

연구논문

출구조사의 체계적인 예측 편향에 대한 분석:

2010년 지방선거 출구조사를 중심으로*

Systematic Forecasting Bias of Exit Poll:

Analysis of Exit Poll for 2010 Local Elections

김영원* · 최윤정***

Young-Won Kim · Yun-Jung Choi

본 연구에서는 선거 출구조사에서 발생하는 편향을 분석하기 위해, 먼저 2010년 전국동시지방선거 출구조사의 표본설계와 표본추출오차, 그리고 무응답 현황 및 예측오차 등을 살펴보고, 이를 토대로 출구조사에서 체계적으로 발생하는 지역별 편향 문제를 다루었다. 출구조사에서 발생하는 편향을 통계적으로 검증하기 위해 Martin et al.(2005)이 제안한 예측 정확성 척도인 통계량 A 를 사용하였다. 2010년 지방선거를 포함해 2006년 지방선거와 2007년 대통령 선거 방송사 출구조사 자료를 토대로 시·도 단위에서 지역별 편향을 분석해 본 결과, 여당 성향이 강한 지역에서는 여당 후보를 과대 추정하는 편향이 체계적으로 발생하고 있으며, 여당 성향이 강해질수록 이런 편향이 더 강해진다는 것을 확인할 수 있었다. 이런 연구결과는 향후 출구조사의 정확성 제고를 위한 방안을 모색하는 데 크게 기여할 수 있을 것으로 기대된다.

주제어: 출구조사, 응답 편향, 예측 정확성 척도, 2010 지방선거

In this paper, we overview the sample design, sampling error, non-response rate and prediction errors of the exit poll conducted for 2010 local elections and discusses how to detect a prediction bias in exit poll. To investigate the bias problem in exit poll in regional(Si-Do) level, we analyze exit poll data for 2007 presidential election and 2006 local elections as well as 2010 local elections in Korea. The measure of predictive accuracy A proposed by Martin et al.(2005) is used to assess the exit poll

- 3) 본 연구는 숙명여자대학교 2010학년도 교내연구비 지원에 의해 수행되었음.
- 4) 교신저자(corresponding author): 숙명여자대학교 통계학과 교수 김영원.
E-Mail: ywkim@sookmyung.ac.kr
- 5) 숙명여자대학교 통계학과 대학원생

bias. The empirical studies based on three exit polls clearly show that there exists systematic bias in exit poll and the predictive bias of candidates affiliated to conservative party (such as Hannara-Dang) is serious in the specific regions. The result of this study on systematic bias will be very useful to improving the exit poll methodology in Korea.

key words: exit poll, response bias, measure of predictive accuracy, 2010 local election

I. 서론

2010년 제5회 전국동시지방선거(이하, 2010 지방선거)는 그 동안의 선거 출구조사와는 달리 최초로 지상파 방송 3사가 공동조사에 합의하여 방송협회 산하에 KEP(Korea Election Pool)라는 공동예측조사위원회를 구성하여 수행하였다. KEP에서는 심사과정을 거쳐 3개 회사(밀워드브라운리서치/코리아리서치/TNS-R)를 조사기관으로 선정하여 선거예측 출구조사를 실시하였으며, 16개 시·도지사뿐만 아니라 교육감에 대한 선거예측이 이루어졌다. 이번 출구조사의 경우 16개 전 지역에서 시·도지사 당선자를 정확히 예측했을 뿐만 아니라 예측오차도 매우 작았다.

통계학적인 관점에서 출구조사의 정확성 평가를 위해서는 단순히 당선자를 예측했다는 것보다 1위와 2위 후보자의 득표율을 얼마나 정확하게 예측하였는지에 대해서 더 큰 관심을 갖게 된다. 선거예측조사의 정확성에 대한 분석은 적절한 예측오차의 산출을 통해 이루어진다. 흔히 생각할 수 있는 선거예측조사의 예측오차는 각 후보자에 대한 예측 득표율(\hat{p})과 실제득표율(p)의 차이, 즉 $(\hat{p} - p)$ 로 설명될 수 있을 것이다. 그동안 우리나라 출구조사 결과를 살펴보면, 특히 총선의 경우 $(\hat{p} - p)$ 가 통계이론으로 설명될 수 있는 표본추출오차(sampling error)의 수준을 벗어나는 경우를 많이 볼 수 있다. 이와 같이 출구조사에서 표본추출오차로 설명할 수 없는 오차가 특정 정당에 대해 양(+) 또는 부(-)의 방향으로 체계적으로 발생한다면 이는 해당 지역에서 발생한 특정 정당에 대한 편향(bias; 다른 연구에서는 'bias'를 편파라고 부르기도 함)으로 보는 것이 타당할 것이다. 그 동안 우리나라 방송사에서 실시한 출구조사와 관련해 표본추출오차로 설명될 수

없는 편향이 존재하고 있다는 점을 지적한 학계나 조사회사 전문가들이 많이 있었지만, 이런 현상을 통계적으로 분석한 사례는 찾아 볼 수 없다. 따라서 본 연구에서는 2010 지방선거 출구조사 자료를 중심으로 정당별 편향이 지역에 따라 어떤 패턴을 보이는지 구체적으로 분석해 보고자 한다.

선거예측조사를 수행하게 되면 응답거절이나 거짓응답 등 주로 비표본추출오차(non-sampling error)와 관련된 다양한 요인에 의해 발생하는 특정 정당에 대한 응답 편향(response bias)에 대해 많은 관심을 갖게 된다. 현경보(2005)는 “선거예측에서 편파(bias)는 특정 정당을 체계적으로 과대 또는 과소 평가하는 현상으로 당선자 예측에 결정적인 영향을 미치는 요인이다.”라고 설명하고 있다. 또한 이준웅(2004)은 17대 총선 선거 예측조사 사후분석을 통해 “지난 16대 총선과 마찬가지로 이번 선거에서도 표집이나 조사과정에 내재하는 무작위 오차가 문제였다기보다는 체계적으로 기울어지는 편파가 예측 실패의 주요한 원인이었던 것으로 보인다.”라고 지적하면서 우리나라 총선 예측조사에서 그 동안 지속적으로 발생해 온 ‘여당 과대평가’라는 체계적인 편향 문제를 다루고 있다. 이와 같이 선거예측에서 편향이라고 정의하는 예측오차의 방향성에 대한 관심은 계속되어 왔지만 편향에 대한 분석은 연구자의 직관이나 오랜 경험에 의한 주관적인 해석에 의존하는 경향이 있었기 때문에 계량적으로 이런 편향을 검증하기 위한 연구가 미흡한 것이 현실이다.

현경보(2005)는 “일반적으로 출구조사는 전화조사에 비해 예측력이 높은 것으로 알려져 있다. 하지만 출구조사 과정에서 응답자의 민감한 반응이나 응답 회피 등으로 인해 발생하는 비표본오차의 문제를 해결하지 못함으로써 특정 정당을 체계적으로 과소, 과대 예측하는 오류를 범하기도 한다. 따라서 선거예측조사에서 지역적 특성이나 응답자의 반응 등을 고려하지 않은 채 단순히 전화조사보다 출구조사의 예측력이 높다고 일반화하기는 어렵다고 본다.”라고 설명하고 있다. 하지만 김세용·허명희(2009)는 2007년 대통령 선거 관련 11개 전화여론조사 사례들에 대한 분석을 통해 우리나라에서 보편적으로 사용되는 할당추출에 의한 전화조사에서의 체계적 편향 문제를 다루었으며, 메타분석 결과 이명박 후보에 대한 지지율이 과다 추정되었다는 사실을 밝히고 있다.

한편 김영원·곽은선(2010)은 2008년 총선 출구조사 자료를 토대로 총조사오차라는 개념을 도입해 출구조사에서 발생하는 오차를 투표소 추출오차와 응답과정에서의 오차로 구분해 응답과정에서 발생하는 응답 편향 문제를 다루었다.

