일반논문

읍·면·동 멀티레벨 데이터를 이용한 정당우세율 분석모형

A Predictive Model of Electoral Outcomes Using Town-level Data

최필선[a] · 민인식[b]
Pilsun Choi · Insik Min

본 연구에서는 최근 치러진 세 번의 선거(2012년 총선, 2012년 대선, 2014년 지방선거)에서 나타난 정당지지 결과를 활용하여 읍·면·동 단위의 멀티레벨 데이터를 구축한 다음, 이를 기반으로 정당우세율을 분석하는 모형을 개발한다. 읍·면·동 단위의 모형이라는 점에서 대통령선거, 국회의원선거 및 지방선거 등 다양한 선거에 적용할 수 있는 범용성을 지닌 모형이라고 할 수 있다. 또한 선거 분석 및 예측모형의 경우 득표율에 영향을 미치는 모든 변수를 포함시키기 어렵다는 점을 고려하여 멀티레벨 데이터를 이용한 회귀모형을 구축함으로써 관찰되지 않은 읍·면·동의 이질성을 포착하고자 하였다. 본 연구모형에서는 읍·면·동의 인구구조와 지역 변수를 주요 설명변수로 사용했을 뿐만 아니라, 정당지지 분석에서는 경제정책 변수가 매우 중요하다는 점을 고려하여 지역별 주택가격 변수를 모형에 포함시킨 점도 선행연구와 차별성을 지닌다. 이처럼 읍·면·동 단위의 멀티레벨 데이터를 이용한 정당우세율 분석모형을 추정하여 어떤 설명변수들이 득표율에 영향을 미치는지 파악할 수 있으며, 이를 통해 선거예측에 도움을 얻을 수 있다. 추정결과를 가까운 미래에 치러질 전국 단위 선거에 적용하여 정당우세율 분석 시뮬레이션을 예시적으로 수행한다.

a) 건국대학교 상경대학 국제무역학과 교수.

b) 교신저자(corresponding author): 경희대학교 정경대학 경제학과 교수 민인식.
E-mail: imin@khu.ac.kr

주제어: 우세율, 읍·면·동, 멀티레벨 데이터, 주택가격

This study proposes a predictive model of electoral outcomes, using major parties' vote revealed in last three nationwide elections (2012 parliamentary election, 2012 presidential election, and 2014 local election). In particular, we construct town-level election outcomes in the form of a two-level data structure. Because our model generates election outcomes at the town level, it can be applied to any types of elections, and in this respect, this modeling approach can yield a methodological contribution. By capturing a town-level heterogeneity, our multi-level regression model can be a complement to traditional prediction models. The economic policy variables as well as demographic and region-specific variables are explicitly considered in our model. Specifically housing sales and rental prices are used as proxy variables for the performance of economic policy in the election year. We apply the model to "Jongno-gu" parliamentary district as an exemplary simulation.

Key words: predictive model, town-level, multi-level data, housing price

Ⅰ. 서론

선거는 민주주의에서 유권자의 의사를 표현하는 가장 중요한 수단이다. 유권자가 어떤 후보, 어떤 정당을 선택하느냐는 선거와 관련된 연구에서 가장 주목받는 연구이다. 선거는 그 결과에 영향을 미치는 변수가 다양하고 선거의 종류마다 변하기 때문에 분석 및 예측이 쉽지 않으며, 일관된 접근방법을 유지하

기도 쉽지 않다. 일반적으로 선거예측의 기준으로 사전성, 정확성, 경제성이 거론되지만(김재한 1995c) 본 연구에서는 여기에 추가적으로 '범용성'을 포함시키고자 한다. 범용성은 간단히 말해 가급적 다양한 유형 및 시점의 선거에 대해 광범위하게 적용될 수 있어야 한다는 것이다. 선거예측모형이 가령 대통령, 국회의원, 지방선거에 모두 활용될 수 있다면 범용성 기준을 충족하게 된다.

선거예측은 크게 여론조사를 이용한 예측과 모형에 기초한 예측으로 구분할 수 있다. 여론조사는 모형이라기보다는 단순히 조사를 통한 결과이기 때문에 선거에 출마할 후보들이 결정되고 난 뒤 선거가 가까워진 시점에서 이루어져야 하는 제약이 있다. 즉 사전성 측면에서는 단점이 있다. 대통령선거의 경우에는 대규모의 표본을 이용하기 때문에 예측결과가 잘 맞지만 상대적으로 소규모 지역의 선거에는 표본오차(sampling error)가 크다. 또한 여론조사는 후보자가 결정된 상태에서 직접 유권자를 조사하여 발표하기 때문에 정확성 측면에서는 우수하다고 볼 수 있지만, 경제성과 사전성, 그리고 범용성 측면에서는 단점이 있다.

반면, 과거 선거결과 데이터에 기초하여 모형을 개발하고 그에 따른 예측결과를 제시하는 모델링 접근방법은 사전성과 경제성 측면에서 장점이 존재한다. 사전성 측면을 보면, 해당 선거구에 출마후보가 결정되지 않은 상황일지라도 주요 정당의 득표율 또는 상대적 우세율을 예측하는 모형을 적용할 수 있다. 경제성 측면에서는 과거 선거결과 데이터와 통계청 등에서 얻을 수 있는 설명변수 값을 사용하기 때문에 거의 비용이 들지 않는다. 그러나 유권자를 직접 조사하지 않고 과거 선거결과에 의존하기 때문에 정확성은 여론조사 접근법에 비해 낮은 편이다. 또한 한국의 선거예측모형은 특정 선거유형에 초점을 맞추고 개발되기 때문에 범용성도 역시 떨어진다(김재한 1995c; 김길수 2013 등).

