

# A case study for alternative methods of election forecasting

Jea-Bok Ryu<sup>a,1</sup>

<sup>a</sup>Department of Statistics, College of Science & Engineering, Cheongju University

(Received May 12, 2020; Revised May 22, 2020; Accepted May 22, 2020)

---

## Abstract

We compared and analyzed the relationship between vote intention, vote expectation, and party approval rate using the 19th Presidential Election and the 7th Nationwide Simultaneous Local Election (Regional Local Government Election) poll data. The case study provides an alternative method of predicting the winner using vote expectation and a party approval rate that can improve the accuracy of election forecasting.

Keywords: vote expectation, vote intension, party approval rate, mean absolute error

---

## 1. 서론

선거에서 국민들의 가장 큰 관심사는 어느 후보가 당선될 것 인가이고, 다음으로는 그 후보가 국민들로부터 받은 지지의 정도이다. 그렇다면, “당신은 이번 선거에서 누가 당선될 것이라고 생각합니까?”라는 당선가능성 질문이 우선되어야 한다. 그러나 중앙선거여론조사심의위원회 홈페이지에 등록된 제19대 대통령선거(2017년 5월 9일)에서 후보자가 확정된 이후 실시된 67건의 선거여론조사에서 당선가능성 질문이 포함된 조사는 19건으로 전체의 28.4%에 불과하다. 그리고 2018년 6월 13일에 실시된 제7회 전국동시지방선거에서도 총 571건의 선거여론조사가 등록되었지만 120건의 광역단체장선거 여론조사를 제외하고는 당선가능성 질문이 거의 없었다. 광역단체장 선거여론조사에서도 30.8%인 37건의 선거여론조사에서만 당선가능성 질문이 포함되었다. 물론 과거 선거여론조사에서는 그 비율이 더욱 낮았기 때문에 우리나라에서 당선가능성을 이용한 선거예측 사례는 없는 실정이다.

미국은 선거여론조사에서 당선가능성 질문이 1932년 대통령선거부터 사용되었지만 당선가능성 자료를 이용하여 선거예측에 사용된 경우는 매우 적었다. 그 이유는 후보자의 당선 여부가 유권자들이 실제 선거에서 투표함으로써 결정되기 때문에 후보자의 득표율을 추정하게 된다. 지금까지 대부분의 선거예측 조사에서는 “만약 내일이 선거일이라면 당신은 누구에게 투표할 것입니까?” 또는 “당신은 후보들 중에서 어느 후보가 당선되었으면 합니까?” 등과 같이 투표의향(vote intension)이나 선호도(preference) 질문을 통한 후보지지율로 후보들에 대한 득표율을 추정하여 당선자를 예측하였다. 그러나 투표의향이나 선호도 질문에 대해 응답자들은 실제 그 후보의 당선 여부와 관계없이 응답할 수 있다. 예를 들어 응답자가 후보자가 속한 정당의 당원이거나 후보자와 개인적 관계(혈연, 지연, 학연 등)가 있는 경우 또는

This work was supported by the research grant of Cheongju University (2019.03.01.–2021.02.28.).

<sup>1</sup>Department of Statistics, College of Science & Engineering, Cheongju University, 298 Daeseong-ro, Cheongwon-gu, Cheongju, Chungbuk 28503, Republic of Korea. E-mail: [jbryu@cju.ac.kr](mailto:jbryu@cju.ac.kr)

후보자와 정치적 성향이 같을 때는 자신이 선택한 후보자의 당선 여부와 관계없이 응답하게 된다. 당선 가능성(vote expectation) 질문이 실제 투표행위와는 다르기 때문에 당선가능성에 대한 응답으로 득표를 추정이 모호하고 이 질문에 대한 응답의 신뢰성을 담보하기 어려웠다. 또한 당선가능성 질문에 대한 응답은 단순히 선거결과에 대해 응답자 자신이 생각하고 있는 주관적인 추측이지 그 이상은 없다고 생각하였다 (Hayes, 1936). 그러나 실제로 당선가능성 질문에 대해 응답자들은 자신의 투표의향과 친구, 가족, 동료는 물론 언론 등의 다양한 정보를 참고하여 보다 객관적으로 응답하게 된다. 따라서 최근 들어서는 당선가능성을 이용한 선거예측이 점차 증가하고 있다 (Lewis-Beck과 Skalaban, 1989; Lewis-Beck과 Stegmaier, 2011; Lewis-Beck과 Tien, 1999; Graefe, 2014; Miller 등, 2012; Murr, 2011, 2015; Rothschild와 Wolfers, 2013).