이렇게 특정 정당을 체계적으로 과소, 과대 예측하는 것을 조사의 편향 혹은 편파라

고 부르며, 본 연구에서는 우리나라 출구조사에서 지역별로 정당에 따른 편향이 실제로 존재하였는지 통계적으로 분석해 보기로 한다. 이를 위해 2010 지방선거 출구조사 자료를 중심으로 2006년 지방선거와 2007년 대통령선거 출구조사 자료를 함께 활용한다.

본 연구의 2절에서는 출구조사의 특성을 정리하기 위해 2010 지방선거 출구조사의 전반적인 개요와 더불어 성별·연령대별 응답률 및 응답자 구성비율 등에 대해 살펴본다. 3절에서는 2010 지방선거의 출구조사 예측결과와 실제결과의 차이를 지역별로 분석함으로써 예측오차 방향성, 즉 편향에 대한 과학적인 접근을 통해 출구조사에서 지역별 정치 성향에 따라 발생하는 특정 정당 후보자에 대한 편향 문제를 고찰해 볼 것이다. 또한 2006년 제4회 지방선거와 2007년 대통령선거 출구조사 자료를 이용해 지역별 정당 편향의 발생 여부 및 정당별 편향이 지역별로 어떤 패턴을 보이는지 종합적으로 분석해 본다.

II. 2010 지방선거 출구조사 현황

1. 투표소 표본 추출¹⁾

2010 지방선거 출구조사에서는 각 시·도별로 투표소를 지역적 인접성 등을 고려해 층화하고, 각 층 내에서 과거 선거에서 나타난 투표소별 정치적인 성향을 기준으로 정렬한 후 계통추출하는 층화 후 정렬계통추출법으로 표본 투표소를 추출하였다. 각 시·도별 투표소 모집단과 표본 투표소 현황은 <표 1>에 제시되어 있다.

<표 1> 시·도별 표본 투표소 추출 현황

지역		서울	부산	대구	인천	광주	대전	울산	경기
투표소수	모집단	2,198	871	575	617	345	335	279	2,714
	표본	50	35	25	40	25	40	25	55
지역		강원	충북	충남	전북	전남	경북	경남	제주
투표소수	모집단	655	467	734	619	870	950	933	226
	표본	41	40	53	25	35	37	40	41

1) 이 내용은 한국방송협회가 주관한 방송3사 공동예측조사위원회(Korea Election Pool)의 “2010 지방선거 방송사 공동 예측조사 평가보고서” 내용을 일부 발췌한 것임.

〈표 2〉 표본 응답자 성별·연령대별 구성비

(단위:%)

구 분		서울		부산		대구		인천	
		선관위	출구조사	선관위	출구조사	선관위	출구조사	선관위	출구조사
남	20대	8.9	6.2	8.2	4.7	8.9	5.4	8.3	5.2
	30대	9.6	9.4	6.6	6.6	6.9	6.4	9.0	8.5
	40대	9.9	10.8	9.5	9.1	10.3	11.3	12.1	11.2
	50대	9.8	11.0	11.1	12.7	10.1	12.0	10.3	12.3
	60대 이상	11.1	11.1	12.6	14.3	11.9	13.2	9.8	11.3
여	20대	8.6	8.2	6.9	6.0	6.8	6.3	7.9	7.2
	30대	10.1	10.5	8.2	8.0	9.0	8.1	10.3	10.0
	40대	10.4	11.3	11.1	10.5	11.8	12.5	12.4	12.0
	50대	10.3	10.9	12.1	13.4	10.6	12.2	9.6	11.2
	60대 이상	11.4	10.6	13.7	14.6	13.7	12.8	10.2	11.2
구 분		광주		대전		울산		경기	
		선관위	출구조사	선관위	출구조사	선관위	출구조사	선관위	출구조사
남	20대	9.0	5.1	8.8	5.9	8.6	5.0	8.6	4.7
	30대	8.4	8.0	9.0	8.7	8.9	9.3	9.9	9.1
	40대	12.1	11.6	12.1	11.9	13.9	14.9	12.8	14.2
	50대	9.4	10.7	10.4	12.3	11.8	13.8	9.7	11.4
	60대 이상	10.1	11.8	10.5	11.6	7.8	8.4	9.5	10.7
여	20대	7.7	7.2	7.0	6.6	6.8	6.0	7.7	6.4
	30대	10.9	10.8	10.4	9.5	10.3	9.9	10.9	10.7
	40대	13.0	13.1	11.9	12.1	14.0	14.2	12.2	13.1
	50대	9.0	10.7	9.7	11.0	10.2	10.8	8.7	9.8
	60대 이상	10.4	11.0	10.3	10.4	7.8	7.8	10.0	10.1
구 분		강원		충북		충남		전북	
		선관위	출구조사	선관위	출구조사	선관위	출구조사	선관위	출구조사
남	20대	6.5	4.4	7.2	4.9	6.6	3.6	7.0	4.8
	30대	7.9	7.4	8.4	7.5	7.3	7.3	6.8	6.4
	40대	12.1	11.2	11.8	9.9	11.1	11.3	10.2	10.8
	50대	11.0	13.4	10.3	12.8	10.9	12.3	10.1	11.9
	60대 이상	12.7	14.4	12.6	14.4	15.0	16.8	15.3	16.0
여	20대	4.9	4.4	5.8	5.2	4.9	3.9	5.2	5.2
	30대	8.6	8.4	9.0	8.0	7.8	7.7	7.4	7.4
	40대	11.8	10.5	11.1	10.6	10.6	10.7	10.0	10.4
	50대	10.3	11.7	9.7	11.8	10.1	10.9	9.8	10.7
	60대 이상	14.1	14.1	14.3	15.0	15.7	15.6	18.1	16.5
구 분		전남		경북		경남		제주	
		선관위	출구조사	선관위	출구조사	선관위	출구조사	선관위	출구조사
남	20대	6.1	3.2	6.3	3.7	7.4	5.1	7.5	4.8
	30대	5.9	7.1	6.3	6.1	7.9	6.8	8.5	8.2
	40대	9.8	11.7	10.3	10.2	11.5	10.5	11.3	11.9
	50대	10.1	11.8	10.9	13.3	10.6	12.6	9.9	11.9
	60대 이상	16.9	15.1	14.9	15.8	11.5	14.1	12.2	13.2
여	20대	4.6	3.6	4.7	3.6	5.8	5.5	6.0	6.1
	30대	5.7	8.7	6.9	7.6	8.6	7.4	8.4	8.3
	40대	9.2	11.0	10.2	10.9	11.8	10.8	10.6	11.5
	50대	9.9	10.5	10.9	12.1	10.2	11.7	10.0	10.1
	60대 이상	21.8	17.4	18.6	16.9	14.7	15.5	15.7	14.0

2010년 출구조사의 경우 시·도지사과 교육감에 대한 조사가 동시에 이루어졌으며, 본 연구에서는 지역에 따른 정당별 편향 분석에 초점을 맞추기 위해 시·도지사 출구조사 자료만을 분석대상으로 하였다. 실제 출구조사 예측은 최종 투표마감 시간인 6시 기준 자료가 아니라 5시까지 현장에서 조사된 자료를 기준으로 한다는 점을 고려해, 여기서는 방송사 예측에 실제로 사용되는 5시 기준 출구조사 자료를 이용했다는 점을 참고하기 바란다.

2. 표본 응답자 현황

출구조사를 통해 조사된 응답자의 시·도별, 성별·연령대 분포를 정리하면 <표 2>와 같다. 출구조사 응답자의 성별·연령대 분포를 전체 투표자 모집단의 성별·연령대와 비교해 보면 출구조사 표본의 대표성을 어느 정도 가늠해 볼 수 있을 것이다. 하지만 투표자 전체에 대한 성별·연령대 분포는 공식적으로 집계되지 않기 때문에 여기서는 중앙선거관리위원회(이하 선관위)의 투표율 분석 자료와 비교해 보기로 한다. 선관위 자료는 전체 투표소의 약 10%에 해당하는 표본 자료이기 때문에 전체 투표자 자료와 일부 차이가 있을 수 있다는 점을 유의하기 바란다(중앙선거관리위원회 2010).