모형에 기초한 선거예측 접근이 여러 장점에도 불구하고 범용성 측면에서 갖는 한계를 극복하기 위해 본 연구는 읍·면·동 단위에서 정당우세율을 분석 및 예측하는 모형을 제안한다. 이 모형은 읍·면·동 단위이기 때문에 대통령, 국회의원, 지방선거에 모두 적용할 수 있는 범용성 기준을 만족시킬 수 있으리라 기대된다.

본 연구에서는 최근 치러진 전국 단위 선거인 18대 대통령선거, 19대 국회의원선거, 그리고 6회 지방선거(광역단체장) 결과를 이용하여 읍·면·동 단위 멀티레벨 데이터를 구축한다. 전국 단위 또는 시·군·구 단위 데이터를 이용하는 것에 비해 읍·면·동 단위 데이터는 추정에 사용할 수 있는 표본 수가 증가되는 장점이 있다.[1] 또한 멀티레벨 데이터를 이용한 회귀모형에서는 관찰되지 않는(unobserved) 읍·면·동의 이질성(heterogeneity)을 고려할 수 있다. 즉, 선거예측모형에서는 득표율에 영향을 미치는 모든 변수를 포함시키기 어렵지만, 본 연구는 읍·면·동(상위 레벨)-선거(하위 레벨)의 2단계 멀티레벨 데이터를 구축함으로써, 해당 읍·면·동에서 특정 정당의 득표율에 영향을 미칠 수 있으면서도 관찰될 수가 없어서 설명변수로 포함되지 않은 요인들을 감안할 수 있게 된다.

한국의 선거결과는 유권자 연령(세대)과 지역 변수를 이용해 상당부분 예측가능하다고 알려져 있다(김재한 1993). 본 연구의 정당우세율 분석모형에서도 읍·면·동의 인구구조와 지역 변수를 주요 설명변수로 사용한다. 뿐만 아니라 선거 시점의 경제정책(환경) 변수를 모형에 포함시켰다는 점에서도 기존 연구와 차별화 된다. 경제정책 변수로는 주택 매매가격과 전세가격을 고려하였다. 주택 매매가격은 주택을 보유한 유권자의 자산가격 상승으로 간주될 수 있기 때문에 현 집권당의 지지율을 높일 수 있는 반면, 전세가격 상승은 주거비용의 상승으로 연결되고 따라서 집권당의 주택정책 실패로 간주됨으로써 집권당 후보의 지지율 하락으로 이어질 수 있기 때문이다. 정당 간 우세율 분석에서는 경제정책 변수가 매우 중요하다고 여겨짐에도 불구하고 지역 단위에서 적절한 경제정책 변수를 발굴하는 것이 쉽지 않은 상황에서 지역별 주택가격 변수를 모형에 포함시켰다는 점에서 의의가 있다.

본 연구에서는 최근 치러진 세 번의 선거(2012년 총선, 2012년 대선, 2014년 지방선거)에서 행정동 단위의 정당지지 결과를 활용하여 우세율 예측모형을 구축한다. 모형의 우수성을 판단하는 기준인 경제성, 사전성, 정확성뿐만 아니라 범용성까지 고려한 모형을 개발하기 위한 것이다. 읍·면·동 단위의 멀티레

1) 시·군·구 단위는 251개이고, 읍·면·동 단위는 대략 3,500개이다.

벨 데이터를 이용한 모형 추정을 통해 어떤 설명변수들이 득표율에 영향을 미치는지 파악할 수 있으며, 이를 통해 후보자들은 선거대책을 마련할 수 있다. 이러한 접근방법은 기존의 선거예측모형에서 고려되지 않은 방법론적 기여라고 볼 수 있으며 추정결과에 기초한 시뮬레이션을 가까운 미래에 치러질 전국단위 선거에 예시적으로 적용해 본다.

본 논문의 구성은 다음과 같다. 먼저 Ⅱ장에서는 모형에 기초한 선거예측 연구가 우리나라와 미국에서 어떻게 진행되어 왔는지 선행연구를 정리하고, Ⅲ장에서는 본 논문의 실증분석에 활용된 데이터와 주요 변수에 대해 논의한다. Ⅳ장에서는 멀티레벨 모형 추정을 통해 득표율에 영향을 주는 변수에 대해 확인하고 추정결과를 활용한 시뮬레이션 방법과 그 결과를 제시한다. 마지막으로 Ⅴ장에서는 실증분석 결과를 요약하고, 그 결과가 가진 이론적·학술적 함의를 논한다.

Ⅱ. 선행연구 검토

서론에서 설명하였듯이 본 연구의 목적은 읍·면·동 수준의 데이터를 이용하여 정당우세율 분석모형을 개발하는 것이다. 따라서 이번 장에서는 모형에 근거한 선거예측과 관련된 선행연구를 중점적으로 살펴보기로 한다.