본 연구에서는 후보지지(투표의향) 조사와 당선가능성 조사의 정확성을 비교하고, 당선가능성으로부터 득표율에 대한 추정식을 유도한다. 또한 Ryu (2019)에서 다룬바와 같이 정당지지율을 함께 고려하였다. 당선가능성, 후보지지율, 그리고 정당지지율과의 관계들을 살펴보고, 3가지 방법들을 비교한다. 2장에서는 당선가능성에 대한 선행 연구들을 검토하고 3장에서는 제19대 대통령선거자료와 제7회 전국동시지방선거자료를 이용해서 새로운 선거예측 방법의 적합성을 분석한다.

## 2. 선행 연구

1932년 미국대통령선거에서 선두 세 후보에 투표하겠다는 사람들에게 자신의 지지후보와 관계없이 어느 후보가 당선될 것인가라는 질문을 하여 얻은 결과가 Table 2.1에 제시되어 있다 (Hayes, 1936; 위키백과). Table 2.1에서 경합을 벌이고 있지 않은 Thomas 지지자들은 보다 객관적으로 Roosevelt의 당선가능성을 매우 높게 보고 있다.

지지자들의 당선가능성과 후보지지율(또는 후보선호율)의 차이가 발생하는데, 이를 기대편향(wishful thinking bias)이라 한다. 하지만, 유권자들이 생각하는 당선가능성은 그들의 후보지지율에 영향을 받는다. Granberg와 Brent는 당선가능성과 후보지지율과의 상관이 0.8로 양의 상관관계가 있음을 보여주고 있다 (Graef, 2014, p.208). 한편 Rothschild와 Wolfers (2013)에 의하면 1952년부터 2008년까지 모든 주 단위의 미국대통령선거자료를 사용한 경우는 0.42이다. 그러나 응답자(일반 시민들)가 증가하면 개인 편향은 그룹에 의해 상쇄되어 기대편향은 없어지게 된다. 일반 시민들의 판단이 매우 현명하다는 집단지성으로 당선자를 매우 정확하게 예측할 수 있는데, 이를 입증하기 위해 Condorcet's Jury Theorem이 이용된다 (Murr, 2011, 2015; Graefe, 2014).

Lewis-Beck과 Skalaban (1989)은 1956년부터 1984년까지 8번의 미국대통령선거에서 유권자들의 69%가 대통령당선자를 정확하게 예측했음을 보여주고 있다. 물론 시민들의 예측은 1976년 대통령 선거의 경우를 제외하고는 그들의 예측이 랜덤(귀무가설이 50:50)하지 않았는데, 이는 시민들의 예측이 정확하다는 것을 의미한다. 그리고 이들은 집계된 자료를 이용해서 시민들의 예측 정확성을 대통령 당선자가 얻은 총 득표율(%)에 대한 회귀식으로 설명하였다.

$$F = -130.85 + 3.63S \quad (2.1)$$

$$(-4.65) \quad (7.14)$$

$$R^2 = 0.895, \quad N = 8,$$

여기서  $F$ 는 대통령 당선자를 정확하게 예측한 예비선거조사의 백분율이고  $S$ 는 대통령 당선자가 얻은 총 득표율(%)이며 괄호 속은  $t$ 값이다. 그리고  $R^2$ 은 결정계수이고  $N$ 은 사례수이다.

식 (2.1)은 대통령 당선자가 얻은 총 득표율( $S$ )이 커질수록 대중들의 예측 정확성( $F$ )이 커진다는 가설

**Table 2.1.** 1932 US presidential election

(단위: %)

투표의향	실제 득표결과	Hoover 당선가능성	Roosevelt 당선가능성
Hoover(공화당)	39.7	72.0	23.4
Roosevelt(민주당)	57.4	6.5	90.7
Thomas(사회당)	2.2	25.3	70.0

을 뒷받침해준다. Lewis-Beck과 Tien (1999)은 1956년부터 1996년까지 11번의 미국 대통령선거에서 당선가능성 조사는 당선자를 71% 정확하게 예측함을 보여주었고, 당선가능성으로부터 양당제에서 집권당 후보의 득표율  $V$ 를 예측하기 위한 회귀식을 유도하였다.

$$V = 39.45^* + 0.21^*E \quad (2.2)$$

$$(13.55) \quad (5.11)$$

$$R^2 = 0.74, \quad SEE = 3.2, \quad N = 11.$$

$E$ 는 American National Election Studies (ANES)에서 집권당 후보의 당선가능성(%)이다. \*는 유의수준 1%에서 통계적으로 유의함을 나타내며 괄호 속은  $t$ 값이고, SEE는 추정치의 표준오차이다. 그리고  $R^2$ 과  $N$ 은 각각 결정계수와 사례수이다.

