

연구논문

다수 후보에 대한 선거예측의 정확성과 당선 가능성: 2017년 대통령선거의 경우*

The Accuracy of Election Forecasts and the Chance of Winning for
Multiple Candidates: In Case of the 2017 Presidential Election

류 제 복^{a)}
Jea-Bok Ryu

본 연구에서는 2017년도 대통령선거에 입후보한 다수 후보들에 대한 선거여론조사의 정확성을 분석하고 당선 가능성을 측정하였다. 다수 후보에 대한 선거예측의 정확성을 평가하기 위해서 오즈비를 이용한 Martin et al.(2005)의 척도를 일반화한 Arzheimer & Evans(2014)의 척도를 사용하였다. 중앙선거여론조사심의위원회에 등록된 58개 선거여론조사를 분석한 결과 보수 후보에 대해서는 과소 예측이 극심하였고, 진보나 중도 후보들은 과대 예측되었다. 그리고 선거여론조사결과를 바탕으로 오즈(odds)를 이용해서 후보들의 당선 가능성을 측정하였다. 향후 선거여론조사에서 후보들에 대한 지지율과 함께 당선 가능성을 제시해주는 것도 바람직하다.

주제어: 다수 후보, 예측오차, 당선 가능성, 오즈

In this study, we analyzed the accuracy of the election polls and measured the chance of winning for the multiple candidates in the 2017 presidential election. In order to evaluate the accuracy of the election forecast for multiple candidates, we used the measure of Arzheimer and Evans(2014) which generalized the measure of Martin et al. (2005)

* 이 논문은 2017학년도 청주대학교가 지원하는 연구년으로 연구되었음.

a) 청주대학교 통계학과 교수 류제복.

E-mail: jbyu@cju.ac.kr

using odds ratio. The result of analysis for 58 election polls registered in National Election Survey Deliberation Commission revealed that progressive or centrist candidates were overestimated and conservative candidates were severely underestimated. Based on the results of the election polls, odds were used to measure the chance of winning an election for candidates. In future election polls, it is also desirable to present the chance of winning an election as well as approval ratings for candidates.

Key words: multiple candidates, forecasting error, chance of winning, odds

I. 서론

선거가 끝나고 난 후에는 선거예측을 위한 각종 여론조사들에 대한 분석이 반드시 필요하다. 이는 다음 선거예측의 정확성을 높이기 위한 중요한 단계이다. 물론 선거예측이 실제 결과와 크게 차이가 나서 예측의 신뢰성이 떨어진 경우는 물론이고 예측이 잘된 경우도 마찬가지이다. 세밀한 분석을 통해 예측이 잘못된 경우는 그 원인을 찾아 향후 유사한 오류의 재발을 방지하고, 예측이 정확한 경우에는 장점을 최대한 살릴 수 있기 때문이다. 2017년도에 실시된 제19대 대통령 선거는 매우 이례적이었다. 박근혜대통령의 탄핵 사태로 당초 2017년 12월 20일로 예정된 선거가 6개월 이상 앞당겨진 2017년 5월 9일에 실시되었다. 이러한 특수한 상황에서 집권당이었던 새누리당은 자유한국당과 바른정당으로 분당되었다. 따라서 제19대 대통령 후보로 더불어민주당 문재인, 자유한국당 홍준표, 국민의당 안철수, 바른정당 유승민, 그리고 정의당 심상정 등이 출마하였다. 그러나 초반부터 더불어민주당 문재인 후보가 유력 후보로 부상하였고, 자유한국당 홍준표 후보와 국민의당 안철수 후보가 2위 다툼을, 그리고 바른정당 유승민 후보와 정의당 심상정 후보가 4위 다툼을 벌이는

형국이 되었다. 제19대 대통령선거는 과거와 다른 양상이어서 기존의 방법으로 선거예측의 정확성을 평가하는 데는 한계가 있다. 통상적으로 당선을 예측한 후보가 낙선하거나 반대인 경우에, 예측치가 오차 범위를 벗어났는지의 여부와는 관계없이 선거여론조사의 정확성 문제가 제기되곤 하였다. 그러나 이번에는 당선자를 정확히 예측한 때문인지 선거예측의 정확성에 대한 논란이 크게 제기되지 않았다. 하지만 통계학적으로 제19대 대통령선거 예측조사의 정확성을 상세히 분석할 필요가 있다.

〈표 1〉 제13대~제19대 대통령선거결과 5%이상 득표한 후보

(단위: %)

제 19 대 (2017 년)	더불어민주당 문재인 41.08	자유한국당 홍준표 24.03	국민의당 안철수 21.41	바른정당 유승민 6.76	정의당 심상정 6.17
제 18 대 (2012 년)	새누리당 박근혜 51.55	민주통합당 문재인 48.02	-	-	
제 17 대 (2007 년)	한나라당 이명박 48.67	대통합민주신당 정동영 26.14	창조한국당 문국현 5.82	-	
제 16 대 (2002 년)	새천년민주당 노무현 48.91	한나라당 이회창 46.58	-	-	
제 15 대 (1997 년)	새정치국민회의 김대중 40.27	한나라당 이회창 38.74	국민신당 이인제 19.20	-	
제 14 대 (1992 년)	민주자유당 김영삼 41.96	민주당 김대중 33.82	통일국민당 정주영 16.31	신정당 박찬중 6.37	
제 13 대 (1987 년)	민주정의당 노태우 36.64	통일민주당 김영삼 28.03	평화민주당 김대중 27.04	신민주공화당 김종필 8.06	

대한민국에서 선거여론조사가 실시되고 그 결과가 공표된 경우는 1987년의 제13대 대통령선거부터다. 1987년 민주화항쟁으로 국민이 직접 대통령을 선출한 제13대 대통령선거에서부터 2017년 제19대 대통령선거까지 선거결과를 정리해 보면 <표 1>과 같다. 1987년 제13대 대통령선거에서는 민주화 열기로 3김이 각자 대권에

도전하여 선거전도 치열하였으며, 그 결과도 과반수를 얻은 후보가 없었다. 제14대와 제15대에도 다수 후보로 선거가 실시되었다. 그러나 제16대부터 제18대까지는 선두 두 후보가 각축을 벌였고 군소 후보들은 선거에 거의 영향을 미치지 못하였다. 다수 후보에 대한 연구는 제13대부터 제15대에 걸쳐 이루어졌어야 했다. 하지만 당시에는 이에 대한 연구 기반이 갖추어져 있지 못했고, 연구에 필요한 자료들도 제대로 정리되어 있지 않았다. 제19대 대통령선거는 5% 이상 득표한 후보가 5명이었고, 대통령후보도 13명이나 되어 과거 어느 선거보다도 후보들이 많았다.