출구조사와 선관위 자료의 성별·연령대 구성비를 보면, 다른 범주에서는 큰 차이를 보이지 않지만 특히 20대 남자의 경우 상대적으로 차이가 크다는 것을 볼 수 있다. 이는 출구조사의 경우 군인들에 대한 부재자 투표가 출구조사 대상에서 누락되어 있기 때문에 나타나는 현상이다.

3. 무응답 현황

출구조사를 통한 선거예측에 있어서도 무응답, 즉 조사거절자에 의한 편향은 심각할 수 있다. 그 동안 우리나라 출구조사에서 조사거절에 따른 무응답률은 개략적으로 15~20% 수준인 것으로 알려져 있다. 실제 출구조사에서 무응답률 및 무응답 처리방법에 따라 후보자별 득표율 예측값이 상당히 달라질 수 있으며, 이는 분명 편향요인이 될 수 있다. 이런 측면에서 김영원·곽은선(2010)은 2008년 총선에서 각 선거구별 무응답률과 표본추출오차와 비표본추출오차를 모두 포함하는 개념인 총조사오차와의 관계를 분석하여 출구조사에서 무응답률이 커짐에 따라 총조사오차가 증가한다는 사실을 보였다. 이런 관점에서 2010 지방선거 출구조사의 성별·연령대별 무응답률 현황을 정리하면 <표 3>과 같다.

〈표 3〉 출구조사 성별·연령대별 무응답률

(단위:%)

구 분		서울	부산	대구	인천	광주	대전	울산	경기
남	20대	3.3	3.0	1.9	2.2	1.9	2.5	2.7	2.9
	30대	6.3	5.1	4.5	4.4	5.3	5.3	8.2	6.8
	40대	7.5	7.7	10.1	7.0	8.2	7.8	10.1	10.1
	50대	11.7	12.5	12.1	13.2	12.8	13.0	15.2	10.9
	60대 이상	15.3	15.9	16.0	15.8	15.0	17.4	11.0	12.9
여	20대	5.4	4.8	3.6	3.1	3.7	4.2	3.1	3.9
	30대	6.9	7.8	7.0	5.5	7.8	7.5	8.2	9.7
	40대	9.5	7.8	11.6	10.1	11.9	10.2	12.8	11.0
	50대	13.5	11.8	13.0	13.9	12.2	11.9	10.6	11.6
	60대 이상	20.7	23.6	20.3	24.9	21.3	20.2	18.2	20.3
구 분		강원	충북	충남	전북	전남	경북	경남	제주
남	20대	2.1	2.1	1.4	1.6	1.2	1.2	2.2	2.8
	30대	4.2	4.6	3.7	2.3	3.4	4.4	3.4	5.3
	40대	8.4	5.8	7.0	6.1	7.1	6.0	6.7	9.2
	50대	12.6	11.6	11.3	9.7	11.1	12.5	10.1	10.6
	60대 이상	17.1	15.8	21.5	19.5	18.5	17.0	16.8	14.9
여	20대	1.7	2.7	1.4	3.4	1.1	1.4	2.5	3.5
	30대	5.3	6.4	5.5	4.5	4.3	5.8	4.3	6.4
	40대	7.7	8.9	8.7	7.5	6.2	8.4	8.9	9.4
	50대	12.3	12.7	10.2	10.9	10.3	11.8	13.1	10.2
	60대 이상	28.7	29.3	29.4	34.5	36.7	31.5	32.1	27.7

〈표 3〉의 성별과 연령대에 따른 무응답률을 비교해 보면, 남성보다 여성의 무응답률이 높으며, 연령대가 높아질수록 무응답률이 높아지는 것을 알 수 있다. 20대 남성의 무응답률이 가장 낮고 60대 이상 여성의 무응답률이 가장 높은 것으로 나타났으며, 이준웅(2004), 김영원·김지현(2006) 및 김영원·곽은선(2010)에 의하면 2006년 지방선거 출구조사나 2004년 및 2008년 총선 출구조사에서도 이와 유사한 무응답 패턴이 나타났다.

무응답에 따른 편향과 관련해 이준웅(2004)가 지적한 것처럼, 우리나라 선거예측에서 연령이 많을수록 보수 성향이 강해진다는 것은 일종의 상식이다. 하지만 문제는 동일한 성별·연령대 그룹 내에서 조사거절자와 응답자의 투표성향이 유사하다고 볼 수 있는지에 대해서는 확실하게 밝혀진 것이 없다는 점에 유의해야 할 것이다.

〈표 4〉 시·도별 득표율 추정치 및 표준오차

(단위: %, 명)

지 역		추정치(\hat{p})	방송사 예측결과	개표결과	표준오차 SE.(\hat{p})	표본 투표소수	표본 응답자 수	
							전 체	무응답 제외
서울	1위	48.7	47.4	47.4	0.97	50	20,396	15,615
	2위	46.1	47.2	46.8	0.92			
부산	1위	58.7	57.0	55.4	1.14	35	12,038	9,578
	2위	41.3	43.0	44.6	1.14			
대구	1위	77.4	76.4	72.9	0.95	25	7,858	5,740
	2위	14.6	15.3	16.9	0.85			
인천	1위	54.1	52.1	52.7	0.97	40	14,481	11,521
	2위	43.8	45.5	44.4	1.02			
광주	1위	58.8	58.8	56.7	1.56	25	7,268	5,659
	2위	13.8	14.5	14.5	1.12			
대전	1위	48.5	48.4	46.7	0.87	40	13,982	10,962
	2위	28.6	28.2	28.5	0.83			
울산	1위	64.6	63.0	61.3	1.22	25	8,211	6,346
	2위	26.8	27.9	29.3	1.11			
경기	1위	52.6	52.1	52.2	1.20	55	18,406	14,398
	2위	45.8	47.9	47.8	1.21			
강원	1위	52.2	53.4	54.4	1.55	41	11,484	8,957
	2위	46.9	46.6	45.6	1.57			
충북	1위	49.1	49.6	51.2	1.33	40	12,067	9,215
	2위	49.0	48.5	45.9	1.31			
충남	1위	43.6	41.4	42.3	1.72	53	14,628	11,007
	2위	37.9	38.8	39.9	1.45			
전북	1위	75.1	72.9	68.7	1.04	25	7,397	5,985
	2위	16.0	16.4	18.2	0.81			
전남	1위	73.5	72.2	68.3	1.15	35	10,067	7,627
	2위	12.0	11.6	13.4	0.85			
경북	1위	82.5	78.0	75.4	1.22	37	9,840	7,572
	2위	9.2	13.0	11.8	0.76			
경남	1위	51.7	51.5	53.5	1.51	40	13,358	9,846
	2위	48.3	48.5	46.5	1.51			
제주	1위	40.2	42.0	41.4	0.48	41	10,519	7,940
	2위	41.0	40.8	40.6	0.65			

4. 후보자별 득표율 추정 결과

2010 지방선거의 시·도지사 1, 2위 후보자에 대한 지역별 득표율 추정과 관련된 다양한 통계는 <표 4>에 정리되어 있다. 여기에는 본 연구를 위해 산출된 지역별 1위와 2위 후보자의 득표율 추정치(\hat{p}), 방송사 예측결과, 실제 개표결과 득표율(p), 득표율 추정치의 표준오차, 각 지역별 투표소 수 및 출구조사 전체 응답자 수와 무응답자를 제외한 응답자 수가 정리되어 있다.

<표 4>에 제시된 방송사 예측결과는 5시까지의 출구조사 자료를 사용한 것이지만, 최종 분석단계에서 각 지역별로 무응답 및 부재자 투표에 대한 조정과 출구조사에서 누락되는 5시 이후 투표자의 성향을 고려한 복잡한 보정 과정을 거치기 때문에 본 연구에서 계산한 후보자 득표율 추정치(\hat{p})와 차이가 있다는 점에 유의하기 바란다. 제시된 득표율 추정치(\hat{p})와 표준오차($SE(\hat{p})$)는 SAS의 Surveymeans 프로시저를 이용해 무응답을 제외한 5시까지의 출구조사 자료를 바탕으로 하여, 출구조사 표본설계에 적용된 층화 및 집락 효과를 고려하는 동시에 층별 추출확률에 따른 가중치를 반영해 구한 것이다.