Graefe (2014)는 1988년부터 2012년까지 7차례의 미국 대통령선거에서 선거 마지막 100일에 걸쳐, 당선가능성 조사가 4개의 입증된 방법들(후보지지율조사, 예측시장, 양적모델, 그리고 전문가판단)보다 득표율과 당선자를 보다 정확하게 예측했음을 보여주었다. 1932년부터 2012년까지 217개의 당선가능성 조사를 사용해서 득표율을 추정하였다. 추정된 회귀식은 다음과 같다.

$$V = 41.0 + 0.17E \quad (2.3)$$

$$(20.4) \quad (78.5)$$

$$R^2 = 0.66, \quad SEE = 2.2$$

여기서  $V$ 는 득표율(%)이고,  $E$ 는 (양당제의 경우)집권당 후보에 대한 당선가능성이고 괄호 속은  $t$ 값이다. 그리고  $R^2$ 는 결정계수이고 SEE는 표준오차이다.

식 (2.2)와 식 (2.3)은 당선가능성을 이용해서 후보자의 득표율을 추정한 식인데, 둘 다 최소 득표율이 40% 수준이다. 이는 Campbell (1996)이 주장한 “미국과 같은 양당제에서는 선거전을 아무리 못해도 최소 40%를 득표할 수 있고, 또한 아무리 잘 해도 60%를 넘기 어렵다”는 주장과 일치한다. 실제로 1932년부터 2016년까지 22번의 미국 대통령선거에서 당선자의 득표율이 60%를 초과한 경우는 1936년 민주당 Roosevelt(60.8%), 1964년 민주당 Johnson(61%), 그리고 1972년 공화당 Nixon(60.7%)의 단 3번이었고 최소득표율이 40%에 못 미치는 경우는 5번(36.5%–39.7%)이었다(위키백과 미국대통령선거 결과).

영국의 경우에도 당선가능성에 근거를 둔 집단예측(citizen forecasting)의 정확성 분석을 위해 Lewis-Beck과 Stegmaier (2011)는 1951년부터 2005년까지 13번의 영국 총선거 자료를 사용하였는데, 응답자의 60.2%가 선거결과를 정확하게 예측하였다. 그리고 그들은 2005년도까지의 자료를 이용해서 2010년 선거에서 국회의원 의석수를 예측하는 식도 유도하였다.

Rothschild와 Wolfers (2013)도 다양한 방법들을 이용해서 득표율을 예측하였는데, 당선가능성을 기반으로 한 예측이 후보지지율보다 정확하고 당선가능성과 후보지지율과의 차이도 통계적으로 유의하였다. 회귀분석을 이용하거나 최적가중을 이용한 득표율 예측의 경우도 당선가능성이 후보지지율보다 훨씬 기

**Table 3.1.** Correlation in the 19th presidential election poll

	당선가능성	후보지지율	정당지지율
당선가능성	1		
후보지지율	0.5326	1	
정당지지율	0.5092	0.8734	1

여도가 높았다. 또한 이들에 의하면, 2004년 Bush와 Kerry의 경우 후보지지율은 변동이 심했지만 당선가능성은 시종일관 Bush가 재선에 성공할 것이라고 예측했다. 2008년 Obama의 경우도 선거 막바지 17개 조사의 당선가능성은 Obama의 승리를 예상하였다. Rothschild와 Wolfers (2013)는 당선가능성 질문으로부터 얻는 정보의 양이 후보지지 질문으로부터 얻는 정보양보다 훨씬 많음을 수리적 모형을 이용하여 유도하였다. 그리고 그들은 극단적으로 치우친 표본의 경우와 소표본에서도 당선가능성 조사를 이용해서 선거예측을 정확하게 할 수 있다는 것을 보여주었다.

Graefe (2016)는 2013년 독일 선거예측자료로부터 대표성이 부족한 당선가능성 조사도 유용한 저비용의 예측방법을 제공할 수 있고, 또한 소규모 선거에서는 기존의 대표적인 여론조사나 예측시장(prediction markets)같은 인정된 방법들이 적합하지만 비용면에서는 효과적이지 않을 수 있음을 주장하였다.

### 3. 사례연구

2장에서 언급한 바와 같이 국내에서는 당선가능성 질문을 이용한 선거예측에 대한 연구가 거의 전무한 상태이다. 물론 미국이나 독일, 영국 등과 같은 일부 국가에서는 당선가능성 질문을 이용한 연구가 점차 증가하고 있다.

이번 연구에서는 중앙선거여론조사심의위원회 홈페이지에 수록되어 있는 선거여론조사자료(<http://www.nesde.go.kr/>) 중에서 제19대 대통령선거와 제7회 전국동시지방선거에서 실시된 선거여론조사자료들을 사용하였다. 이들 자료를 이용하여 당선가능성, 후보지지율, 그리고 정당지지율과의 관계분석을 통해 새로운 선거예측 방법의 가능성을 검토해 본다.