다수 후보의 경우는 실제 선거결과와 선거 전에 실시되는 여론조사 결과와의 차이가 발생할 수밖에 없다. 왜냐하면, 대부분의 유권자들은 선거 전에는 가능성이 있거나 자신이 지지하는 후보에게 투표하겠다는 의지를 나타내지만, 막상 선거일에 가까워지면 전략적인 선택을 하게 된다. 즉, 경합을 벌이는 선두 후보들 중에서 자신의 투표 의향을 변경하게 된다(Hummel 2014). 마찬가지로 후보들의 경우도 선거 막바지에 사퇴하고, 타 후보를 지지하는 상황도 발생한다. 대한민국의 경우, 제16대에는 정몽준 후보와 노무현 후보의 단일화를 통해 정몽준 후보가 사퇴하였고, 제18대에는 안철수 후보가 막판에 사퇴하여 다수 후보 체계가 최종에는 두 후보로 압축되기도 하였다. 따라서 대통령선거에서 다수 후보에 대한 선거예측의 정확성 분석이 제대로 이루어지지 못하였다. 류제복(2017)은 제15대와 제20대 국회의원선거에서 다수 후보에 대한 선거예측의 정확성을 분석하였다. 본 논문에서는 다수 후보에 대한 선거예측의 정확성과 당선 가능성을 제19대 대통령선거에서 분석해 본다.

제19대 대통령선거는 선두 후보가 확고하였지만 2위와 3위 후보는 경쟁을 벌이는 상태였고, 5% 이상을 득표한 후보도 5명이나 되었기 때문에 다수 후보의 경우를 고려해서 선거예측의 정확성을 살펴본다. 다수 후보들이 참여한 제19대 대통령 선거예측조사의 정확성을 평가하기 위해서 Martin et al.(2005)의 척도를 일반화한 Arzheimer & Evans(2014)의 척도와 총체적인 척도 B 와 B_w 를 사용한다.

한편 대한민국의 경우와 마찬가지로 국회의원 비례선거제도를 운영하고 있는 나라들에서 소수 정당의 성공과 실패는 비례대표 기준에 도달하느냐의 여부로 나누어진다. Wright et al.(2014)은 2011년 뉴질랜드 국회의원 선거자료를 이용해서 소규모 정당의 비례대표 기준인 5%를 넘어서는 가능성을 측정하기 위해서 오즈를

이용한 새로운 방법을 제시하였다.

본 연구에서는 58개의 2017년 대통령선거 여론조사자료(<http://www.nesde.go.kr/>)를 이용하여 득표율 상위 세 후보들에 대한 대통령 당선 가능성을 측정하였다.¹⁾ <표 1>에서 우리나라 대통령선거 결과를 보면 양자 대결의 경우는 득표율 50%를 그리고 다수 후보의 경우는 40%를 당선권으로 볼 수 있다. 2017년 제19대 대통령선거는 다수 후보로 치러졌으므로 40%를 당선권으로 간주하여 당선 가능성을 계산하였다.

2장에서는 선거여론조사에서 발생하는 예측오차를 측정하는 척도들과 척도들의 특징을 검토하고, 다수 후보의 경우 후보들의 당선 가능성을 측정하는 방법을 소개한다. 그리고 3장에서는 제19대 대통령선거에서 실시된 58개 선거여론조사에 대해 다수 후보를 3인과 5인의 경우로 나누어서 선거여론조사의 정확성을 비교·분석한다. 한편 선거여론조사결과를 바탕으로 후보들의 당선 가능성을 오즈를 이용해서 계산한다. 끝으로 4장에서는 연구결과를 정리하고 향후 선거여론조사에 대한 연구 방향을 검토한다.

II. 다수 후보에 대한 연구방법

1. 정확성 평가

선거예측의 정확성을 측정하기 위해 가장 널리 사용되는 척도는 Mosteller et al. (1949년)의 8가지 척도이다. 이 척도들은 1948년 미국 대통령선거에서 발생한 가장 대표적인 선거예측의 오류에 관한 연구에서 제시되었다.

Mitofsky(1998)는 Mosteller et al.(1949)이 제시한 8가지 척도들의 장·단점들을 상세히 기술하였다. 또한 예측오차는 부동층을 어떻게 배분하는가에 따라 영향을 받으므로 부동층을 적절히 배분하여 비교해야 한다고 주장하였다. 그러나 Mosteller et al.(1949년)의 척도들은 편향의 방향성을 파악하기가 어렵고, 최근과 같이 무응답

1) 여론조사결과의 결과공표금지 등에 대한 선거법 제108조 ①에 의해 선거일 전 6일부터 선거일의 투표마감 시각까지 여론조사결과를 공표할 수 없으므로, 2017년 5월 3일 전까지 선거여론조사공정심의위원회 홈페이지에 등록된 자료를 기반으로 하였다.

률이 높은 경우 무응답 처리방법에 따라 결과가 달라질 수 있다. 따라서 정확성과 함께 편향의 방향성을 동시에 측정할 수 있는 척도가 Martin et al.(2005)에 의해 제시되었다. 이들이 제시한 척도는 오즈비를 사용한 것으로 많은 장점을 갖고 있다. 최근에는 선거예측의 정확성 평가에 Martin et al.(2005)의 척도가 주로 사용되고 있다.

Martin et al.(2005)은 예측치와 결과치와의 차이를 기반으로 한 척도가 아니고 오즈비(odds ratio)를 근거로 한 척도를 제시하였다.

$$A = \ln[(p_r/p_d)/(v_r/v_d)] \quad (1)$$

이고 $Var(A) = \frac{1}{np_r p_d}$ 이다. 여기서 $p_r(p_d)$ 는 미국 대통령선거 여론조사에서 공화당(민주당) 후보를 지지하는 응답자의 비율이고 $v_r(v_d)$ 는 실제 선거에서 공화당(민주당) 후보의 지지율(득표율)이며 n 은 두 후보자를 지지하는 응답자 수이다. 척도 A 가 0이면 예측과 실제 투표결과가 완전 일치함을 나타내고, 0에서 멀어질수록 예측의 정확성은 떨어진다. 또한 A 가 유의적으로 음수이면 예측이 민주당 편향을 A 가 유의적으로 양수이면 반대로 예측이 공화당 편향을 나타낸다. 한편 후보가 3인의 경우, 선거여론조사에서의 3번째 후보를 지지하는 응답자 비율을 p_t , 실제 선거득표율을 v_t 라 하면 A' 는 다음과 같다.

$$A' = \ln[(p_t/(p_r + p_d))/(v_t/(v_r + v_d))] \quad (2)$$

이고 $Var(A') = \frac{1}{np_t(p_r + p_d)}$ 이다. A 의 경우와 유사하게, A' 가 유의적으로 양수이면 3번째 후보(정당)에 편향되고 유의적으로 음수이면 다른 두 후보(정당)에 편향됨을 보여준다.