III. 출구조사 정당별 편향 분석

1. 선거조사 예측 정확성과 편향 평가 척도

선거여론조사의 예측 정확성은 오래전부터 관심의 대상이 되어 왔다. 특히 여론조사의 정확성에 관심을 갖게 된 계기는 선거예측을 위한 여론조사의 대표적인 실패 사례로 꼽히는 1948년 미국 대통령선거 당선자 예측 사건이었다고 할 수 있다.

Mitofsky(1998)는 여론조사의 정확성 관련 척도들에 대한 기존 연구결과들을 다음과 같이 잘 정리하고 있다. 제시된 정확성 측정방법은 Mosteller 등이 1949년도에 SSRC(Social Science Research Council)를 설립한 후 고안한 것들로, 선거예측의 정확성을 판정하는 기준으로 8가지 예측오차 평가 척도를 제시하고 있다. 이 중 주요한 정확성 평가 기준들을 수식으로 표현하면 다음과 같다. 여기서 $\hat{p}_1(\hat{p}_2)$ 는 1위(2위) 후보자의 예측 득표율, $p_1(p_2)$ 는 1위(2위) 후보자의 실제 득표율, k 는 전체 입후보자의 수를 나타낸다.

$$\begin{aligned}
 \textcircled{1} \quad & \frac{\hat{p}_1}{\sum_{i=1}^k \hat{p}_i} - \frac{p_1}{\sum_{i=1}^k p_i} \\
 \textcircled{2} \quad & \frac{\hat{p}_1}{\hat{p}_1 + \hat{p}_2} - \frac{p_1}{p_1 + p_2} \\
 \textcircled{3} \quad & \frac{\sum_{i=1}^k |\hat{p}_i - p_i|}{k} \\
 \textcircled{4} \quad & \frac{\sum_{i=1}^k |(\hat{p}_i/p_i) - 1|}{k} = \frac{\sum_{i=1}^k |(\hat{p}_i - p_i)/p_i|}{k} \\
 \textcircled{5} \quad & (\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)
 \end{aligned}$$

하지만 Mitofsky(1998)가 지적한 대로 이런 척도들은 ‘부동층’(undecided vote)의 처리 방법에 따라 정확성 측정 결과가 달라질 수 있다는 문제가 있다. 부동층 배분방법이 정확성 척도에 미치는 영향에 대해서는 Crespi(1988)가 수행한 사례연구를 참고하기 바란다.

Mitofsky(1998) 또는 Crespi(1988) 등 대부분의 여론조사 정확성 평가를 목적으로 하는 연구에서는 결론적으로 ③과 ⑤를 선호하고 있으며, 우리나라의 경우 조성겸·김지연(2004)과 김영원·엄윤희(2005) 등 대부분의 선거조사 예측오차 관련 연구에서 ⑤에 절대값을 취한, $|(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)|$ 을 예측오차 평가 기준으로 사용하고 있다. 하지만 이런 %p 개념의 차이를 사용하는 척도들은 부동층 배분방법에 따라 평가 결과가 달라질 수 있고, 척도를 통해 얻은 수치의 의미를 해석하기 어려울 뿐만 아니라 오차의 방향성, 즉 정당별 편향의 발생 여부를 판정하기 어렵다는 문제도 있다.

이와 같이 기존 척도들이 갖고 있는 한계를 해결하기 위해 Martin et al.(2005)은 선거 여론조사의 예측 정확성 측정을 위해 두 개 정당(R과 D)²⁾ 소속 후보자의 득표율에 대한

2) Martin et al.(2005)의 연구에서는 미국의 공화당(Republican)과 민주당(Democratic) 양당 체제에서 각 정당의 후보자를 나타내기 위해 R과 D라는 기호를 사용하고 있으며, 본 연구에서는 편의상 한나라당을 포함한 보수지향 후보자와 민주당을 포함한 진보지향 후보자로 구분해 편향을 분석하기로 함.

오즈비(odds ratio)를 기초로 한 예측 정확성을 설명하기 위해 식 (1)과 같은 통계량 A 를 제시하였으며, 최근 여론조사 정확성 평가를 목적으로 하는 연구에서는 통계량 A 가 폭넓게 사용되고 있다(Mokrzycki et al. 2009; Traugott & Christopher 2009; Traugott 2005).

$$A = \log\left(\frac{r/d}{R/D}\right) \quad (1)$$

여기서 r 과 d 는 여론조사를 통해 얻은 득표율 예측값, R 과 D 는 개표결과 두 후보자들의 득표수를 나타낸다. r 과 d 는 여론조사에서 나타난 후보자별 단순 지지율로 보아도 되고, 전체 응답자(n) 중 무응답이나 기타 정당 지지자를 제외한 두 후보자 지지자만(n')을 기준으로 $r + d = 1$ 이 되도록 재계산된 후보자별 득표율 예측값으로 보아도 된다. 왜냐하면 두 경우 모두 통계량 A 는 같은 값을 갖기 때문이다. 본 연구에서는 지금부터 조사결과로부터 $r + d = 1$ 이 되도록 재계산된 두 후보자별 예측 득표율을 사용하기로 한다.

통계량 A 는 오즈비에 로그(log)를 취한 것으로 몇 가지 장점을 갖고 있다³⁾. 우선 오즈(r/d)를 사용하기 때문에 앞에서 언급한 대로 여론조사에서 부동층이나 기타 정당(1, 2위 후보자 정당을 제외한 나머지 정당) 후보자의 득표율을 무시하거나 응답자 지지율에 따라 1, 2위 후보자에게 비례배분하는 경우 모두 같은 값을 갖게 된다. 이는 다른 척도들이 부동층 처리방법 등에 따라 정확성이 달리 계산되는 것과 비교하면 통계량 A 가 갖는 큰 장점이 될 수 있다. 또한 해석이 쉽다는 장점도 있는데, 오즈비가 1이면 여론조사결과가 개표결과를 정확히 예측한 것을 나타내며, 이 값이 1에서 멀어질수록 정확성이 떨어지는 것으로 보면 된다. 아울러 오즈비가 1보다 커지면(작아지면) 여론조사결과가 R정당(D정당) 후보를 과대 추정한 것을 의미한다. 만약 이 값이 표본추출오차로 설명될 수 있는 수준을 벗어나는 경우 해당 선거에서 특정 정당에 대해 편향이 발생한 것으로 해석할 수 있다. 여기서 오즈비를 로그(log) 변환한 이유는 통계량 A 가 대칭성을 갖는 동시에 분산추정이 용이하다는 측면을 고려한 것이다. 단순확률추출을 가정하는 경우 통계량 A 의 분산을 유도해 보면 근사적으로 다음과 같다(Martin et al.(2005) 참조).

3) 통계량 A 와 같은 형식의 통계량은 범주형 자료분석에서 자주 사용되는 것으로 '로그오즈비'라고 부르며, 이와 관련된 내용은 정광모·최용석(2002) 등을 참고하기 바람.

$$\text{Var}(A) = \text{Var}\left(\log\left(\frac{r/d}{R/D}\right)\right) \cong \text{Var}\left(\frac{r/d}{R/D}\right) = \frac{1}{n'pq} \quad (2)$$

여기서 p 와 q 는 정당 R과 D 후보자들에 대한 실제 득표율을 나타낸다($p + q = 1$).

그러면 본 연구에서 관심을 갖는 특정 정당에 대한 편향 발생 여부는 어떻게 판정할 수 있는가? 편향 발생 여부는 산출된 통계량 A 가 식 (2)의 근사적인 분산을 이용해 구한 '0'을 기준으로 한 신뢰구간을 벗어나는지 여부를 기준으로 판정이 가능할 것이다. 이런 관점에서 본 연구에서는 특히 편향 여부를 판정하기 위한 목적으로 통계량 A 를 활용하고자 한다.