#### 3.1. 제19대 대통령선거

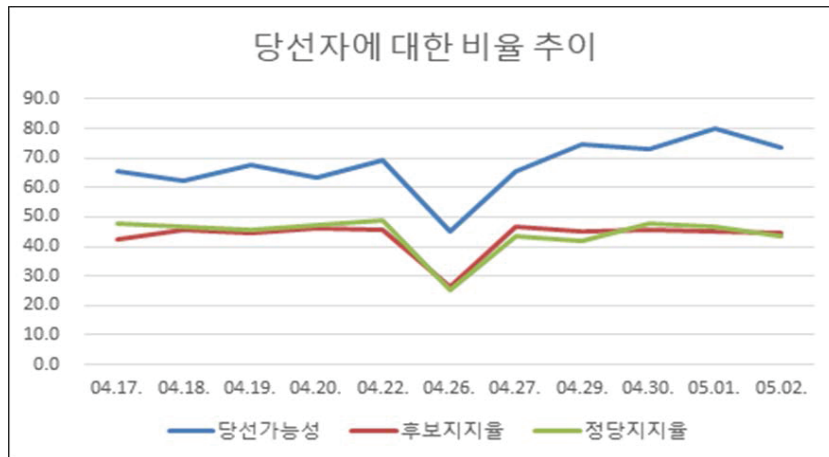
공직선거법 제108조 제7항에 따라 중앙선거여론조사심의위원회 홈페이지에 등록된 제19대 대통령선거 관련 여론조사 중 후보자 확정 이후인 2017년 4월 17일부터 5월 9일까지 실시된 67건 중에서 당선가능성, 후보지지율 그리고 정당지지율 조사가 동시에 이루어진 19건의 조사를 대상으로 분석하였다.

Ryu (2019)는 정당지지율을 이용한 선거예측의 정확성과 편향성을 후보지지율과 비교분석하였다. 시험적 연구였지만 후보지지율을 기반으로 한 전통적인 방법에 대한 대안으로 연구 검토할 가치가 있다고 판단하였다. 본 연구에서도 정당지지율을 함께 고려하였다.

Table 3.1은 제19대 대통령선거여론조사자료로부터 당선가능성, 후보지지율, 그리고 정당지지율에 대한 상관관계를 보여주고 있다. 여기서 사용한 자료는 응답자 개인별 자료를 이용할 수 없어 조사별 집계자료를 사용하였다. 따라서 외국의 사례연구와 직접 비교하는 데는 한계가 있다. 하지만 제19대 대통령선거자료에서 얻어진 상관관계 0.53은 Rothschild와 Wolfers (2013)가 구한 상관관계 0.42보다는 높고, Granberg와 Brent가 구한 상관관계 0.8보다는 낮은 수준이다 (Graefe, 2014, p.208). 당선가능성과 정당지지율과의 상관도 후보지지율과의 상관과 비슷한 수준인 0.51이다. 그러나 후보지지율과 정당지지율과의 상관은 0.87로 매우 높다.

**Table 3.2.** The 19th presidential election poll

	당선가능성	후보지지율	정당지지율
평균	70.01	44.20	44.88
표준편차	9.44	4.74	5.45
최소값	44.88	26.35	25.15
최대값	84.77	49.39	50.91
관측수	19	19	19



**Figure 3.1.** The trend of ratio for president-elect by date.

대통령선거여론조사에서 3가지 질문들에 대한 기초적인 통계분석결과는 Table 3.2에 있다. 문제인 당선자가 실제 얻은 득표율은 41.08%였고, 후보지지율과 정당지지율의 평균은 각각 44.20%와 44.88%였다. 그러나 당선가능성은 평균 70.01%로 매우 높았다. 비록 우리나라는 과거의 자료들이 축적되어 있지 않아 미국과 같이 긴 기간 동안의 자료 분석을 실시하지 못하였지만, 한 번의 대통령선거자료를 분석한 결과는 Lewis-Beck과 Skalaban (1989)의 69%와 Lewis-Beck과 Tien (1999)의 71%와 거의 같은 수준이다.

또한 대통령 선거일에 근접한 날짜에 따른 3가지 질문에 대한 비율은 Figure 3.1에 제시되어 있다. 여기서 같은 날짜에 여러 개의 조사가 실시된 경우는 표본크기에 따른 가중평균값을 사용하였다. Figure 3.1에서 후보지지율과 정당지지율은 거의 같은 수준을 유지하고 선거일에 근접하여도 큰 차이가 없다. 반면에 당선가능성은 후보지지율이나 정당지지율보다 훨씬 높은 수준이고, 선거일에 가까울수록 높아지고 있다. 물론 이러한 패턴이 다른 선거에서 유지되는 지는 더 많은 조사 자료들이 축적되어야 가능할 것이다.

