nr. 130. Beverly Hills and London: Sage.
- Gallup (2002b). *Reykjavík: Töluverðar breytingar*.
Vefsíða: gallup.is/main/view.jsp?branch=429688&e175=453388&nonCached=1. Sótt: 21. maí 2002.
- Gallup (2002c). *Reykjavík 20.–21. maí: Svipuð staða*.
Vefsíða: gallup.is/main/view.jsp?branch=429688&e175=458812&nonCached=1. Sótt: 23. maí 2002.
- Gallup (2002d). *Reykjavík 21.–22. maí: Óbreytt bil milli R og D*.
Vefsíða: gallup.is/main/view.jsp?branch=429688&e175=469208&nonCached=1. Sótt: 24. maí 2002.
- Gallup: *R-listinn með 52,1% fylgi* (2002). Vefsíða: www.ruv.is [Nánara veffang óþekkt. Frétt dagsett: 24.05.2002 17:41]. Sótt: 24. maí 2002.
- Hagstofa Íslands (2005). *Hagtiðindi*, 90 (15).
- Helgi Tómasson (2005). Spágildi skoðanakannana. *Tímarit um raunvísindi og stærðfræði*, 3. Vefskjal: <http://www.raust.is/2005/1/04/raust2005-1-04.pdf>. Sótt 15. mars 2006.

Hnílfjafnt með D- og R-lista (2002).

Vefsíða: www.strik.is/dv/frettir/efni.ehtm?id=3561&cat=Fréttir. Sótt 17. maí 2002.

Jacoby, W. G. (1977). *Statistical graphics for univariate and bivariate data. Sage university paper series on quantitative applications in the social sciences, nr. 117*. Beverly Hills and London: Sage.

Jónas Kristjánsson (2002, 24. maí). Kjósendum sveiflað. *Fréttablaðið*, bls. 8.

Kannanir: Enn minnkar bilið í borginni (2002).

Vefsíða:

<http://www.heimur.is/index.php?frettir=single&newsflokkur=Kannanir&fid=545>. Sótt 22. maí 2002.

Kannanir: R-listinn að stinga af? (2002).

Vefsíða:

<http://www.heimur.is/index.php?frettir=single&newsflokkur=Kannanir&fid=550>. Sótt 22. maí 2002.

Klockars, A.J., & Sax, G. (1986). *Multiple comparisons. Sage university paper series on quantitative applications in the social sciences, nr. 61*. Beverly Hills and London: Sage.

Könnun í Reykjavík: Bilið að lokast (2002).

Vefsíða: www.heimur.is/index.php?frettir=single&newsflokkur=Kannanir&fid=546. Sótt 21. maí 2002.

Mikil hreyfing á fylginu (2002, 24. maí). *Fréttablaðið*, bls. 8.

Ólafur Þ. Harðarson og Indriði H. Indriðason (2005). Hvenær verður minnihluti atkvæða að meirihluta fulltrúa? *Stjórnsmál og stjórnsýsla: Vef tímarit, 1* (1), bls. 3–21.

The R Project for Statistical Computing (e.d.). Vefsíða: www.r-project.org. Sótt 8. maí 2006.

R-listinn með 8 gegn 7 (2002).

Vefsíða: www.heimur.is/?frettir=single&newsflokkur=kannanir&fid=69. Sótt: 24. maí 2002.

Schenker, N., & Gentleman, J.F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55*(3), 182–186.

Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115–129.

Seaman, M.A.; & Serlin, R.C. (1998). Equivalence confidence intervals for two-group comparisons of means. *Psychological Methods, 3*, 403–411.

SPSS (e.d.). Vefskjal: www.spss.com/spss. Sótt: 8. maí 2006.

Sveiflukenndri baráttu lýkur í dag (2002, 25. maí). *Fréttablaðið*, bls. 8.

Systat (e.d.). Vefskjal: www.systat.com/products/Systat/. Sótt: 8. maí 2006.

Wang, M.C. & Bushman, B.J. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods, 3*, 46–54.

Wlezien, C., & Erikson, R.S. (2002). The timeline of presidential election campaigns. *The Journal of Politics, 64*, 969–993.

Viðauki

Birtingar- dagur	Þyngdar- punktur	Könnunar- aðili	Fjöldi	D	R	F
22.04.02	21.04.02	DV	490	44,3	51,4	3,0
— ^a	22.04.02	Gallup	792	39,3	56,3	3,9
26.04.02	23.04.02	Talnakönnun	362	49,9	48,3	1,3
30.04.02	23.04.02	Félagsvísindastofnun	339	43,4	51,7	3,5
08.05.02	06.05.02	Talnakönnun	361	42,7	54,0	2,8
11.05.02	07.05.02	Félagsvísindastofnun	480	45,6	48,8	4,0
14.05.02	13.05.02	Talnakönnun	380	42,8	51,7	4,3
18.05.02	13.05.02	Gallup	400	45,2	48,5	5,2
18.05.02	15.05.02	Talnakönnun	416	44,7	52,5	2,3
17.05.02	16.05.02	DV	506	46,4	48,0	4,7
21.05.02	19.05.02	Gallup	686	41,8	52,4	3,7
24.05.02	20.05.02	Félagsvísindastofnun	738	43,5	51,6	4,2
22.05.02	21.05.02	Talnakönnun	364	41,3	53,5	4,2
24.05.02	22.05.02	Gallup	854	42,2	52,1	4,0
24.05.02	22.05.02	Talnakönnun	499	44,4	51,2	3,4
25.05.02	23.05.02	Fréttablaðið	— ^b	39,7	56,2	3,3
24.05.02	23.05.02	DV	929	42,6	51,1	5,3

Þyngdarpunktur er meðaltal upphafs- og lokadags könnunartímabils gefinn hér upp í heilum dögum. *Fjöldi* er fjöldi svarenda sem tóku afstöðu og er metinn með hliðsjón af heildarfjölda í úrtaki, svarhlutfalli og hlutfalli þeirra sem tóku afstöðu. Byggt er á frumheimildum þar sem þær eru tiltækar. *D*, *R* og *F*: Fylgi viðkomandi kjörlista meðal þeirra sem tóku afstöðu.

^a Nákvæmur birtingardagur fyrstu Gallupkönnunar er óþekktur. ^b Enginn fjöldi var gefinn upp fyrir Fréttablaðskönnunina.