2. 선거별 정당 편향에 대한 실증분석

1) 지역별 정당 편향 분석 방법

우선 2010 지방선거 출구조사 예측결과와 실제 개표결과를 비교하기 위해 통계량 A 를 이용한 95% 신뢰구간을 설정해 16개 시·도별로 정당 편향 발생 여부를 살펴보았다. 또한 우리나라의 특수한 정치적 지역구도에 따라 지역별 정당 편향이 그 동안의 출구조사에서도 지속적으로 발생해 왔는지 확인해 보기 위해 2006년 제4회 지방선거 출구조사 자료와 2007년 제17대 대선 출구조사 자료를 기초로 한 지역별 정당 편향을 분석하였다. 참고로 2007년 대선 출구조사의 경우 지역별 분석을 위해 전국 대상 출구조사 자료를 16개 시·도로 구분하여 분석하였다.

본 연구에서는 2010년 지방선거의 경우 방송 3사 공동예측조사위원회(KEP)에서 실시한 16개 시·도지사 선거에 대한 출구조사 자료를 사용하였으며, 2006년 제4회 지방선거의 경우 KBS/SBS가 공동으로 16개 시·도 중 경합지역으로 분류된 6개 시·도를 대상으로 수행한 출구조사 자료를 사용하였다. 아울러 2007년 17대 대통령 선거는 KBS/MBC가 공동으로 전국 단위로 수행한 한나라당 이명박 후보와 열린우리당 정동영 후보를 대상으로 한 대통령선거 출구조사 자료를 16개 시·도로 구분하여 각 시·도별로 두 후보자에 대한 예측 지지율을 산출해 출구조사의 정당 편향 발생 여부를 분석한 것이다.⁴⁾

4) 2010년 지방선거의 경우 밀워드브라운미디어리서치(MBMR), 코리아리서치(KRC), TNS-RI가 수행한 출구조사자료를 사용했으며, 2006년 지방선거의 경우 밀워드브라운미디어리서치(MBMR)와 TNS-RI에서 수행한 출구조사 자료를 사용하였고, 2007년 대통령선거의 경우 밀워드브라운미디어리서치(MBMR)와 코리아리서치(KRC)가 수행한 출구조사 자료를 사용하였음.

한편 식 (2)의 Martin et al.(2005)이 정의했던 통계량 A 에 대한 분산은 여론조사가 단순확률추출로 이루어졌다는 것을 전제로 한 것이기 때문에 우리나라 출구조사 편향 분석에 그대로 이용할 수는 없다. 따라서 여기서는 출구조사에서 지역별로 집락 및 층화 추출에 따라 설계효과(design effect: D_{eff})를 계산하고 이를 근거로 유효표본크기(effective sample size; n_{eff})를 구해 다음과 같은 방식으로 출구조사에서 산출된 통계량 A 에 대한 표준오차를 추정하였다.

$$A = \log\left(\frac{\hat{p}_a/\hat{p}_b}{p_a/p_b}\right), \quad S.E.(A) = \sqrt{\frac{1}{n_{eff} p_a p_b}}, \quad n_{eff} = \frac{n'}{D_{eff}} \quad (3)$$

여기서 \hat{p}_a 는 한나라당 후보를 포함한 보수지향 후보자 예측 득표율, \hat{p}_b 는 민주당 후보를 포함한 진보지향 후보자 예측 득표율($\hat{p}_a + \hat{p}_b = 1$), n' 는 출구조사에서의 1, 2위 후보자 총득표 수를 나타낸다. 여기서 설계효과(D_{eff})는 출구조사에서 추정량의 분산과 동일한 표본크기를 갖는 단순확률추출에서 분산의 비(ratio)를 나타낸다.

2) 2010 지방선거 지역별 정당 편향 분석

2010년 지방선거에는 지역별로 다양한 정당 후보자들이 입후보하였지만 최종적으로 1, 2위 후보자가 정해졌다. 통계량 A 를 적용한 지역별 분석을 위해 여기서는 16개 시·도의 1, 2위 후보자를 보수지향 후보자와 진보지향 후보자로 구분하여 분석하고자 한다. 우리나라 정당을 보수지향 정당과 진보지향 정당으로 구분하는 것이 쉽지는 않지만 여기서는 한나라당과 자유선진당 후보는 보수지향, 민주당과 민주노동당 및 국민참여당 후보는 진보지향으로 분류하고, 무소속 중 경남의 김두관 후보는 진보지향, 제주의 경우 가장 최근 소속 정당에 따라 무소속 후보 중 우근민 후보는 진보지향, 현명관 후보는 보수지향으로 구분하기로 한다. 대전의 경우 한나라당과 자유선진당 후보가 1, 2위를 차지해, 보수와 진보 지향 구분이 어렵지만 여당인 한나라당 후보를 보수지향, 다른 후보를 진보지향으로 구분해 분석했으며, 광주의 경우 민주당과 국민참여당 후보가 1, 2위를 한 상황이라 보수와 진보 지향 구분이 불가능해 분석대상에서 제외하였다.

2010 지방선거에 대한 1, 2위 후보자의 출구조사 예측 득표율과 최종 개표결과 득표율, 소속정당 및 지역별 설계효과는 <표 5>와 같다. <표 5>의 예측결과와 실제 결과는 시·도별 1위와 2위 후보자의 출구조사 예측 득표율과 개표결과 득표율을 1, 2위 후보자 득표율의 합이 100%가 되도록 재계산한 것이다. 음영으로 진하게 표시된 후보는 소속정당 등을 고려할 때 보수지향 후보로 구분되는 후보를 나타내며, 통계량 A 산출을 위한 기준이 되는 후보이다. 2010 지방선거 출구조사에서 1위로 예측된 후보자가 실제 개표결과 모두 당선되었음을 확인할 수 있다.

<표 5> 2010년 지방선거 득표율 및 설계효과

(단위: %)

지 역	예측결과		실제결과				D_{eff}	
	1위	2위	1위	2위	1위 후보 성향(정당)	2위 후보 성향(정당)	1위	2위
서울	51.4	48.7	50.3	49.7	보수(한)	진보(민)	5.9	5.3
부산	58.7	41.3	55.4	44.6	보수(한)	진보(민)	5.1	5.1
대구	84.1	15.9	81.2	18.8	보수(한)	진보(민)	2.9	3.3
인천	55.3	44.7	54.3	45.7	진보(민)	보수(한)	4.4	4.8
대전	62.9	37.1	62.1	37.9	진보(자)	보수(한)	3.3	3.7
울산	70.7	29.3	67.7	32.3	보수(한)	진보(민노)	4.1	4.0
경기	53.4	46.6	52.2	47.8	보수(한)	진보(국)	8.3	8.5
강원	52.7	47.3	54.4	45.6	진보(민)	보수(한)	8.6	8.9
충북	50.1	49.9	52.7	47.3	진보(민)	보수(한)	6.5	6.4
충남	53.5	46.5	51.4	48.6	진보(민)	보수(자)	13.3	9.8
전북	82.4	17.6	79.1	20.9	진보(민)	보수(한)	3.4	2.9
전남	86.0	14.0	83.6	16.4	진보(민)	보수(한)	5.2	5.2
경북	90.0	10.0	86.4	13.6	보수(한)	진보(민)	7.8	5.3
경남	51.7	48.4	53.5	46.5	진보(무)	보수(한)	9.0	9.0
제주	49.5	50.5	50.5	49.5	진보(무)	보수(무)	0.7	1.4

〈표 6〉 2010년 지방선거 예측 정확성(A) 및 편향 분석

(단위: %)

지역	정확성			신뢰구간	
	A	n_{eff}	$S.E(A)$	상한	하한
서울	0.0412	2,773	0.0380	0.0745	-0.0745
부산*	0.1354	1,870	0.0470	0.0921	-0.0921
대구*	0.2032	1,840	0.0638	0.1250	-0.1250
인천	-0.0398	2,498	0.0402	0.0789	-0.0789
대전	-0.0341	3,134	0.0370	0.0725	-0.0725
울산*	0.1414	1,564	0.0555	0.1089	-0.1089
경기	0.0489	1,717	0.0484	0.0948	-0.0948
강원	0.0671	1,028	0.0625	0.1225	-0.1225
충북*	0.1075	1,428	0.0529	0.1037	-0.1037
충남	-0.0834	956	0.0648	0.1271	-0.1271
전북*	-0.2187	1,894	0.0604	0.1184	-0.1184
전남*	-0.1847	1,466	0.0752	0.1474	-0.1474
경북*	0.3406	1,161	0.0977	0.1914	-0.1914
경남	0.0741	1,096	0.0605	0.1185	-0.1185
제주	0.0397	7,395	0.0233	0.0456	-0.0456

* : 출구조사 편향 지역

〈표 6〉은 2010년 지방선거 출구조사에 대한 통계량 A와 A의 표준오차, 유효표본크기(n_{eff}) 및 편향 판정을 위한 신뢰구간을 정리한 것이다. 15개 지역 중 A값이 (+)인 지역은 10개 지역이고, (-)인 지역은 5개 지역이다. 이 중 5% 유의수준에서 통계적으로 유의한, 다시 말해 출구조사에서 편향이 발생한 것으로 볼 수 있는 지역은 A가 (+)인 10개 중 5개 지역, A가 (-)인 5개 중 2개 지역이다.