제19대 대통령선거여론조사자료를 이용해 당선자의 득표율을 예측하기 위해서 회귀분석을 실시하였다. 독립변수는 당선가능성이고 종속변수는 대통령 당선자의 득표율이다. 추정된 득표율은 다음과 같다.

$$V = 22.49* + 0.31*E \tag{3.1}$$

$$R^2 = 0.32, \quad SEE = 4.04, \quad N = 19.$$

$E$ 는 문제인 후보가 당선될 것이라고 생각하는 응답자들의 비율(%)이고, \*는 유의수준 5%에서 통계적으로 유의함을 나타내며 SEE는 추정치의 표준오차이다. 그리고  $R^2$ 과  $N$ 은 결정계수와 사례수이다.

**Table 3.3.** Correlation in the 7th regional local government election poll

	당선가능성	후보지지율	정당지지율
당선가능성	1		
후보지지율	0.5232	1	
정당지지율	0.4145	0.8833	1

식 (2.2)와 식 (2.3)에서는 최소득표율이 40% 수준이었으나, 우리나라의 제19대 대통령선거자료만을 사용해서 얻은 당선자의 최소득표율은 22% 수준에 불과하다. 모형의 적합성도 상대적으로 낮은 편이다. 이는 우리나라의 경우 사용된 자료가 매우 제한적이고, 19대 대통령선거의 특수성에 기인한 것으로 볼 수 있다. 향후 자료들이 축적되면 보다 정확한 예측이 가능할 것으로 판단된다. 한편 정당지지율을 이용한 당선자의 추정 득표율은 식 (3.2)와 같다.

$$V = 10.11* + 0.76*P \quad (3.2)$$

$$R^2 = 0.76, \quad SEE = 2.38, \quad N = 19.$$

$P$ 는 문재인 후보가 소속된 정당에 대한 지지율이고, \*는 유의수준 5%에서 통계적으로 유의함을 나타내며 SEE는 추정치의 표준오차이다. 그리고  $R^2$ 과  $N$ 은 각각 결정계수와 사례수이다.

정당지지율을 이용한 회귀식이 당선가능성을 이용한 회귀식보다 모형의 적합성이 우수하다. 이는 외국의 사례연구에서도 다루지 않은 특성이다. 물론 정당지지에 대한 질문은 당선가능성 질문이나 후보지지 질문보다 무응답이 매우 높게 나타나므로 이들을 어떻게 처리하는가에 따라 결과에 차이가 있을 수 있다.

### 3.2. 제7회 전국동시지방선거- 광역단체장선거

제7회 전국동시지방선거여론조사결과 중 후보자등록(2018년 5월 25일) 이후 실시되고 중앙선거여론조사심의위원회 홈페이지에 등록된 총 571건의 선거여론조사자료 중에서 광역단체장선거에 대한 120건의 자료를 대상으로 분석하였다. 교육감선거 104건, 기초단체장선거 335건, 그리고 정당지지도조사 12건에는 당선가능성에 대한 질문이 거의 전무하여 분석에서 제외하였다. 17개 광역단체장선거에서도 광주광역시, 대전광역시, 세종특별시, 강원도, 충청북도, 전라북도, 전라남도 등 7개 광역단체장선거여론조사에는 당선가능성을 묻는 질문이 포함되지 않았다. 따라서 10개 광역단체장선거여론조사자료를 사용하고 이 중에서도 3개의 질문이 동시에 포함된 37개 조사자료만을 분석대상으로 하였다. 대통령선거는 전국을 대상으로 하는 반면에 광역단체장선거는 지역별 후보를 대상으로 하기 때문에 지역별 분석 자료가 적어 대통령선거보다 분석에 제한이 있다.

광역단체장선거여론조사에서 3가지 질문의 상관관계는 Table 3.3과 같다. 제주특별시장후보인 원희룡은 자유한국당을 탈당하고 무소속으로 출마하여 당선되었기 때문에 당선자의 소속 정당에 대한 지지도를 반영할 수 없어서 제외하였다. 제7회 전국동시지방선거의 광역단체장선거여론조사에서 3가지 질문에 대한 상관관계는 제19대 대통령선거여론조사의 경우와 비슷하다.

광역단체장선거여론조사에서 3가지 질문들에 대한 기초적인 통계분석결과는 Table 3.4에 있다. 분석에 사용된 10개 광역단체 당선자들의 실제 득표율 평균은 54.79%이다. 한편 득표율 추정치인 후보지지율의 평균은 56.84%이고 정당에 대한 평균지지율은 56.22%로 차이가 없다(제주특별시의 원희룡후보는 무소속이었기 때문에 제외하였음). 반면에 당선가능성 평균은 65.96%로 후보지지율이나 정당지지율보다 훨씬 높았다.