김재한(1995a)은 투표일에 임박한 여론조사는 선거결과를 정확히 예측할 수 있을지 몰라도 선거에 대한 대비 또는 대책에 아무런 도움을 주지 못한다고 설명한다. 투표선택에 대한 직접적 조사는 그 결과를 액면 그대로 믿기 어려운 문제가 있다. 가령 투표율 조사를 살펴보면 대부분의 조사에서 예상 투표율이 실제 투표율을 상회한다(김재한 1995b). 미국의 경우에는 투표율이 50% 정도이기 때문에 단순한 인기도 조사를 선거결과 예측에 그대로 반영하는 것에 신중을 기한다. 투표선택에 대한 직접적 조사는 투표자의 전략적 투표성향을 반영하지 못하기 때문에 선거예측에 도움이 되지 못하기도 한다. 따라서 여론조사 결과를 그대로 받아들이기보다는 이 결과를 해석하고 유추하는 과정이 필요하다. 즉, 유권자의 성향까지 고려해서 예측할 필요가 있다. 김재한(1993)

은 투표선택의 간접적 유추를 위해 유권자 투표행태에 대한 회귀모형을 설정하였다. 그 결과, 가령 14대 대통령선거의 경우 후보선택은 지역과 연령만으로 3/4 정도가 설명되었다.

미국식 선거예측모형 역시 사전성에서는 좋지 않은 평가를 받고 있다. Greene (1993)은 기존의 선거예측모형 모두가 탁상공론식 전망에 비해 더 나은 예측결과를 보여주지 못 한다고 설명한다. 그렇지만 여전히 선거예측모형의 유용성은 논의될 필요가 있다고 주장한다. 과거 선거에 대한 적중률을 높이기 위해 모형과 데이터를 일치하게끔 모형을 수정하다 보면 미래 선거에 대한 적중률은 떨어지는 경향이 있다. 과거 선거에 대한 설명력이 증가되었을지라도 다가올 선거를 전망하는 데 도움이 되지 않을 수 있다. 따라서 과거 사실에 귀납적으로 모형을 맞추다 보면 미래예측에 도움이 되지 않는 경우가 있다.

우리나라의 선거예측이 미국에 비해 상대적으로 우수한 편이기는 하지만 여전히 오차가 발생한다. 송근원(2011)은 2010년 6월 2일 실시된 제5회 지방선거 자료를 이용하여, 광역단체장 후보자의 득표율을 종속변수로 삼고 득표율에 영향을 미치는 설명변수를 둔 회귀모형을 설정하였다. 후보의 가시성(인지도)이 득표율과 매우 유의한 관계가 있다는 점을 고려하였고, 현직자 효과,[2] 지역효과 및 견제효과를 이용하여 득표율 예측모형을 설정하였다. 하지만 송근원의 예측모형은 지방선거에만 적용될 수 있고 대통령선거나 국회의원선거에는 적용될 수 없는 한계가 있다.

우리나라 선거예측모형에서는 지역주의 영향을 무시할 수 없기 때문에 선거에 미치는 지역지배 정당효과를 고려한다. 이러한 지역효과는 영·호남 간에 아주 강하게 나타나며 여전히 지역지배 정당효과는 거의 모든 선거에서 찾아볼 수 있다(조진만 외 2006; 송근원·정봉성 2007). 정책 변수 역시 선거예측모형에서 중요하게 고려되어야 한다. 유권자는 자신의 정책방향과 일치하는 후보를 선택할 가능성이 크기 때문이다. 선거 기간 중이나 바로 직전에 주요 정책 입장에 대한 유권자들의 찬반 의견이 조사된다면 그러한 의견의 찬반비율을 정책변수로 포함할 수 있다. 송근원(2011)에서는 대북정책에 대한 찬반비율

2) 우리나라의 선행연구에서는 대부분 현직자 효과를 인정하고 있다(김석우 2006; 한정택 2007; 박명호·김민선 2008).

을 정책변수로 포함한 모형을 설정하였다. 다만 이러한 정책변수는 미래 선거 예측에서는 활용할 수 없다는 단점이 있다.

김길수(2013)의 연구에서는 기존의 득표율 예측모형들이 설명변수를 일률적으로 결정하는 것과 달리 탐색적(heuristic) 방법을 사용하여 확립된 알고리즘을 사용하는 대신 다양한 변수를 탐색하고 적용한다. 선거의 종류와 지역에 따라 고려해야 할 변수들이 달라지며, 한정된 정보 내에서 최적해가 아닌 만족할 만한 선거결과를 예측하고자 한다. "서울특별시 종로구"만을 연구대상으로 선정하여 해당 지역구의 특성을 파악하고 선거에 영향을 미치는 변수를 발견한다. 그러나 많은 지역구와 다양한 선거에서 활용하기에는 한계가 있는 접근방법이다.

신계균(2012)은 1970년대부터 2000년대까지 미국 대통령 선거예측 방법론을 정리하였다. 미국은 1970년대 후반부터 본격적인 대선 예측에 관한 연구가 시작되었으며 이 시기에는 대통령의 지지율과 선거결과에 초점을 맞추었다. Sigelman(1979)은 투표 전 시행된 갤럽조사의 대통령 지지율과 실제 투표에서 현직 대통령 후보가 얻은 득표율의 회귀분석을 시행하였다. Lewis-Beck & Rice (1982)는 지지율 변수가 어떠한 자료를 통해서 얻어지는지가 중요하다고 강조하였다. 그들은 예비선거 이후와 전당대회 이전에 이루어진 갤럽 설문조사를 이용하여 대선예측을 시도하였다.