한편 Arzheimer & Evans(2014)는 Martin et al.(2005)이 제시한 척도 A 가 다음과 같은 문제점이 있음을 지적하였다. 첫째는 로그오즈비가 기저정당에 대한 편향의 척도이기 때문에 A 에 대한 해석이 잘못될 수 있다. 예를 들면, 어떤 선거여론조사에서 기저정당 r 에 대한 지지율은 과대 예측된 반면에 다른 정당 a 에 대한 지지율은 정확하게 예측되었다고 가정하자. 즉, $p_a = v_a$ 인 반면에 $p_r > v_r$ 이 된다. 그러면 기저정당 r 에 관해 정당 a 에 대한 오즈비는 다음과 같이 된다.

$$OR_a = \left(\frac{p_a}{p_r} / \frac{v_a}{v_r} \right) = \left(\frac{p_a v_r}{p_r v_a} \right) = \frac{v_r}{p_r} \Leftrightarrow OR_a < 1 \quad (3)$$

이는, 정당 a 에 대한 지지율 예측이 편향없이 측정되었어도 결과는 부의 로그오즈비(negative log odds ratios)가 된다. 따라서 OR 에 기반한 척도는 기저정당의 선거편향에 비례해서 편향된다(물론 두 정당인 경우에는 한 정당을 과대 예측하면, 다른 정당을 과소 예측하게 되므로 이러한 경우는 발생하지 않는다. 그러나 다수 정당인 경우는 가능하다). 두 번째는, OR 값들이 기저정당에 관해서 계산되기 때문에, 이들의 합(또는 평균)은 기저정당의 선택에 의존하게 되어 전체적인 편향에 대한 유일한 척도를 얻을 수 없게 된다. 이러한 이유로, Arzheimer & Evans(2014)는 다수 정당(후보)으로 일반화한 다음과 같은 척도를 제시하였다.

$$\begin{aligned} A_i' &= \ln \left(\frac{p_i}{\sum_{j=1}^k p_j} / \frac{v_i}{\sum_{j=1}^k v_j} \right) \quad \text{for } j \neq i \\ &= \ln \left(\frac{p_i}{1-p_i} / \frac{v_i}{1-v_i} \right) \\ &= \ln \left(\frac{p_i}{1-p_i} \times \frac{1}{O_i} \right) \end{aligned} \quad (4)$$

여기서 A_i' 는 정당(후보) i 에 관한 편향을 나타내고 O_i 는 이 정당(후보)에 관한 투표결과에 대한 오즈(odds)(즉, $\frac{v_i}{1-v_i}$)이다. A_i' 는 delta 방법을 사용하면 점근적으로 정규분포를 한다. 즉 평균이 0이고 분산이 $\frac{1}{n v_i (1-v_i)}$ 이다.(Agresti 2002) 이때 n 은 분석에 사용될 후보들을 지지하는 표본 수이다. 그리고 $\sum_{i=1}^k p_i = \sum_{i=1}^k v_i = 1$ 이 되도록 조정해 준다.

A_i' 는 정당(또는 후보)에 대한 정확성과 편향의 정도를 나타내는 척도지만, 양수와 음수 값을 가질 수 있으므로 $\overline{A'}$ 는 부호가 서로 상쇄될 수 있어 전체적인 오차수준을 파악하기 어렵다. Arzheimer & Evans(2014)는 선거예측의 총체적인 편향성을 측정하기 위해서 절대값을 이용한 비가중 척도 $B = \frac{1}{k} \sum_{i=1}^k |A_i'|$ 와 득표율에 비례하는 가중척

도 $B_w(k) = \sum_{i=1}^k w_i |A'_i|$ 를 제시하였다. 다수 후보의 경우 득표율이 낮은 후보에 대한 부정확한 조사가 전체오차에 미치는 영향이 크게 되어 B 의 값이 부풀리게 될 수 있다. 이러한 점을 보완하기 위해서 득표율에 비례하는 가중척도 B_w 를 사용한다.

프랑스 대통령선거는 예비선거를 통해 최종 후보를 선정한다. 예비선거에서 다수의 후보들이 경합을 벌이고, 많은 선거여론조사가 실시된다. Durand(2008)는 2002년과 2007년 프랑스 대통령예비선거에서 다수 후보에 대한 선거예측의 정확성 분석에 Martin et al.(2005)의 척도 A' 를 사용하였다. 한편 Arzheimer & Evans(2014)는 다수 후보에 대한 선거여론조사의 정확성 평가를 위해 Martin et al.(2005)의 척도를 일반화한 새로운 척도를 제시하고, 이를 2012년 프랑스 대통령 예비선거조사에 적용하였다. 그리고 Cantú et al.(2016)은 2006년과 2012년 멕시코 대통령 선거에서 다수 후보에 대한 선거여론조사의 정확성을 평가하기 위해 Mosteller et al.(1949년)의 척도3과 척도5, Martin et al.(2005)의 척도, 그리고 Arzheimer & Evans(2014)의 척도 B 와 B_w 를 사용하였다.

대한민국과 같이 국회의원선거에서 비례대표제를 사용하고 있는 나라들에서도 다수 후보에 대한 선거예측의 정확성 평가가 이루어지고 있다. Wright et al.(2014)은 2011년 뉴질랜드 국회의원선거에 Martin et al.(2005)의 척도를 세 후보의 경우로 확장해서 사용하였다. 이들은 소수 정당들도 비례대표 선출이 가능하기 때문에 다수 정당(후보)에 대한 문제를 다루게 된 것이다. 한편 국내에서는 류제복(2017)이 Mosteller et al.(1949년)과 Martin et al.(2005)의 척도를 이용해서 다수 후보들이 각축을 벌였던 대한민국의 제15대와 제20대 국회의원선거 예측조사의 정확성을 비교·분석하였다.

이번 연구에서는 다수 후보에 대한 선거예측의 정확성과 편향성을 동시에 측정하기 위해 Martin et al.(2005)의 척도를 일반화한 Arzheimer & Evans(2014)의 척도를 사용하였다. 그리고 대한민국의 2017년도 대통령선거에서 실시된 선거여론조사결과를 이용하여 다수 후보에 대한 선거예측의 정확성과 편향성을 살펴보았다.

2. 당선 가능성 예측

선거여론조사를 실시하는 목적 중의 하나는 선거결과를 예측하는 것이다. 이를

위해서 지지율을 사용하지만, 지지율로 그 후보가 실제 당선될 가능성이 어느 정도인지를 정확하게 판단하기는 어렵다.