〈표 6〉에서 출구조사 편향이 발생한 지역 중 부산, 대구, 울산, 충북, 경북 지역은 A 값이 유의적으로 (+)이므로, 보수지향 후보자 편향이 발생한 지역으로 볼 수 있다. 즉, 출구조사가 실제결과보다 보수지향 후보자를 과대 추정했다고 볼 수 있으며, 이들 지역은 일반적으로 보수성향이 강하다고 알려진 지역이다. 한편 전북, 전남 지역은 A

값이 유의적으로 (-)이므로, 진보지향 후보자 편향이 발생한 지역으로 분류될 수 있다. 즉, 출구조사가 실제결과보다 보수지향 후보자를 과소 추정했다고 볼 수 있다. 충북 지역을 제외한 4개의 보수지향 후보자 편향 지역은 모두 보수지향 후보자가 당선되었으며 전북, 전남 지역은 편향의 방향대로 진보지향 후보자가 당선되었다. 하지만 충북의 경우 다른 지역과 달리 보수지향 후보(한나라당 후보)가 당선자가 아님에도 불구하고 보수지향 편향이 발생했다는 점에 특별히 유의할 필요가 있다.

3) 2006 지방선거 지역별 정당 편향 분석

2006 지방선거의 경우 KBS/SBS가 공동으로 수행한 6개 경합지역을 대상으로 한 출구조사 자료를 기준으로 지역별 편향을 분석하였다. 통계량 A 를 토대로 한 지역별 편향 분석을 위해 한나라당과 국민중심당 소속 후보는 보수지향 후보로 분류하고 이를 기준으로 통계량 A 를 산출하였다. 민주당과 열린우리당이 경합한 광주를 분석대상에서 제외했고, 무소속 후보는 가장 최근 소속 정당을 기준으로, 또한 1, 2위 후보자가 모두 보수지향인 충남과 제주를 한나라당(현 여당) 후보자를 기준으로 분석했다.

2006 지방선거에 대한 1, 2위 후보자의 출구조사 예측 득표율과 실제 개표결과 및 지역별 설계효과 등은 <표 7>과 같다. 여기서도 예측 득표율은 1, 2위 후보자 득표율의 합이 100%가 되도록 재계산한 것이다.

<표 7> 2006 지방선거 득표율 및 설계효과

(단위: %)

지역	예측결과		실제결과				D_{eff}	
	1위	2위	1위	2위	1위 후보 성향(정당)	2위 후보 성향(정당)	1위	2위
서울	71.8	28.2	69.1	30.9	보수(한)	진보(열)	4.4	3.2
대전	51.8	48.3	51.6	48.4	보수(한)	진보(열)	3.7	2.5
경기	70.5	29.5	66.0	34.0	보수(한)	진보(열)	7.8	5.4
충남	67.0	33.0	64.5	35.5	보수(한)	보수(국)	4.5	7.1
제주	49.9	50.1	50.9	49.1	보수(무)	보수(한)	3.1	2.9

〈표 8〉 2006 지방선거 예측 정확성(A) 및 편향 분석

(단위: %)

지역	정확성			신뢰구간	
	A	n_{eff}	$S.E(A)$	상한	하한
서울*	0.1281	4,771	0.0322	0.0631	-0.0631
대전	0.0062	3,405	0.0343	0.0672	-0.0672
경기*	0.2087	2,734	0.0419	0.0822	-0.0822
충남*	0.1139	1,648	0.0524	0.1027	-0.1027
제주	0.0438	2,954	0.0368	0.0721	-0.0721

* : 출구조사 편향 지역

〈표 8〉은 2006 지방선거 출구조사에 대한 통계량 A와 A의 표준오차, 유효표본 크기 (n_{eff}) 및 편향 판정을 위한 신뢰구간을 정리한 것이다. 이 경우 2010 지방선거와 달리 5개 지역에 대한 분석만 이루어졌기 때문에 사례가 많지 않아 한계가 있지만, 5개 지역 모두 A 값이 (+)이고, 5% 유의수준에서 통계적으로 편향이 유의한, 다시 말해 출구조사에서 편향이 발생한 것으로 볼 수 있는 지역은 서울, 경기, 충남 3개 지역이다.

서울, 경기, 충남은 A 값이 유의적으로 (+)이므로, 보수지향 후보자 편향이 발생한 것으로 볼 수 있다. 즉, 출구조사가 실제 결과보다 보수지향 후보자를 과대 추정한 것이다. 결과적으로 2006 지방선거의 경우 경합지역을 중심으로 실시된 출구조사에서는 당시 여당이었던 열린우리당보다 한나라당 후보자에 대한 득표율을 과대 추정하는 경향이 뚜렷했음을 알 수 있다.

4) 2007 대선 지역별 정당 편향 분석

2007년 대통령 선거 출구조사는 KBS/MBC가 공동으로 전국 단위로 수행하였다. 여기서는 한나라당 이명박 후보와 열린우리당 정동영 후보에 대한 출구조사가 16개 시·도에서 별개로 이루어진 것으로 간주하고, 시·도별로 출구조사 자료를 이용해 두 후보자별 득표율을 추정하고 이를 지역별 개표결과와 비교해 출구조사 과정에서 편향이 발생하였는지 분석하였다.

〈표 9〉는 2007 대선 결과를 시·도별로 구분하여 한나라당 이명박 후보와 열린우리당 정동영 후보의 득표율을 재계산한 결과 및 실제 개표결과와 함께 지역별 설계효과를 정리한 것이다. 지역별로 보수지향 후보인 한나라당 이명박 후보 득표율은 음영으로 진하게 표시하였다.

〈표 9〉 2007년 대선 시·도별 득표율 및 설계효과

(단위: %)

지역	예측결과		실제결과				D_{eff}	
	1위	2위	1위	2위	1위 후보 성향(정당)	2위 후보 성향(정당)	1위	2위
서울	70.3	29.7	68.5	31.5	보수(한)	진보(열)	3.0	2.7
부산	82.1	17.9	81.2	18.9	보수(한)	진보(열)	1.2	2.9
대구	93.1	6.9	92.0	8.0	보수(한)	진보(열)	1.6	0.9
인천	68.4	31.6	67.4	32.6	보수(한)	진보(열)	3.6	1.6
광주	92.0	8.0	90.3	9.7	진보(열)	보수(한)	6.6	5.8
대전	60.1	39.9	60.6	39.4	보수(한)	진보(열)	3.3	0.6
울산	80.6	19.4	79.8	20.2	보수(한)	진보(열)	6.6	1.8
경기	70.9	29.1	68.8	31.2	보수(한)	진보(열)	2.6	2.2
강원	75.9	24.1	73.4	26.7	보수(한)	진보(열)	1.7	2.4
충북	66.3	33.7	63.6	36.4	보수(한)	진보(열)	2.8	0.8
충남	61.0	39.0	61.9	38.1	보수(한)	진보(열)	4.1	2.3
전북	91.1	8.9	90.0	10.0	진보(열)	보수(한)	2.4	3.3
전남	90.0	10.0	89.5	10.5	진보(열)	보수(한)	3.7	2.6
경북	91.4	8.6	91.5	8.6	보수(한)	진보(열)	12.3	3.6
경남	84.1	15.9	81.7	18.3	보수(한)	진보(열)	6.1	3.1
제주	55.8	44.2	54.2	45.8	보수(한)	진보(열)	0.8	0.4