대통령 지지율뿐 아니라 유권자들이 후보자를 평가하는 근거를 포함하는 것이 예측모형의 정확도를 높이는 방법이라고 하는 연구들이 있다. 특히 경제적 수행능력 평가는 예측모형을 향상시키는 가장 중요한 변수라고 주장한다. 경제적 호황기에 대중들은 현직 대통령에게 더 높은 지지를 보일 것이고, 따라서 당선가능성이 높아진다. 이에 따라 현직 대통령의 경제적 수행능력을 파악하는 것은 상당히 중요한 요소라 할 수 있다(Krammer 1971; Lewis-Beck & Rice 1984; Tufte 1978). 미국의 경우 경제적 수행능력은 주로 거시경제 지표에 의존한다. 즉 국가경제 평가가 중요한 변수이다. Fair(1978)은 경제변수로 실질국민총생산, 국민총생산 디플레이터, 그리고 실업률을 예측모형에 포함하였다.

미국의 연구에서는 예측모형의 분석단위를 국가 전체가 아니라 주 단위로 설정하기도 한다(Campbell 1992; Holbrook 1991). Rosenstone(1983)은 1948년

부터 1972년까지 시행된 주 단위 대통령 선거 모형에서 사회복지 그리고 인종 이슈와 같은 다양한 주 단위 변수를 포함하였다. Holbrook(1991)은 각 주의 경제지표, 이념변수, 국가 경제지표와 현직 대통령의 지지율 등 주 단위 변수와 국가 단위 변수를 포함한 모형을 개발하였다.

2000년대 미국 대통령선거 예측연구에서는 기존 예측모형을 보완하는 차원의 연구가 진행되었다. 대부분의 연구들은 기존의 대통령 지지율과 경제변수를 포함시키고 갤럽의 선호조사를 주요 지표로 사용하였다. 다만 Norpoth(2004)는 자기회귀모형을 이용한 예측모형을 설정하고 대통령 후보경선 결과를 포함하였다는 점에서 기존 연구와 차이가 있다. 자기회귀모형은 앞선 두 대통령 선거의 득표율을 설명변수로 사용하고 있다.

미국에서의 다양한 연구와 달리 우리나라에서는 선거예측에 대한 연구는 주로 여론조사에 근거한 예측이 이루어지고 있다. 본 연구에서는 사전성과 경제성 측면에서 장점이 있는 모델링에 기초한 선거예측을 제시한다. 읍·면·동 단위에서 정당우세율 분석모형을 개발함으로써 대통령선거, 국회의원선거 및 지방선거 등 다양한 선거에 적용할 수 있는 모형이라는 점에서 방법론적 기여가 있다. 또한 선거에 영향을 미치는 경제환경 변수로서 주택가격을 포함하였다는 점에서 기존연구와 차별화된다.

Ⅲ. 연구자료 및 변수 설명

1. 연구 데이터 구축

본 연구에서는 최근 치러진 세 차례 전국 단위 선거결과 데이터를 이용한다. 2012년 4월 19대 국회의원 선거(이하 국선), 2012년 12월 18대 대통령선거(이하 대선), 그리고 2014년 6월 6회 지방선거(이하 지선)가 그 대상이다. 지방선거에서는 광역단체장 선거결과 데이터만 활용한다. 세 번의 전국 단위 개표결과를 읍·면·동 멀티레벨 데이터로 만들기 위해서는 먼저 읍·면·동 단위에서 각 후보자의 득표수를 알아야 한다. 이에 대한 엑셀자료는 중앙선거관리위원

회(www.nec.go.kr) 자료실에서 다운로드 받을 수 있다.

멀티레벨 데이터 구조로 만들기 위해서는 상위 레벨(읍·면·동)-하위 레벨(선거) 변수가 필요하다. 우선 하위 레벨 변수는 각 읍·면·동(i)-선거시점(j)으로 만들 수 있다. 선거시점은 선거유형(국선, 대선, 지선)과 일치한다. 다음으로 상위 레벨 변수는 각 읍·면·동이다. 그런데 읍·면·동 행정구역은 시간에 따라 그 이름이 변경될 수 있다. 가령 "여주군"에 속한 읍·면·동은 2013년 9월 '여주시' 읍·면·동으로 승격되었다.[3] 또한 인천광역시 서구에 속한 검단1동은 2013년 9월 검단1동과 검단5동으로 분동되었다. 따라서 2012년 국회의원과 대통령 선거결과에서는 검단5동은 존재하지 않았다. 2014년 지선부터 검단5동 선거결과를 얻을 수 있다. 이처럼 2012년~2014년 동안 행정구역 변경의 문제를 고려하여 데이터를 구축했다.