Wright et al.(2014)은 스포츠내기에서 오즈가 사용된 점에 착안하여, 2011년 뉴질랜드 국회의원 선거자료를 근거로 소규모 정당이 비례대표 기준을 넘어설 가능성을 측정하였다. 본 연구에서는 2017년도 제19대 대통령선거 여론조사결과를 바탕으로 후보들의 대통령 당선 가능성을 Wright et al.(2014)이 제시한 방법을 사용하여 계산한다. 제19대 대통령선거는 선두 후보의 윤곽이 일찍부터 드러났다. 반면에 2위와 3위 후보, 그리고 4위와 5위 후보는 비슷한 수준에서 각축을 벌이는 상황이었다. <표 1>을 보면 과거 대통령선거에서 다수 후보가 경합을 벌인 경우 40% 이상 득표하면 당선되었다. 제19대 대통령선거도 다수 후보인 경우이므로 대통령 당선권을 40%로 가정하고 선거여론조사결과를 바탕으로 선두 후보인 문재인 후보와 2위와 3위 후보인 홍준표 후보, 안철수 후보에 대해서 당선 가능성을 계산하였다.

Wright et al.(2014)에 따른 선거 오즈의 계산과정은 다음과 같다.

1. 신뢰할 수 있는 자료를 얻고(조사들을 결합하면 대규모 표본 크기로 인해서 선호된다) 관련된 후보(또는 정당)에 대한 평균과 표준오차를 계산한다.
2. 선거예측결과와 당선권인 40%와의 차이를 계산하고, 그 차이를 표준오차로 나누어 z -score를 구한다.
3. z -score를 단측 p -값으로 변환한다(Excel에서 표준정규분포함수를 이용).
4. p -값을 오즈로 변환한다; 당선권 이상의 후보에 대해서는 오즈 $p/(1-p)$, 그리고 당선권 이하의 후보에 대해서는 오즈 $(1-p)/p$ 를 계산한다.

Ⅲ. 사례분석

1. 다수 후보에 대한 정확성 분석

예측오차의 계산에 최종 선거결과와 예측값과의 차이를 이용하는 것이 전통적인 방법이다. 그러나 선거예측 당시의 지지율을 미래 시점의 선거결과와 비교하는 것은 조사시점의 상황을 제대로 반영하지 못하는 문제가 있다. 이러한 문제를 보완하

기 위해서 조사시점의 참 지지율을 예측하고 이 예측값과 조사시점의 조사결과를 이용한 예측값과의 차이로 오차를 계산한다.

Jackman(2005)은 확률보행모형(random walk model), 그리고 Cantú et al.(2016)은 역확률보행모형(reverse random walk model)을 이용해서 선거전 각 시점의 참 지지율을 예측하였다. 한편 박종희(2013)는 베이저안 상태공간모형(Bayesian state space model)을, 그리고 최필선·민인식(2013)은 공통확률추세(common stochastic trend)를 이용하였다. 그러나 이러한 방법은 사용된 모형에 따라 달라지고, 조사된 자료의 질에도 의존하게 된다. 따라서 본 연구에서는 각 시점별로 참 지지율을 예측하는 대신 최종 선거결과를 기반으로 선거예측조사의 정확성을 분석한다.

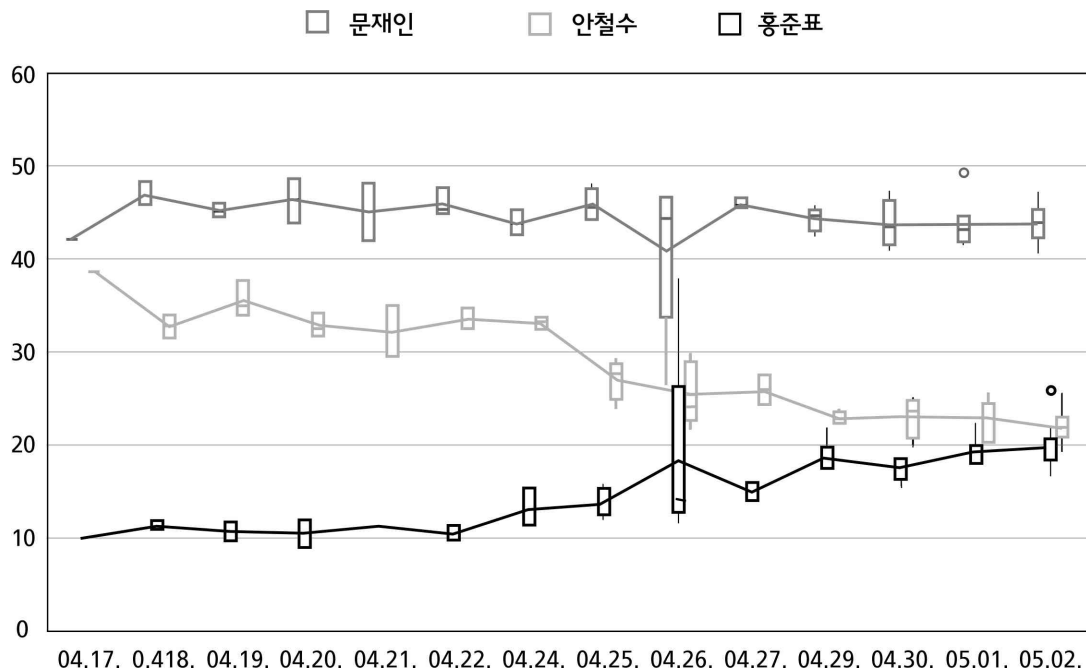
2017년 대통령선거여론조사를 위해 선거여론조사심의위원회에 등록한 총 65개 조사 중에서 지역을 대상으로 조사한 7개를 제외한 58개 조사결과를 정확성 분석에 사용하였다. 조사는 통상적으로 2~3일에 걸쳐서 실시되었기 때문에 최종 조사일을 기준으로 하였다. 조사결과 분석을 위해서 무응답은 각 후보들의 지지율에 비례배분하였다. 무응답을 배분하는 방법은 다양한데, 무응답에 대한 정보가 없는 상태에서는 비례배분이 가장 일반적으로 사용된다. 무응답을 조정하지 않고 실제 선거결과를 비교하게 되면 무응답으로 인해서 차이가 크게 발생하게 된다. 분석에 사용된 58개 조사에서의 평균 무응답률은 8.3%이고 최소 1.2%에서 최대 21.3%로 편차가 크다.