**Table 3.4.** The 7th regional local government election poll

	당선가능성	후보지지율	정당지지율
평균	65.96	56.84	56.22
표준편차	7.39	5.83	8.63
최소값	48.73	47.48	36.65
최대값	82.46	69.39	69.71
관측수	37	37	30

대통령선거에서와 마찬가지로 제7회 광역단체장 선거자료를 이용해서 당선자의 득표율을 예측하기 위해서 회귀분석을 실시하였다. 독립변수는 당선가능성이고 종속변수는 광역단체 당선자의 득표율이다. 추정된 득표율은 식 (3.3)과 같다.

$$V = 26.62^* + 0.49^* E \quad (3.3)$$

$$R^2 = 0.34, \quad SEE = 4.81, \quad N = 37.$$

$E$ 는 후보자가 당선될 것이라고 생각하는 응답자들의 비율(%)이고 \*는 유의수준 1%에서 통계적으로 유의함을 나타내며 SEE는 추정치의 표준오차이다. 그리고  $R^2$ 과  $N$ 은 각각 결정계수와 사례수이다.

한편 정당지지율을 이용한 광역단체 당선자의 득표율 추정식은 식 (3.4)와 같다.

$$V = 24.73^* + 0.59^* P \quad (3.4)$$

$$R^2 = 0.78, \quad SEE = 2.76, \quad N = 30$$

$P$ 는 후보자가 소속된 정당에 대한 지지율이고, \*는 유의수준 1%에서 통계적으로 유의함을 나타내며 SEE는 추정치의 표준오차이다.

대통령선거자료의 경우와 유사하게 정당지지율을 이용한 회귀식이 당선가능성을 이용한 회귀식보다 모형의 적합성이 우수하다.

### 3.3. 토의

선거에서 양자대결의 경우는 득표율 50%, 다수후보의 경우는 득표율 40%를 당선권으로 본다 (Ryu, 2018). 제19대 대통령선거는 다수후보로 치루어졌기 때문에 득표율이 40% 이상이 되어야 당선된다. 반면에 제7회 전국동시지방선거에서 광역단체장 선거도 다수 후보였지만, 실제로는 양자대결의 성격이었다. 따라서 득표율이 50% 이상이어야 당선된다 (실제 당선한 후보들의 득표율은 모두 50% 이상이다).

당선자를 맞힌다는 것은 당선가능성이 50% 이상으로 조사된 경우와 후보지지율에서는 양자대결 50%, 다자대결 40% 이상을 얻는 경우이다. 반면에 후보지지율 조사와 당선가능성 조사의 정확성은 당선자의 실제 득표율이 후보지지율의 오차범위 내에 있으면 후보지지율 조사의 예측은 성공한 것이고, 당선가능성은 당선가능성 50% 이상을 얻는 것을 대립가설로 두고 대립가설이 채택된 경우를 예측이 성공한 것으로 하여 정확성 척도로 사용하였다. 모든 조사의 신뢰수준이 95%이므로, 당선가능성 검정(단측검정)은 유의수준을 5%로 하였다.

Table 3.5에 의하면, 제19대 대통령선거여론조사에서 당선가능성과 후보지지율 조사의 당선자 예측성공률은 같았다. 반면에 제7회 광역단체장선거에서는 당선가능성이 후보지지율보다 당선자에 대한 예측성공률이 높았다.

**Table 3.5.** Predicted success rate for winner

	제19대 대통령선거			제7회 광역단체장선거		
	N	성공	%	N	성공	%
후보지지율	19	18	94.7	37	32	86.5
당선가능성	19	18	94.7	37	36	97.3

**Table 3.6.** Accuracy of vote intention and vote expectation

	제19대 대통령선거			제7회 광역단체장선거		
	N	성공	%	N	성공	%
후보지지율	19	4	21.1	37	20	54.1
당선가능성	19	18	94.7	37	33	89.2

**Table 3.7.** Mean absolute error of 3 methods

	제19대 대통령선거			제7회 광역단체장선거		
	당선가능성	후보지지율	정당지지율	당선가능성	후보지지율	정당지지율
평균절대오차	3.6965	4.6691	4.4305	5.0596	4.3860	4.4230
N	19	19	19	37	37	30

한편 당선가능성과 후보지지율에 대한 정확성은 상당한 차이가 있다. Table 3.6에 의하면 제19대 대통령선거여론조사에서 당선가능성의 정확성은 94.7%인 반면에 후보지지율은 21.1%에 불과하였다. 당선자 득표율에 대한 추정값인 후보지지율이 제시된 오차의 범위를 벗어나는 경우가 19개 조사에서 15개나 되었다. 제7회 광역단체장선거의 경우도 당선가능성 조사가 후보지지율 조사보다 훨씬 정확하였다. 이러한 결과는 기존에 사용된 후보지지율을 이용한 선거예측의 한계라 볼 수 있으며, 당선가능성을 이용한 예측을 새로운 대안으로 생각할 수 있다.