우리의 관심변수는 각 시점(j)의 선거에서 후보자가 특정 읍·면·동(i)에서 얻는 득표수이다. 그러나 매 선거마다 출마하는 후보자가 다르기 때문에 후보자 대신 정당의 득표수를 관심변수로 구축한다.[4] 가령 새누리당 후보가 각 선거에서 '서울특별시 강남구 삼성동'에서 얻는 득표수가 실증분석에서 필요한 변수이다. <표 1>은 저자에 의해 구축된 데이터의 일부분을 보여준다. 맨 상단에 변수 이름이 나와 있다. 'region5' 변수가 상위 레벨에 해당하는 읍·면·동 변수이다. 'year' 변수는 선거가 치러진 연도 변수이다. 'vote새누리당' 변수는 해당 읍·면·동에서 새누리당 후보의 득표수이고, 'vote새정치민주연합' 변수는 새정치연합(이하 새정연) 후보가 얻은 득표수이다.[5] 앞서 언급하였듯이 후보자의 소속정당에 초점을 맞추어 득표수 변수를 생성하였다.

---

3) 멀티레벨 데이터로 병합하기 위해서 "여주시"는 모두 "여주군"으로 변경하였다.

4) 따라서 본 연구모형은 선거 예측이라기보다는 정당우세율 예측모형이라고 할 수 있다.

5) 19대 국회의원선거에서는 새정연 정당이 창당되기 이전이기 때문에 민주통합당 후보를 새정연 후보로 간주한다.

<표 1> 구축된 멀티레벨 데이터의 일부분

region5	election	year	vote 새누리당	vote 새정치민주연합
서울특별시 강남구 도곡2동	18대 대선	2012	15,129	5,202
서울특별시 강남구 도곡2동	19대 국선	2012	11,890	3,252
서울특별시 강남구 도곡2동	6회 지선	2014	10,473	5,125
서울특별시 강남구 삼성1동	18대 대선	2012	6,060	3,155
서울특별시 강남구 삼성1동	19대 국선	2012	4,550	1,982
서울특별시 강남구 삼성1동	6회 지선	2014	4,128	2,651
서울특별시 강남구 삼성2동	18대 대선	2012	10,918	7,206
서울특별시 강남구 삼성2동	19대 국선	2012	7,897	4,624
서울특별시 강남구 삼성2동	6회 지선	2014	6,826	6,055

2. 종속변수

본 연구모형의 종속변수는 각 선거에서 새누리당 후보와 새정연 후보의 득표수 비율(배수)이다.

$$y_{ij} = \frac{\text{새누리당 후보 득표수}_{ij}}{\text{새정연 후보 득표수}_{ij}} \quad (1)$$

어떤 선거에서는 특정 읍·면·동에 새누리당 또는 새정연에서 후보가 출마하지 않는 경우가 있다. 이런 경우에는 결측치(missing value)가 생성되기 때문에 추정 표본에서 제외된다. $y_{ij} > 1$이면 새누리당 후보가 새정연 후보에 비해 더 많이 득표했다는 것이고, $y_{ij} < 1$이면 새정연 후보가 새누리당 후보에 비해 더 많이 득표했다는 의미이다. 즉, y_{ij}는 새누리당 후보의 우세비율(배수)이다.

<표 2>에서는 선거유형별로 y_{ij} 변수의 기초통계량을 정리하여 제시한다. 평균은 가중치를 고려하지 않은 단순 평균이다. 19대 국선에서 2,655개[6] 읍·면

·동의 평균 y_{ij} 값은 1.88이다. 평균적으로 새누리당 후보의 득표수가 새정연 후보의 득표수보다 1.88배 더 많았다는 것을 의미한다. 이러한 경향은 18대 대선과 6회 지선에서도 거의 유사하다. 다음으로 표준편차를 보면, 19대 국선이 제일 크고, 6회 지선, 그리고 18대 대선 순으로 그 값이 작아진다.

<표 2> 새누리당 후보 우세비율(배수) 기초통계량

선거	표본 수	평균	표준편차	최소	최대
19 대 국선	2,655	1.88	2.57	0.02	32.6
18 대 대선	3,478	1.87	1.89	0.06	14.4
6 회 지선	3,202	1.81	2.37	0.04	15.7

3. 주요 설명변수

멀티레벨 선형회귀모형에서 설명변수는 읍·면·동 단위에서 관찰가능한 변수를 사용하고자 한다. 그러나 시간(각 선거시점)에 따라 변하는 읍·면·동 특성 변수를 얻는 것은 데이터 제약으로 인해 현실적으로 매우 어려운 일이다. 읍·면·동 단위에서 설명변수를 얻을 수 없는 경우에는 부득이하게 상위 레벨인 시·군·구 단위에서 관찰가능한 변수를 선택하였다.

1) 50세 이상 인구비율

최근 우리나라 선거의 특징은 세대투표 성향이 두드러진다는 점이다. 65세 이상 고령인구가 급속히 늘어나고 있는 현 시점에서 고령인구의 정치적 성향은 투표에서 매우 중요한 요소이다(최필선·민인식 2015). 최근 드러난 세대투표 성향을 모형에 반영하기 위해 다음과 같이 50세 이상 인구비율(x_{1ij}) 변수를 생성한다.

---

6) 국회의원선거의 표본수가 적은 이유는 새정연은 영남지역에서 후보를 내지 않는 경우가 많기 때문이다.

$$x_{1ij} = \frac{\text{읍·면·동 } i\text{의 } j \text{ 선거에서의 50세 이상 인구}}{\text{읍·면·동 } i\text{의 } j \text{ 선거에서의 전체 인구}} \times 100 \quad (2)$$

각 읍·면·동의 인구는 통계청의 통계정보시스템(www.kosis.kr)에서 구할 수 있다. 그러나 현재 2012년과 2013년 주민등록상 연령별 인구 데이터만 올라와 있다. 2012년에 치러진 국선과 대선은 2012년 인구를 사용할 수 있지만, 2014년 6월에 치러진 지선에서는 2014년 주민등록 인구가 필요하다. 하지만 데이터의 제약으로 2013년 12월 시점의 연령별 인구를 2014년 지선에 적용하기로 한다.