<그림 1>은 선두 3후보에 대한 조사시점별 지지율의 변동을 보여주고 있다. 같은 날 조사가 여러 차례 있었던 경우는 지지율에 대한 상자그림을 그렸고, 평균값들을 실선으로 연결하였다. 더불어민주당 문재인 후보의 지지율은 조사기간 내내 큰 변동이 없었다. 반면에 자유한국당 홍준표 후보는 시간이 경과함에 따라 지지율이 상승하고, 국민의당 안철수 후보의 지지율은 하락하는 추세였다. <그림 1>에서 4월 26일의 후보 지지율에서 편차가 매우 컸다. 이는 문재인 후보의 지지율은 급락한 반면에 홍준표 후보의 지지율이 급등한 26일 조사결과의 특이값 때문이다.²⁾

다음은 다수 후보에 대한 정확성과 편향을 평가하기 위해 Martin et al.(2005)의 척

2) 2017년 4월 26일에는 5차례의 조사가 있었는데, 영남일보가 의뢰하고 리얼미터가 실시한 조사결과가 문재인 후보 23.4%, 홍준표 후보 33.7%(무응답 보정 후는 각각 26.9%와 38.8%)로 나타났기 때문이다. 이것은 각 후보들의 정확성 척도인 A' 값에서도 특이값으로 나타났다.

도를 일반화한 Arzheimer & Evans(2014)의 척도 식(4)를 사용하였다. 식(4)를 사용하기 위해서 예측치와 실제 결과를 조정하였다. 3인 후보와 5인 후보의 경우 모두 예측치와 실제 선거결과의 합이 1이 되게 하였다. 즉, $\sum_{i=1}^k p_i = \sum_{i=1}^k v_i = 1$ ($k=3$ 또는 5). 그리고 표본크기는 3인 후보 또는 5인 후보를 지지한 응답자 수로 조정한다. 예를 들어 1,000명을 조사했는데, 3인 후보의 예측 지지율 합이 0.9이면 표본 수는 $1,000 \times 0.9 = 900$ 명으로 조정한다. 5인 후보도 3인 후보의 경우와 유사하다.



〈그림 1〉 조사 시점별 3후보의 지지율 변화(무응답 비례조정)

<표 2>에서는 일반화 척도인 식(4)를 이용하여 후보 수에 따른 정확성을 비교하였다. 58개 조사에서 후보들에 대한 예측오차는 문재인 후보가 가장 작았고, 홍준표 후보가 가장 컸다. 3인 후보든 5인 후보든 예측오차의 차이는 크지 않았지만 5인 후보의 경우에는 심상정 후보와 홍준표 후보가 비슷한 수준이었다. 예측의 편향 정도를 나타내는 $\overline{A_i}$ 는 양수인 경우는 과대 예측을 음수인 경우는 과소 예측을 나타낸다. 3인 후보의 경우, 문재인 후보와 안철수 후보는 과대 예측되었으며, 안철수

후보의 과대 예측 수준은 더 높았다. 반면에 홍준표 후보는 $\overline{A}_i' = -0.5673$ 으로 지나치게 과소 예측되었다. 5인 후보인 경우에도 홍준표 후보에 대한 과소 예측이 가장 크고, 유승민 후보도 지나치게 과소 예측되었다. 전체적으로 보수 후보에 대해서 과소 예측의 정도가 심하였는데, 이번 선거가 박근혜대통령 탄핵으로 인한 특수한 상황에 기인한 것으로 판단된다.

〈표 2〉 일반화 척도(A')를 이용한 후보별 정확성 비교

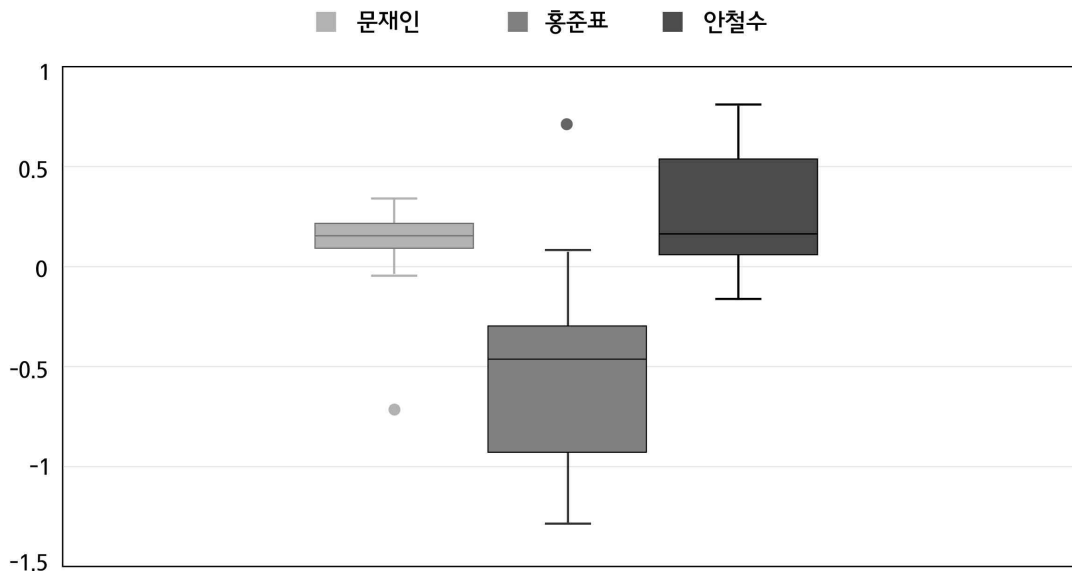
($N=58$)

구분	3인 후보			5인 후보				
	문재인	홍준표	안철수	문재인	홍준표	안철수	유승민	심상정
$\overline{A}_i'$	0.1473	-0.5673	0.2678	0.1458	-0.5401	0.2662	-0.3837	0.1413
표준오차	0.0195	0.0513	0.0338	0.0185	0.0471	0.0360	0.0307	0.0464
과대(n^+)	40	1	34	38	2	31	0	32
과소(n^-)	1	54	0	1	54	0	39	11
n	41	55	34	39	56	31	39	42
편향비율	0.707	0.948	0.586	0.672	0.966	0.534	0.672	0.741