Table 3.7은 당선자에 대한 득표율 추정치와 실제 득표율과의 평균절대오차(mean absolute error)를 나타내고 있다. 당선가능성은 식 (3.1)과 식 (3.3)을 그리고 정당지지율은 식 (3.2)와 식 (3.4)를 이용하여 당선자 득표율을 회귀추정 한 결과에 대한 평균절대오차이다. 대통령선거는 전국에 걸쳐서 후보가 동일한 반면에 광역단체장선거는 지역별로 후보가 다르고 지역별 정당에 대한 지지가 다르기 때문에 지역 특성이 크게 반영된다. 따라서 대통령 당선자에 대한 득표율 추정은 당선가능성을 이용하여 득표율을 추정할 경우 평균절대오차가 가장 작고, 정당지지율을 이용한 경우도 후보지지율을 이용한 경우보다 평균절대오차가 작다. 반면에 광역단체장선거의 경우는 광역단체별로 조사 사례가 매우 적어 전체 자료를 통합하여 하나의 추정식을 유도하고 이를 사용하여 광역단체장후보의 득표율을 추정하였기 때문에 평균절대오차가 크게 된다. 하지만 광역단체별로 조사 사례가 많아져서 개별 추정식을 사용하게 되면 지역의 특성이 잘 반영되어 오히려 대통령선거의 경우보다 오차가 작아질 것으로 판단된다. 향후 조사 자료들이 축적되면 이러한 현상들에 대한 다양하고 심도 있는 연구가 가능하게 된다.

#### 4. 결론

본 연구에서는 기존의 선거여론조사에서 사용된 “오늘이 선거일이라면 당신은 누구에게 투표할 것입니까?”라는 투표의향 질문과 “이번 선거에서 당신은 누가 당선될 것이라고 생각합니까?”라는 당선가능성을 묻는 질문, 그리고 정당지지율도 함께 고려하여 선거여론조사결과를 비교 분석하였다.

당선가능성이 후보지지율보다 당선자를 더 정확하게 예측하고 예측의 정확성면에서도 우수함을 외국의 사례에서 볼 수 있다. 그러나 우리나라의 경우는 당선가능성을 묻는 질문이 상대적으로 적고 이를 이용



한 예측 사례도 없기 때문에, 본 연구에서는 제19대 대통령선거여론조사와 제7회 전국동시지방선거(광역단체장선거) 여론조사를 이용하였다. 당선가능성과 후보지지율, 그리고 정당지지율과의 상관관계를 보면 당선가능성과 후보지지율의 상관은 대통령선거에서는 0.53이고 광역단체장 선거에서는 0.52로 거의 같은 수준이다. 그러나 후보지지율과 정당지지율의 상관은 대통령선거는 0.87이고 광역단체장선거는 0.88로 매우 높다. 19개의 대통령선거여론조사결과 당선자의 당선가능성은 평균 70.01%인 반면에 후보지지율 평균은 44.20%이고, 37개 광역단체장선거여론조사결과에서 당선자의 당선가능성 평균은 65.96%, 후보지지율 평균은 56.84%로 당선가능성 조사가 후보지지율 조사보다 당선자 예측이 정확하였다. 당선자를 예측하는데 있어서는 당선가능성이 대통령선거에서는 동일하였는데, 광역단체장 선거의 경우는 월등히 높았다. 또한 제19대 대통령선거와 제7회 광역단체장선거에서 당선가능성 조사가 후보지지율 조사보다 훨씬 정확하였다.

당선가능성과 정당지지율을 독립변수로 해서 당선자의 득표율을 예측하는 회귀식을 유도하고 당선자의 득표율을 추정하여 평균절대오차를 구하였다. 제19대 대통령선거에서는 당선가능성을 이용한 경우의 평균절대오차가 가장 작았다. 반면에 제7회 광역단체장선거에서 후보지지율 조사의 평균절대오차가 작았는데, 이는 광역단체장선거 조사의 사례가 너무 적고, 지역적 특성이 큼에도 불구하고 조사 자료를 하나로 통합한 회귀식을 사용한 것이 원인으로 판단된다.