<표 3>에서는 50세 이상 인구비율(x_{1ij}) 변수의 기초통계량을 정리하여 보여준다. 2012년 19대 국선과 18대 대선에는 3,466개 읍·면·동의 50세 이상 인구비율이 평균 40.9%이다. 최소값은 6.6%이고 최대값은 78.4%이다. 2014년 초에는 50세 이상 인구비율이 42.1%로 증가한다. 고령화 비율이 점차 늘어나고 있다고 판단할 수 있다. 최대값과 최소값은 2012년과 거의 유사하다.

〈표 3〉 50세 이상 인구비율 (단위: %)

선거	표본 수	평균	표준편차	최소	최대
19 대 국선 18 대 대선	3,466	40.9	13.8	6.6	78.4
6 회 지선	3,468	42.1	14.1	6.7	78.9

2) 주택가격

정당간 우세율을 분석함에 있어 핵심변수는 선거시점에서의 경제적 상황변수이다. 본 연구에서는 이에 해당하는 변수로 선거시점의 주택가격 변수를 사용한다. 주택가격은 매매인 경우 매매가격, 그리고 임대인 경우에는 전세가격을 각각 계산하여 사용한다. 국토교통부에서 제공하는 주택실거래가 사이트(http://rt.molit.go.kr/)에서는 2016년 1월 현재 주택매매는 2006년 1월부터

2015년 7월까지, 그리고 전세/월세 거래는 2011년 1월부터 2015년 12월까지 모든 거래 정보를 공개하고 있다. 주택유형은 아파트, 연립·다세대, 그리고 단독·다가구로 구분하고 있다. 모든 주택유형을 포함하여 매매와 전세가격을 계산하였다.

주택가격 변수를 읍·면·동 단위에서 평균하여 계산하고자 하였으나 특정 읍·면·동은 1년 동안 주택거래가 전혀 없거나 2~3건의 주택거래만 있는 경우도 있다. 이런 경우에는 해당 읍·면·동의 주택가격 변수를 생성하는 것이 적절하지 않다. 따라서 읍·면·동 단위의 주택가격 대신 해당 읍·면·동이 속한 시·군·구 레벨의 주택가격 변수를 생성하여 사용한다.[7] i 읍·면·동의 j 선거시점에서의 주택 매매가격이 x_{2ij} 변수이고, 전세가격이 x_{3ij} 변수이다. 매매와 전세 모두 $3.3m^2$당 가격이다. 월세의 경우에는 전월세 전환율을 6%로 가정하고 월세가격을 전세가격으로 변환하였다.[8] 세 번의 선거에 매칭하는 평균 주택가격을 계산하기 위해 각 선거시점을 기준으로 직전 6개월 동안의 가격을 평균하였다. 구체적인 설정 기간이 <표 4>에 나와 있다.

<표 4> 평균 주택가격 계산을 위한 기간 설정

선거	주택가격 계산 기간
19 대 국선 (2012. 4 월 선거)	2011 년 10 월~2012 년 4 월
18 대 대선 (2012 년 12 월 선거)	2012 년 7 월~2012 년 12 월
6 회 지선 (2014 년 6 월 선거)	2014 년 1 월~2014 년 6 월

7) 읍·면·동의 상위 레벨인 시·군·구 단위는 모두 251개로 구분된다.

8) 월세 거래를 전세가격으로 전환하는 공식은 [(월세 × 12)/0.06 + 월세보증금]이며, 여기서 0.06은 전월세 전환율이다.

<표 5>에서는 서울시 강남구와 은평구의 선거별 평균 주택 매매가격과 전세가격을 예로 보여준다. 서울시 강남구의 경우, 주택 매매가격은 19대 국선에 비해 18대 대선과 6회 지선에서는 하락한다. 그러나 전세가격은 꾸준히 상승하는 추세이다. 서울시 은평구의 경우, 매매가격은 큰 차이가 없지만 전세가격은 큰 폭으로 상승하고 있는 것을 알 수 있다.

〈표 5〉 시·군·구 매매가격과 전세가격 예시 (단위: 만원/$3.3m^2$)

선거	서울시 강남구		서울시 은평구	
	평균 매매가	평균 전세가	평균 매매가	평균 전세가
19 대 국선	3,380	1,392	1,221	678
18 대 대선	3,105	1,414	1,206	711
6 회 지선	3,155	1,607	1,219	786

3) 투표율

투표율 역시 보수정당 지지성향을 설명할 수 있는 주요한 변수이다. 고령층일수록 투표율이 높기 때문에 투표율이 높을수록 보수정당 지지가 높아질 가능성이 있다. 이런 점을 고려하여 회귀모형에서 설명변수로 투표율을 포함시키기로 한다. 투표율은 읍·면·동 단위에서는 관찰되지 않기 때문에 시·군·구별 투표율을 중앙선거관리위원회 선거관리통계시스템에서 얻을 수 있다. j 선거에서 읍·면·동 i의 투표율(x_{4ij})은 해당 읍·면·동이 속한 시·군·구의 투표율(%)로서, 투표자수를 선거인수로 나눈 것에 100을 곱한 값이다.