〈표 10〉은 2007 대선 출구조사를 시·도별로 구분해 구한 통계량 A 와 A 의 표준오차, 유효표본크기(n_{eff}) 및 편향 판정을 위한 신뢰구간을 정리한 것이다. 여기서 통계량 A 는 보수지향 후보자인 한나라당 이명박 후보를 기준으로 산출한 것이다. 15개 지역 중 A 값이 (+)인 지역은 10개 지역이고, (-)인 지역은 6개 지역이다. 〈표 6〉에 제시된 2010 지방선거 분석결과와 지역별로 통계량 A 의 부호만을 비교해 보면 인천과 경북을 제외하고(경북은 〈표 10〉에서 A 값이 거의 0인 것으로 볼 수 있음) 다른 지역에서는 일치한다는 것을 확인할 수 있다. 통계량 A 의 부호가 시·도별로 일치한다는 것은, 통계량 A 가 (+) 또는 (-) 방향으로 통계적으로 유의적인 편향을 입증해 주지는 못하더라도, 최소한 지역에 따라 출구조사가 일정한 방향으로 치우침이 발생할 소지가 높다는 것을 시사한다. 2007 대선 출구조사의 경우 통계량 A 를 기준으로 분석해 보면 5% 유의수준

에서 통계적으로 유의한, 다시 말해 출구조사에서 편향이 발생한 것으로 볼 수 있는 지역은 A 가 (+)인 10개 지역 중 4개, A 가 (-)인 6개 지역 중에는 없다.

<표 10>을 보면 서울, 경기, 강원, 충북은 A 값이 유의적으로 (+)이므로 보수지향 후보자 편향이 발생했다고 할 수 있다. 즉, 이들 지역에서는 출구조사에서 실제 결과보다 한나라당 이명박 후보를 과대 추정하는 편향이 발생했다는 것을 확인할 수 있다. 나머지 지역에서는 유의적이지 않은 것으로 나타나 정당 편향은 없었던 것으로 볼 수 있다.

<표 10> 2007년 대선 시·도별 예측 정확성(A) 및 편향 분석

(단위: %)

지 역	정확성			신뢰구간	
	A	n_{eff}	$S.E(A)$	상한	하한
서울*	0.0838	4,739	0.0318	0.0623	-0.0623
부산	0.0614	1,534	0.0666	0.1305	-0.1305
대구	0.1487	1,856	0.0914	0.1791	-0.1791
인천	0.0461	1,116	0.0644	0.1262	-0.1262
광주	-0.2090	426	0.1782	0.3492	-0.3492
대전	-0.0217	1,232	0.0582	0.1140	-0.1140
울산	0.0506	713	0.0948	0.1858	-0.1858
경기*	0.0995	5,086	0.0309	0.0605	-0.0605
강원*	0.1377	1,156	0.0688	0.1349	-0.1349
충북*	0.1187	1,514	0.0544	0.1066	-0.1066
충남	-0.0379	823	0.0715	0.1401	-0.1401
전북	-0.1252	846	0.1207	0.2366	-0.2366
전남	-0.0514	724	0.1237	0.2425	-0.2425
경북	-0.0057	368	0.1858	0.3642	-0.3642
경남	0.1694	761	0.0991	0.1942	-0.1942
제주	0.0662	2,061	0.0444	0.0869	-0.0869

* : 출구조사 편향 지역

〈표 11〉 세 선거 출구조사에서 편향 발생 현황

2010년 지방선거 (15개)	2007년 대선 (16개)	2006년 지방선거 (5개)
7(47%)	4(25%)	3(60%)

3. 지역별 정당 편향 발생 경향

제시된 세 가지 선거의 출구조사에 대한 편향 분석결과를 비교해 보면, 전반적으로 통계량 A 값의 부호가 지역별로 비슷한 경향을 보이고 있는 것을 알 수 있다. 서울 및 수도권, 충청권, 영남권은 대체적으로 보수지향 후보자 편향 경향을 보이고 있고 호남권은 대체적으로 진보지향 후보자 편향 경향을 보이고 있다.

종합적으로 지역별 정당 편향 현상을 살펴보기 위해 앞에서 분석한 세 가지 선거의 36개 지역⁵⁾의 출구조사 자료를 종합적으로 분석해 보기로 한다. 우선 세 가지 선거에서 보수 또는 진보지향 여부와 관계없이 출구조사에서의 편향 발생 현황을 요약해 보면 〈표 11〉과 같다. 표를 보면 2010 지방선거 출구조사의 경우 예측결과가 매우 정확했다는 평을 받고 있지만 편향 문제에 있어서는 다른 선거에 비해 아주 좋은 성과를 보였다고 할 수는 없을 것 같다. 다시 말해 매우 정확한 예측결과를 얻은 2010 지방선거에서도 출구조사 응답과정에서 지역에 따른 편향이 발생했다고 볼 수 있다.

〈표 12〉는 보수지향 후보자 과대 편향 지역 분포를 보수지향 우세/열세 지역으로 나눠서 정리한 것이다. 보수지향 우세 지역은 당선자가 보수지향 후보자인 지역을 의미하고, 열세 지역은 당선자가 진보지향 후보자인 지역을 말한다. 보수지향 우세 지역에선 대부분이 보수지향 과대 편향을 보이며, 편향이 유의한 지역의 비율도 꽤 높다. 반대로 보수지향 열세 지역(진보지향 우세 지역)에선 보수지향 과소 편향이 다소 있지만 유의한 지역의 비율은 그다지 높지 않다. 따라서 우리나라 출구조사에서 보수지향 우세 지역의 보수지향 과대 편향 문제가 가장 심각한 현상이라고 할 수 있으며, 이는 대체적으로 현 집권당인 여당에 대한 과대 편향 문제와 연결해 해석해 볼 수 있다. 이런 현상은 김영원·곽은선(2010)에서도 지적한 것처럼 자신을 소수파라고 생각하는 유권자들이 자신이 주변의 다수 여론과 다르다고 생각하면 아예 조사 참여를 거절하거나 거짓 응답을 하는 경우가 생길 수 있다고 하는 흔히 말하는 ‘침묵의 나선(spiral of silence)’ 효과 때문에 발생하는 것으로 설명될 수 있을 것이다.

5) 2006년 지방선거의 5개 지역, 2007년 대선의 16개 지역, 2010년 지방선거의 15개 지역.

〈표 12〉 보수지향 후보자 과대·과소 편향 지역 분포

보수지향 우세 지역(23개)	보수지향 후보 (+) 추정 지역	20(87%)
	과대 편향 유의 지역	11(48%)
보수지향 열세 지역(13개)	보수지향 후보 (-) 추정 지역	8(62%)
	과소 편향 유의 지역	2(15%)

〈표 13〉 보수지향 후보 우세 지역 편향 발생 현황

1, 2위 득표율 격차	30% 초과(14)	보수지향 후보 (+) 추정 지역	13(93%)
		과대 편향 유의 지역	8(57%)
	10~30%(5)	보수지향 후보 (+) 추정 지역	3(60%)
		과대 편향 유의 지역	3(60%)
	10% 미만(4)	보수지향 후보 (+) 추정 지역	4(100%)
		과대 편향 유의 지역	0(0%)

〈표 14〉 보수지향 후보 열세 지역 편향 발생 현황

1, 2위 득표율 격차	30% 초과(5)	보수지향 과소 편향 경향	5(100%)
		보수지향 과소 편향	2(40%)
	10~30%(1)	보수지향 과소 편향 경향	1(100%)
		보수지향 과소 편향	0(0%)
	10% 미만(7)	보수지향 과소 편향 경향	2(29%)
		보수지향 과소 편향	0(0%)

〈표 13〉은 ‘침묵의 나선’ 효과를 보다 구체적으로 살펴보기 위해, 보수지향 후보자 우세 지역을 1, 2위 득표율 격차를 기준으로 3개 범주로 나누고, 각 범주에 해당하는 사례 중 보수지향 후보 편향이 발생한 비율을 비교해 본 것이다. 범주별로 편향이 발생한 건수(비율)를 비교해 보면, 1, 2위 격차가 클수록 보수지향 과대 편향의 경향이 뚜렷해진다 는 것을 알 수 있다. 즉 보수지향 우세 지역에서 보수지향 후보(1위 후보)의 득표율과 2위 후보 득표율의 차이가 클수록, 다시 말해 보수성향이 강할수록 출구조사에서 보수지향 후보에 대한 과대 편향이 발생할 가능성이 높다는 것을 알 수 있다.