본 연구에 사용된 자료가 매우 제한되고 또한 모든 자료들이 개인별 응답 자료가 아니고 조사별로 집계된 자료이기 때문에 분석의 범위가 제한되었다. 향후 선거여론조사에 당선가능성 문항을 포함한 조사가 많아지고 또한 응답자별 자료들이 축적되면 다양하고 심도있는 연구가 가능할 것이다. 장기적으로 당선가능성과 정당지지율을 사용하면 선거예측을 보다 정확하게 할 수 있다. 한편 Graefe (2014, 2015)와 Graefe 등 (2014)들은 여러 조사들을 결합하여 예측하는 방법을 제시하고 있는데, 여기서 다른 당선가능성도 다양한 조사 방법들과 결합하여 예측에 사용하는 것도 바람직하다. 또한 최근의 조사에서 응답자들이 응답을 회피하거나 꺼려하고 심지어 솔직하게 응답하지 않는 경우도 발생하므로 확률화응답기법의 적용도 고려할 필요가 있다 (Lehrer 등, 2019).

## References

- Campbell, J. (1996). Polls and votes - The trial-heat presidential election forecasting model, certainty, and political campaigns, *American Politics Research*, **24**, 408-433.
- Hayes, S. P. (1936). The predictive ability of voters, *Journal of Social Psychology*, **45**, 183-191.
- Graefe, A. (2014). Accuracy of vote expectation surveys in forecasting elections, *Public Opinion Quarterly*, **73**, Special Issue, 204-232.
- Graefe, A. (2015). Accuracy gains of adding vote expectation surveys to a combined forecast of US presidential election outcomes, *Research and Politics*, January-March, 1-5.
- Graefe, A. (2016). Forecasting proportional representation elections from non-representative expectation surveys, *Electoral Studies*, **42**, 222-228.
- Graefe, A., Armstrong, J. S., Jones, Jr. R. J., and Cuzan, A. (2014). Combining forecasts: An application to elections, *International Journal of Forecasting*, **30**, 43-54.
- Lehrer, R., Juhl, S., and Gschwend, T. (2019). The wisdom of crowds design for sensitive survey questions, *Electoral Studies*, **57**, 99-109.
- Lewis-Beck, M. S. and Skalaban, A. (1989). Citizen forecasting: Can voters see into the future?, *British Journal of Political Science*, **19**, 146-153.
- Lewis-Beck, M. S. and Stegmaier, M. (2011). Citizen forecasting: Can UK voters see the future?, *Electoral Studies*, **30**, 264-268.
- Lewis-Beck, M. S. and Tien, C. (1999). Voters as forecasters: A micromodel of election prediction, *International Journal of Forecasting*, **15**, 175-184.

- Miller, M. K., Wang, G., Kulkarni, S. R., Poor, H. V., and Osherson, D. N.(2012). Citizen forecasts of the 2008 U.S. presidential election?, *Politics & Policy*, **40**, 1019–1052.
- Murr, A. E. (2011). “Wisdom of crowds?” A decentralized election forecasting model that uses citizens’ local expectations, *Electoral Studies*, **30**, 771–783.
- Murr, A. E. (2015). The wisdom of crowds: Applying Condorcet’s jury theorem to forecasting US presidential elections, *International Journal of Forecasting*, **31**, 916–929.
- Rothschild, D. and Wolfers, J. (2013). Forecasting elections: Voter intentions versus expectations (working paper). Available from: <http://users.nber.org/~jwolfers/Papers/VoterExpectations.pdf>
- Ryu, J. B. (2018). The accuracy of election forecasts and the chance of winning for multiple candidates: In case of the 2017 presidential election, *Survey Research*, **19**, 1–20.
- Ryu, J. B. (2019). Alternative methods for election forecast: For the 2017 presidential election, *Journal of Research Institute of Industrial Science*, **36**, 1–8.
- National Election Survey Deliberation Commission; <http://www.nesde.go.kr/>
- Wikipedia; [https://en.m.wikipedia.org/wiki/United\\_States\\_presidential\\_election#Popular\\_vote\\_results](https://en.m.wikipedia.org/wiki/United_States_presidential_election#Popular_vote_results)

# 선거예측의 대안적 방법에 대한 사례연구

류제복<sup>a,1</sup>

<sup>a</sup>청주대학교 통계학과

(2020년 5월 12일 접수, 2020년 5월 22일 수정, 2020년 5월 22일 채택)

---

## 요약

본 연구에서는 제19대 대통령선거와 제7회 전국동시지방선거(광역단체장선거) 여론조사자료를 이용해서 후보지지율, 당선가능성 및 정당지지율의 관계를 비교, 분석하였다. 사례연구결과 당선가능성과 정당지지율을 이용해서 당선자를 예측하는 대안적 방법이 선거예측의 정확성을 높일 수 있을 것으로 기대된다.

주요용어: 당선가능성, 후보지지율, 정당지지율, 평균절대오차

---

이 논문은 2019학년도에 청주대학교 산업과학연구소가 지원한 학술연구조성비(특별연구과제)에 의해 연구되었음.

<sup>1</sup>(28053) 충북 청주시 청원구 대성로 298, 청주대학교 통계학과. E-mail: jbryu@cju.ac.kr