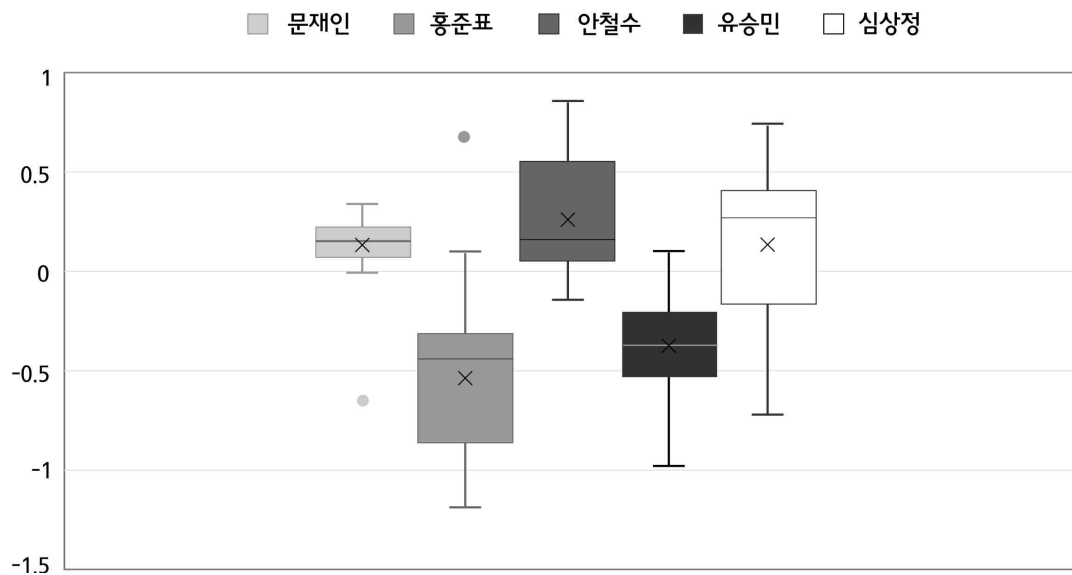
조사결과 편향은 각 후보들에 대한 A_i' 들을 구하고 이들의 95% 신뢰구간을 구해서 A_i' 들이 신뢰구간을 벗어난 경우가 된다. 과소 예측은 A_i' 가 신뢰하한보다 작은 경우이고 과대 예측은 신뢰상한보다 큰 경우이다. 3인 후보의 경우 홍준표 후보는 58개의 조사 중에서 55개가 편향되어 편향비율이 94.8%나 되었고 문재인 후보도 70.7%였다. 편향된 조사 중에서 문재인 후보와 홍준표 후보는 1개씩을 제외하고 모두 과대 또는 과소 예측되었다. 이는 <그림 1>을 통해서도 쉽게 알 수 있는데, 1개의 과소 또는 과대 예측된 것은 앞에서도 언급한 바와 같이 4월 26일 조사에서의 특이값 때문이다. 5인 후보의 경우도 문재인 후보와 안철수 후보의 편향 형태는 비슷한 수준이다. 하지만 안철수 후보와 유승민 후보는 편향된 조사 모두가 과대 또는 과소 예측되었다.

<그림 2>와 <그림 3>은 각각 3인 후보와 5인 후보의 일반화 척도값들에 대한 상자그림으로 편향의 정도를 쉽게 파악할 수 있다. <그림 2>에서 문재인 후보의

선거예측의 편차가 홍준표 후보나 안철수 후보에 비해 상당히 안정적임을 알 수 있다. 그리고 <그림 3>에서는 보수 후보인 홍준표 후보와 유승민 후보가 과소 예측된 반면에 나머지 세 후보는 과대 예측된 것을 알 수 있다.



<그림 2> 3인 후보의 일반화 척도(A')에 대한 상자그림



<그림 3> 5인 후보의 일반화 척도(A')에 대한 상자그림

<표 3>은 2017년 대통령선거 예측조사에서 나타난 총체적인 예측오차를 보여주고 있다. 58개의 2017년도 대통령선거 예측조사에서 다수 후보에 대한 선거예측의 정확성은 가중평균에 의한 오차가 작았지만, 후보 수에 따른 예측오차에는 차이가 없었다. 제19대 대통령선거 예측오차의 수준을 비교할 국내 자료는 없지만, 2006년과 2012년 멕시코 대통령선거 예측조사(Cantú et al. 2016)과 2012년 프랑스 대통령예비선거 예측조사(Arzheimer & Evans 2014)를 비교하면 대한민국 2017년 대통령선거 예측조사의 전체적인 오차는 상당히 높게 나타났다.³⁾

〈표 3〉 제19대 대통령선거예측조사의 총체적 예측오차

구분	$\bar{B}$	$sd(B)$	$\bar{B}_w$	$sd(B_w)$
3 인 후보	0.3535	0.1852	0.2841	0.1186
5 인 후보	0.3504	0.1391	0.2892	0.1245

2. 후보들의 당선 가능성

2017년 대통령선거에서 실시된 58개의 선거여론조사를 근거로 해서 문재인, 홍준표, 그리고 안철수 선두 세 후보의 대통령 당선 가능성을 살펴본다. 당선 가능성을 측정하기 위해서 사용된 표본은 단순확률추출된 경우로 가정하였다. 실제 사용된 58개의 조사자료에는 표본추출에 대한 명확한 언급이 없어서 단순확률추출을 가정하였는데, 이러한 가정으로 실제보다 표본오차가 크게 반영되어 예측확률이 달라질 수 있다.

선거 오즈를 계산하기 위해서는 58개 조사 자료로부터 가중 예측치를 구한다. 실제 선거조사에 사용될 오차에 대한 표준편차는 이론적인 정규근사에 의한 표준오차보다 크므로 이를 반영해서 수정(수정계수 k)해 주어야 한다. Buchanan(1986)은 155개 조사로부터 계산된 오차들의 표준편차가, 이론적인 정규근사를 이용하여 계산된 표준오차의 2배가 됨을 보였고, Wright et al.(2014)은 $k = 2.07$ 을 사용하였

3) 후보 수와 조사 수 등 세부적인 자료의 부족으로 정확한 비교·분석에는 한계가 있음.

다. 본 연구에서는 세 후보 각각에 대해서 58개의 조사결과를 사용하였다. 문재인, 홍준표, 그리고 안철수 세 후보에 대한 조사오차를 계산하기 위한 수정계수 k 는 각각 1.84, 3.91, 그리고 4.13이다. k 값이 차이가 나는 이유는 문재인 후보에 비해 홍준표 후보와 안철수 후보의 실제 조사오차의 표준편차가 크기 때문이다.

당선 가능성을 계산하는 절차는 다음과 같다.