<표 6>에서는 예시적으로 서울시 강남구와 은평구에 대한 각 선거에서의 투표율을 정리하였다. 대선 투표율이 가장 높고 지선과 국선에서는 50% 중반 수준이다.

〈표 6〉 **투표율 변수 예시** (단위: %)

선거	서울시 강남구	서울시 은평구
19 대 국선	54	53
18 대 대선	74	73
6 회 지선	57	56

4) 시·도 단위 더미변수 및 기타 후보자 출마 변수

우리나라에서는 지역투표 성향이 매우 분명하게 나타난다. 종속변수인 y_{ij}는 영남지역에서는 1보다 훨씬 큰 값이 나타날 것이고, 호남지역에서는 1보다 훨씬 작은 값이 나타날 것이다. 또한 수도권 지역에서는 1에 가까운 값이 나타날 것으로 예상할 수 있다. 따라서 각 읍·면·동이 속한 17개 시·도에 대한 더미변수를 회귀모형에 포함시켜 추정한다. 관심대상인 각 읍·면·동이 도시에 속해 있는지 그렇지 않은지도 정당지지 성향에 영향을 미칠 수 있다. 따라서 각 읍·면·동이 '시' 또는 '구'에 속하면 1의 값을 주고 "군"에 속하면 0의 값을 주는 더미변수를 모형에 포함시킨다.

선거에서 주요 정당인 새누리당과 새정연 후보 외에도 무소속, 진보정당 또는 기타 정당 후보들이 출마할 수 있다. 특정 지역에서는 진보정당 후보의 출마가 새정연 후보의 득표에 영향을 줄 수 있다고 주장한다.[9] 각 선거에서 진보정당 후보가 출마했을 때는 1, 그렇지 않은 경우에는 0의 값을 갖는 더미변수를 모형에 포함한다. 18대 대선에서는 진보정당 후보가 출마하지 않았기 때문에 모든 읍·면·동에서 이 변수는 0의 값을 갖게 된다. 진보정당 후보를 제외한 무소속 및 기타 후보가 출마한 경우에는 그 후보자 수를 역시 설명변수로 포함한다. <표 7>에 실증분석모형에서 사용하는 종속변수와 설명변수가 정리되어 있다.

9) '야권 후보 단일화' 문제와 관련 있는 주장이다.

<표 7> 종속변수와 설명변수

	변수이름	관찰 단위	정의	연속형/ 범주형 구분
종속 변수	y	읍·면·동	새누리당 후보 우세비율(배수)	연속형
설명 변수	x_1	읍·면·동	50 세 이상 인구 비율(%)	연속형
	x_2	시·군·구	평균 주택매매가격(만원)	연속형
	x_3	시·군·구	평균 주택전세가격(만원)	연속형
	x_4	시·군·구	투표율(%)	연속형
	x_5	시·도	17 개 시·도 더미변수 강원도($x_{5,1}$)~충청북도($x_{5,17}$)	범주형
	x_6	시·군·구	도시 더미변수 (시/구=1, 군=0)	범주형
	x_7	선거구	진보정당을 제외한 기타 정당 후보 출마자 수	이산형
	x_8	선거구	진보정당 후보 출마 더미변수 (출마=1, 불출마=0)	범주형
	x_9		선거유형 더미변수 국선($x_{9,1}$), 대선($x_{9,2}$), 지선($x_{9,3}$)	범주형

Ⅳ. 실증분석 및 시뮬레이션

1. 회귀모형 추정결과

이번 절에서는 <표 7>에서 정의한 변수들을 이용하여 선거결과 모형을 설정하고 추정한 결과를 제시하고자 한다. 상위 레벨 (i)은 읍·면·동이고, 하위 레벨(j)은 선거시점(또는 선거유형)으로 정의하고 추정 모형을 다음과 같이 설정한다.[10)]

10) 본 연구모형은 패널모형이라고 부르기 어렵다. 패널모형에서는 시간(j)이 균등하게

$$\log(y_{ij}) = \alpha + \beta_1 x_{1ij} + \beta_2 \log(x_{2ij}) + \beta_3 \log(x_{3ij}) + \beta_4 x_{4ij} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7ij} + \beta_8 x_{8ij} + \beta_9 x_{9j} + u_i + e_{ij} \quad (3)$$

종속변수인 y_{ij}는 로그를 취하여 사용한다. 로그를 취함으로써 y_{ij}의 예측값을 항상 0보다 크게 만들 수 있는 장점이 있다. $\log(y_{ij})$의 의미는 새누리당 후보가 새정연 후보에 비해 몇 퍼센트나 더 많은 득표를 했는지의 (로그)비율을 의미한다. y_{it}가 새누리당 후보의 우세비율 '배수'인 데 반해 $\log(y_{ij})$는 '우세율'이다. 한편, x_{5i}(시·도 더미변수)와 x_{6i}(도시 더미변수)는 시간불변(time-invariant)인 설명변수로 표시하고 있다. 멀티레벨 모형에서는 관찰되지 않는 읍·면·동의 이질성을 u_i 오차항으로 고려할 수 있다. 현실적인 데이터의 제약으로 읍·면·동 단위에서 새누리당과 새정연의 상대적 지지율을 설명하는 모든 변수를 모형에 포함시키기는 불가능하다. 모형에서 제외된 설명변수 중 시간불변 설명변수를 읍·면·동 이질성(town heterogeneity)인 u_i를 통해 통제할 수 있는 장점이 있다.