한편 보수지향 열세 지역(즉, 진보지향 우세 지역)에 대한 분석결과는 〈표 14〉와 같

다. 여기서도 보수지향 후보자 열세 지역을 1, 2위 득표율 격차에 따라 3개 범주로 나누어 살펴보았다. 여기서도 격차의 크기가 클수록(즉, 진보성향이 강한 지역일수록) 보수지향 후보 과소 편향 경향이 뚜렷해진다는 것을 알 수 있다. 즉 보수지향 열세 지역에서 진보지향 후보(1위 후보)의 득표율과 2위 후보(보수지향 후보) 득표율의 차이가 클수록 출구조사에서 보수지향 후보 과소 편향(진보지향 후보 과대 편향)의 발생 가능성이 높다는 것을 알 수 있다.

IV. 맺음말

본 연구에서는 2010 지방선거 출구조사 특성을 간략히 정리하고, 2010 지방선거를 포함해 2006 지방선거와 2007 대선 출구조사 자료를 기초로 지역별 정치적 성향에 따라 특정 정파에 대한 편향이 출구조사에서 발생하는지 여부를 분석해 보았다. 이를 위해 Martin et al.(2005)이 제안한 여론조사 예측 정확성 측정을 위한 통계량 A 를 활용하였다.

선거 예측조사에서 오차관리를 위해서는 무엇보다 편향을 줄이는 것이 중요하다. 이를 위해서는 기존 조사에서의 편향 발생 여부를 분석해 보는 것이 필요하고, 아울러 어떤 유형의 편향이 어떤 원인에 의해 발생하는지를 파악하는 것이 필요하다. 본 연구에서는 향후 출구조사의 정확성 제고 차원에서 그 동안 우리나라에서 수행된 출구조사에서 지역별로 발생한 편향을 분석하고, 이런 편향과 지역별 정치적 성향과의 관계를 분석하였다. 학계나 조사업계의 전문가들은 지역별 정치 성향에 따라 출구조사에서 체계적인 편향이 발생한다는 주장을 많이 하였지만 이에 대한 이론적인 연구는 미흡한 실정이었다는 점을 고려하면 본 연구결과는 계량적인 분석결과를 제시하였다는 측면에서 의의를 찾을 수 있을 것이다.

본 연구에서는 추측은 되지만 명확하게 그 실체를 분석하는 데 한계가 있었던 지역에 따른 정당별 편향 문제를 실제 세 가지 선거의 출구조사 자료를 토대로 분석하였다. 분석결과 일반적으로 추측했던 것처럼 보수성향이 강한 지역에서는 대체적으로 보수지향 후보자의 득표율이 과대 추정되는 현상이 발생하고, 진보성향이 강한 지역에서는 반대로 진보지향 후보자의 득표율이 과대 추정된다는 사실을 통계적으로 확인할 수 있었다. 아울러 1, 2위 후보자의 득표율 격차가 커질수록 다시 말해 보수(진보) 성향이 강하면 강할

수록 출구조사에서 보수(진보) 후보자에 대한 (+) 편향이 심각하게 발생할 수 있다는 사실을 확인했다. 이런 현상은 흔히 말하는 ‘침묵의 나선효과’로 설명될 수 있을 것이다.

물론 본 연구에서 얻은 결론은 세 가지 선거의 출구조사 자료만을 가지고 내린 결론이기 때문에 해석상 한계가 있을 수 있다. 하지만 본 연구에서 얻은 지역별 편향 문제를 제대로 이해하고 발생 요인을 정확히 이해한다면 향후 출구조사의 정확성 제고에 큰 도움이 될 수 있을 것으로 생각한다.

본 연구에서는 출구조사로 연구대상을 제한하였지만, 여기서 다룬 지역별 정치성향에 따른 편향 문제는 전화조사에서도 거의 그대로 발생할 것으로 추측된다. 따라서 그동안 우리나라에서 수행되었던 선거 전화조사의 정확성을 비교·분석하고 향후 전화조사 개선 방안을 도출한다는 차원에서 본 연구와 유사한 연구가 선거여론 전화조사를 대상으로 수행되는 것도 큰 의미가 있을 것이다. 아울러 본 연구에서 사용한 통계량 A 는 척도의 속성상 설명변수로 선거여론조사의 정확성을 결정하는 것으로 판단되는 다양한 요인들(추출방법, 조사시점, 응답률, 재조사횟수 등)을 사용한 로그선형(log linear) 모형에 적용하기 적합하기 때문에 선거여론조사의 정확성을 결정해 주는 요인을 분석하기 위한 연구에서도 활용 가치가 높을 것으로 보인다.

참고문헌

- 김세용·허명희. 2009. “전화조사의 체계적 편향 -2007년 대통령선거 여론조사들에 대한 메타분석-.” 《응용통계연구》 22(2): 375-385.
- 김영원·곽은선. 2010. “2008 총선 출구조사의 총조사오차 분석.” 《조사연구》 11(3): 33-55.
- 김영원·김지현. 2007. “2006년 지방선거 출구조사 현황 및 예측오차.” 《조사연구》 8(1): 55-79.
- 김영원·엄윤희. 2005. “출구조사를 위한 투표소 확률추출 방법.” 《조사연구》 6(2): 1-32.
- 이준웅. 2004. “제17대 총선 예측조사의 문제: 조사 거절자와 응답불성실자 편향을 중심으로.” 《언론정보연구》 41(1): 110-136.
- 정광모·최용석. 2002. 《범주형 자료분석 개론》. 자유아카데미.
- 조성겸·김지연. 2004. “출구조사의 투표소 표집방안 비교.” 《조사연구》 5(2): 3-29.
- 중앙선거관리위원회. 2010. 제5회 전국동시지방선거 투표율 분석 자료. 보도자료. (<http://www.nec.go.kr>)

- 한국방송협회 KEP(Korea Election Pool). 2010. 《2010 지방선거 방송사 공동예측조사 평가보고서》.
- 현경보. 2005. “선거 예측조사의 정확성과 그 영향요인 연구: 17대 총선 예측조사 결과의 사후 분석.” 《언론과학연구》 5(1): 301-336.
- Crespi, Irving. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russel Sage.
- Martin E.A., M.W. Traugott, and C. Kennedy. 2005. “A Review and Proposal for a New Measure of Poll Accuracy.” *Public Opinion Quarterly* 69(3): 342-369.
- Mitofsky, Warren. 1998. “The Polls – Review: Was 1996 a Worse Year for Polls than 1948?” *Public Opinion Quarterly* 62: 230-249.
- Mokrzycki, M., S. Keeper, and C. Kennedy. 2009. “Cell-phone-only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias.” *Public Opinion Quarterly* 73(5): 845-865.
- Mosteller, Frederick, H. Hyman, P.J. McCarthy, E.S. Marks, and D.B. Truman. 1949. *The Pre-Election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. New York: Social Science Research Council.
- Traugott, M.W. 2005. “The Accuracy of the National Preelection Polls in the 2004 Presidential Election.” *Public Opinion Quarterly* 69(5): 642-654.
- Traugott, M.W. and W. Christopher. 2009. “The Dynamics of Poll Performance during the 2008 Presidential Nomination Contest.” *Public Opinion Quarterly* 73(5): 866-894.

<접수 2011/10/22, 수정 2011/11/16, 게재확정 2011/11/18>