1. 58개 조사에서 무응답을 조사결과 지지율에 따라 비례배분한다.
2. 58개 조사에서 각 후보에 대해 $S_i = \sqrt{\frac{p_i(1-p_i)}{n_i}}$ 와 이들의 평균 $S = \frac{1}{58} \sum_{i=1}^{58} S_i$ 를 구한다.
3. 58개 조사에서 각 후보에 대한 $e_i = |\hat{p}_i - p|$ 를 계산하고 이들의 표준편차 S_e 를 구한다.
4. 위의 2와 3으로부터 후보별로 $k = \frac{S_e}{S}$ 를 구한다. 여기서 k 는 수정계수로 실제 조사에 대한 표준오차를 구하는 데 사용된다.
5. 세 후보에 대한 지지율 $\hat{p}$ 를 각 조사의 표본크기로 가중해서 구한다. 가중 지지율은 문재인 후보 $\hat{p} = 0.44$, 홍준표 후보 $\hat{p} = 0.16$ 그리고 안철수 후보 $\hat{p} = 0.26$ 이다. 58개 조사에서의 평균 표본크기는 $n = 1,296$ 이다.

2017년도 대통령선거에서 유력 후보가 3인 경우 <표 1>에서와 같이 당선권을 득표율 40%로 본다면 문재인 후보의 지지율은 전체적으로 당선권인 40%선을 웃돌고 있었고 홍준표 후보와 안철수 후보의 지지율은 40%선에 훨씬 못 미치는 상황이었다.

문재인 후보의 당선 가능성에 대한 계산과정은 다음과 같다.

6. $\hat{p} = 0.44$ 인 경우

$$Z = \frac{\hat{p} - 0.4}{S_e} = \frac{\hat{p} - 0.4}{1.84 \times \sqrt{\frac{0.44 \times 0.56}{1,296}}} = \frac{0.44 - 0.4}{0.0254} = 1.5748 \text{ 이므로 엑셀에}$$

서 표준정규분포를 이용해서 Z 까지의 누적확률 $p = 0.9421$ 를 구한다.

7. 6에서 구한 p 를 이용하면 오즈 $p / (1 - p) = 16.3$ 이 된다. 즉, 문재인 후보의 지지율이 $\hat{p} = 0.44$ 인 경우 대통령 당선 오즈는 16.3:1로 문재인 후보의 대통령 당선 가능성이 낙선 가능성의 16.3배가 된다는 의미이다. 그러나 홍준표 후보와

안철수 후보의 지지율은 당선권인 40% 이하므로 문재인 후보와는 달리 오르는 $(1-p)/p$ 이 된다. 이는 정확하게 표현하면 당선권에 못 미칠(낙선) 가능성이다. 같은 절차로 안철수 후보의 가중 지지율이 $\hat{p} = 0.26$ 인 경우 낙선 오르는 368:1인데 안철수 후보의 낙선 가능성이 당선 가능성의 368배이므로 당선 가능성을 1:368로 표기한다.

〈표 4〉 지지율에 따른 세 후보의 당선 가능성

지지율	문재인	홍준표	안철수
0.45	39.1:1		
0.44	16.3:1		
0.43	7.4:1		
0.42	3.7:1		
0.41	1.9:1		
0.40	1:1	1:1	1:1
0.38		1:1.8	1:1.8
0.36		1:3.5	1:3.3
0.34		1:7.2	1:6.4
0.32		1:16.4	1:13.8
0.30		1:43.7	1:33.9
0.28		1:142.1	1:99.6
0.26		1:599.1	1:368.0
0.24		1:3,505.1	1:1,820.7
0.22		1:31,053.0	1:13,061.6
0.20		1:469,331.5	1:151,302.6
0.18		1:14,360,784.3	1:3,301,064.7

이번 선거예측조사에서의 평균표본크기는 $n = 1,296$ 로 가정하고, 지지율 변화에 따른 세 후보들에 대한 당선 가능성은 <표 4>에 있다. <표 4>에 의하면 2017년도

대통령선거에서 홍준표 후보와 안철수 후보의 지지율이 대부분 20%대 전후였던 것을 감안하면 당선 가능성은 거의 없다고 볼 수 있다. 하지만 같은 지지율에서도 홍준표 후보와 안철수 후보의 당선 가능성이 다른 이유는 두 후보에 대한 조사오차의 차이 때문이다. 즉, 수정계수 k 의 차이로 인한 것이다. <표 4>로부터 2017년도 대통령선거 여론조사에서 홍준표 후보와 안철수 후보에 대한 선거예측의 정확성이 매우 낮았던 것을 재확인할 수 있다(<표 2>, <그림 2>).

당선 가능성은 경쟁 후보가 2인인 경우에도 득표율 50%를 당선권으로 간주하고 계산할 수 있는데, 실제 선거예측이 정확할수록 당선 가능성도 보다 정확하게 측정된다.

지금까지는 선거여론조사에서 각 후보들에 대한 지지율을 제공하였는데, 이러한 지지율만으로는 후보의 당선(또는 낙선) 가능성을 구체적으로 설명할 수 없었다. 그러나 이번과 같이 오즈를 이용한다면, 지지율과 함께 후보의 당선 가능성을 측정할 수 있다. 본 연구를 통해 향후 선거예측조사에서 지지율과 함께 당선 가능성을 측정할 수 있고, 선거기간 중에 조사가 누적되면 많은 조사 자료를 토대로 보다 정확한 정보를 국민들에게 제공할 수 있다.

IV. 결론

2017년에 실시된 제19대 대통령선거는 박근혜대통령의 탄핵이라는 특수한 상황에서 조기에 실시되었고, 경쟁 후보가 없는 상황에서 문재인후보의 독주로 일찍부터 당선이 거의 확실한 상황이었다. <그림 1>에서와 같이 선거기간 내내 많은 선거여론 조사가 실시되었지만 문재인 후보의 지지율은 큰 변동 없이 당선권을 웃돌고 있었다. 반면에 안철수 후보에 대한 지지율은 선거일에 가까울수록 하락하였고 상대적으로 낮은 지지율을 보였던 홍준표 후보의 지지율은 상승하였다.

이러한 지지율의 현격한 차이로 문재인 후보의 대통령 당선이 당연히 되고 타 후보들의 당선 가능성이 희박한 상황이었기 때문에 대통령선거 예측조사에 대한 정확성 분석이 미흡하였다. 그러나 이번 대통령선거는 1987년도의 제13대 대통령선거와 같이 득표율이 20%가 넘는 후보가 3인이었고, 5% 이상 득표한 후보도 5인이거나 되는 특별한 경우였다. 따라서 종전의 2인 후보에 대한 대통령선거 예측조사의

정확성 분석을 다수 후보로 확대하여 검토할 필요가 있다. 다수 후보에 대한 연구는 류제복(2017)이 제15대와 제20대 국회의원선거 예측조사에서 다룬 바 있는데, 본 연구에서는 Martin et al.(2005)의 척도를 일반화한 Arzheimer & Evans(2014)의 척도를 사용하여 3인과 5인 후보인 경우를 다루었다.

3인 후보의 경우, 58개 조사 중에서 문재인 후보는 1개 과소, 40개가 과대 예측되었고, 안철수 후보도 과대 34개로 과대 예측비율이 각각 69.0%와 58.6%였다. 반면에 홍준표 후보는 과대 1개, 과소 54개로 과소 예측비율이 무려 93.1%나 되었다. 5인 후보의 경우에도 문재인 후보, 안철수 후보, 심상정 후보에 대한 과대 예측비율은 각각 65.5%, 53.4%, 55.2%인 반면에 홍준표 후보와 유승민 후보에 대한 과소 예측비율은 93.1%와 67.2%로 극명하게 차이가 났다. 보수 성향의 후보에 대해서 과소 예측이 심하였는데, 과거 선거여론조사에서 집권여당을 과대 예측한 것과는 다른 양상이었다. 이는 이번 선거가 박근혜대통령 탄핵으로 인한 특수한 상황에 기인한 것으로 판단된다. 비록 2017년도 대통령선거의 당선자는 국민들의 기대에 어긋나지 않았지만, 선거예측의 정확성은 매우 낮다. 정확한 비교는 어렵지만, 류제복(2017)의 2016년도 국회의원선거에서 정당별 정확성과 비교하여도 이번 대통령선거 예측조사의 정확성은 매우 낮은 편이다.

한편 <표 3>은 2017년 대통령선거 예측조사에서 나타난 총체적 예측오차를 보여주고 있다. 다수 후보에 대한 선거예측의 정확성은 가중평균에 의한 오차가 작았지만, 후보 수에 따른 예측오차에는 차이가 없었다. 총체적인 오차의 수준을 비교할 국내 자료는 없지만, 2006년과 2012년 멕시코 대통령선거 예측조사(Cantú et al. 2016)과 2012년 프랑스 대통령예비선거 예측조사(Arzheimer와 Evans 2014)를 비교하면 2017년 대통령선거 예측조사의 총체적인 오차는 상당히 높게 나타났다.

선거예측조사에서 후보들에 대한 지지율이 얼마인가가 관심이지만, 한편으로는 그 후보의 당선 가능성이 어는 정도인지도 매우 궁금하다. 그동안 당선 가능성에 대한 측정이 시도되지 않았지만, Wright et al.(2014)은 오즈를 이용해서 국회의원선거에서 소규모 정당의 비례대표 기준을 넘어서는 가능성을 측정하였다. 본 연구에서는 2017년도 제19대 대통령선거 여론조사결과를 바탕으로 후보들의 당선 가능성을 측정하였다. <표 1>의 대한민국 대통령선거에서 양자 대결인 경우는 50% 득표율을, 그리고 3자 대결인 경우는 40% 득표율을 당선권으로 볼 수 있다. 그런데 2017년

대통령선거는 3자 대결로 볼 수 있으므로 40% 득표율을 당선권으로 간주하고, 후보별 당선 가능성을 살펴보았다. 무응답 비례배정 후, 문재인 후보는 선거기간 내내 40% 이상의 지지율을 기록한 반면에, 홍준표 후보와 안철수 후보는 40% 이하에 머물렀으므로 이에 따라 당선 가능성을 계산하였다. 20%대 전후의 지지율을 기록했던 홍준표 후보와 안철수 후보의 당선은 거의 불가능한 상태로 계산되었다. 그러나 홍준표 후보와 안철수 후보 간에도 당선 가능성이 차이가 났는데, 이는 두 후보에 대한 예측오차의 차이로 인한 것이다. <표 2>에서 본 바와 같이 홍준표 후보는 지나치게 과소 예측되었고, 안철수 후보는 과대 예측된 것이 결국 두 후보의 당선 가능성에도 영향을 준 것이다.

본 연구는 다수 후보에 대한 선거예측의 정확성 분석에 Martin et al.(2005)의 척도를 일반화한 Arzheimer & Evans(2014)의 척도를 이용하여 제19대 대통령선거 예측의 정확성을 분석하였고, 그동안 지지율만을 제시한 것에 덧붙여 조사결과를 기반으로 후보들의 당선 가능성을 측정하였다. 국회 교섭단체가 4개가 된 2018년도 상황에서 기존의 양당 체계의 연구를 다수 정당(후보)체제로 확대할 필요가 있고, 이를 바탕으로 당선 가능성을 수량화해서 제공하는 것도 의미가 있을 것이다. 다만, 당선 가능성을 정확하게 측정하려면, 정확한 예측조사가 전제되어야 한다.

참고문헌

- 류제복. 2017. “다수 후보에 대한 선거예측의 정확성 분석: 제15대와 제20대 국회의원선거를 중심으로.” 《조사연구》 18(3): 1-22.
- 박종희. 2013. “제18대 대선 여론조사에서 나타난 조사기관 편향.” 《조사연구》 14(1): 1-30.
- 중앙선거관리위원회. <http://www.nec.go.kr/>
- 중앙선거여론조사심의위원회. <http://www.nesde.go.kr/>
- 최필선·민인식. 2013. “18대 대통령 선거 여론조사의 기관별 정확성 측정 및 비교.” 《조사연구》 14(2): 1-18.
- Agresti, A. 2002. *Categorical Data Analysis*. 2nd ed. John Wiley & Son, Inc.

- Arzheimer, K. and J. Evans. 2014. "A New Multinomial Accuracy Measure for Polling Bias." *Political Analysis* 22: 31-44.
- Buchanan, W. 1986. "Election Predictions: An Empirical Assessment." *Public Opinion Quarterly* 50(2): 222-227.
- Cantú, F., V. Hoyo, and M.A. Morales, 2016. "The Utility of Unpacking Survey Bias in Multiparty Election: Mexican Polling Firms in the 2006 and 2012 Presidential Elections." *International Journal of Public Opinion Research* 28(1): 96-116.
- Durand Claire. 2008. "The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2001 Catastrophe?" *International Journal of Public Opinion Research* 20(3): 275-298.
- Hummel, P. 2014. "Pre-election Polling and Third Party Candidates" *Social Choice & Welfare* 42(1): 77-98.
- Jackman, S. 2005. "Pooling the Polls over an Election Campaign" *Australian Journal of Political Science* 40(4): 499-517.
- 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, W.J. 1998. "The Poll-Review; Was 1996 a Worse Year for Pools than 1948?" *Public Opinion Quarterly* 62(2): 230-249.
- Mosteller, F., 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.
- Wright, M.J., D.P. Farrar, and D.F. Russell. 2014. "Polling Accuracy in Multiparty Election." *International Journal of Public Opinion Research* 26(1): 113-124.

<접수 2018/04/25, 수정 2018/06/04, 게재확정 2018/06/10>