식(3)의 멀티레벨 모형을 추정할 때, u_i를 고정효과(fixed effects) 또는 확률효과(random effects)로 가정할 수 있다. 본 연구에서는 확률효과 추정결과를 제시한다.[11] 확률효과 추정을 위해서는 오차항 u_i를 확률변수(분포)로 가정한다. 일반적으로 $u_i \sim (0, \sigma_u^2)$로 가정한다.[12] 읍·면·동(i)과 시점(t)에 따라 변하는 오차항 e_{it} 역시 확률변수이고 $(0, \sigma_e^2)$로 가정할 수 있다. <표 8>에서는 멀티레벨-확률효과 모형의 추정결과가 나와 있다.[13]

배열(equally-spaced)되어야 하지만 2012년에 국선과 대선이 모두 치러졌기 때문이다. 따라서 패널모형이라기보다는 멀티레벨 모형이라고 부르는 것이 좀 더 정확한 표현으로 판단된다.

11) 설명변수와 관찰되지 않는 그룹 이질성(u_i)이 서로 상관관계가 없다면 확률효과가 고정효과 추정량에 비해서 더 효율적이라고 알려져 있다(민인식·최필선 2012).

12) 최우추정 대신 GLS 추정량을 사용하는 경우에는 정규분포를 가정할 필요는 없다.

13) 멀티레벨 모형 추정결과는 Stata 13.0 통계패키지를 이용하여 얻었다.

<표 8> 멀티레벨 모형 추정결과

설명변수	추정 계수
x_1: 50세 이상 인구비율	0.017*** (0.0004)
$\log(x_2)$: 주택 매매가격	0.313*** (0.024)
$\log(x_3)$: 주택 전세가격	-0.320*** (0.026)
x_4: 투표율	0.0037*** (0.0007)
x_6: 도시 더미변수	-0.020 (0.015)
x_7: 기타 후보 출마자 수	-0.005 (0.005)
x_8: 진보정당 출마 여부	-0.010 (0.012)
$x_{9,2}$: 선거유형 더미(대선)	0.086*** (0.021)
$x_{9,3}$: 선거유형 더미(지선)	0.045** (0.013)
상수항	-0.701*** (0.137)
overall R^2	0.914
표본 크기	9,330
읍·면·동 수	3,498

주: 1) 추정 계수 아래 괄호 안은 표준오차이다.

2) ***, **, *는 각각 1%, 5%, 10% 수준에서 유의함을 의미한다.

3) x_5(시·도 더미변수)는 모형에 포함하여 추정하였으나 지면제약 상 추정치를 제시하지 않았다. 시·도 더미변수는 모두 10% 유의수준에서 통계적으로 유의하다.

먼저 인구구성 변수의 추정계수를 해석하면, 50세 이상 인구비율이 증가할수록 새누리당 후보의 우세율이 유의하게 높아진다.[14] 고령화가 보수지지 성향

을 높인다는 선행연구 결과와 일치한다. 50세 이상 인구비율이 1%p 증가하면 새누리당 우세율이 1.7%p 높아진다. 투표율 역시 1% 수준에서 유의하며, 고령화 변수와 유사하게 작용하는 것으로 나타났다. 즉, 투표율이 높아질수록 새누리당 후보의 우세율이 높아진다. 투표율이 1%p 높아지면 새누리당 우세율이 0.37%p 높아진다. 투표율이 50세 이상 인구비율에 비해 종속변수에 미치는 영향이 작아 보이지만, 실질적으로는 더 크다고 할 수 있다. 왜냐하면 50세 이상 인구비율은 주어진 지역에서 시간적으로 매우 완만하게 변화하는데 반해 투표율은 상대적으로 더 크게 변할 수 있기 때문이다. 가령 투표율이 10%p가 높아질 경우 새누리당 우세율은 3.7%p 높아진다.

주택가격 변수 역시 종속변수에 유의한 영향을 미친다. 주택 매매가격이 높아질수록 새누리당 우세율이 높아지지만 전세가격이 높아지면 오히려 새누리당 우세율이 낮아진다. 매매가격 상승은 유권자의 자산가격이 상승한다고 해석할 수 있다. 자산가격 상승은 보수적인 여당후보 지지성향을 높이는 것으로 추정된다. 그러나 전세가격은 주거불안정으로 이어질 수 있기 때문에 주택정책을 책임지고 있는 여당에 불리하게 작용하는 것으로 분석된다. 매매가격이 1% 높아지면 새누리당 우세율이 0.31%p 증가하고 전세가격이 1% 높아지면 오히려 우세율이 0.32%p 감소한다.

도시 더미변수 역시 예상대로 음(-)의 추정계수가 얻어진다. 도시에 위치한 읍·면·동일수록 새누리당 후보 우세율이 낮아진다. 그러나 근소하게 10% 수준에서 유의하지 않는 것으로 나타났다. 진보정당 출마여부와 기타 후보자 출마자 수 변수는 통계적으로 유의하지 않다. 선거유형 변수를 살펴보면 국선에 비해 대선과 지선에서 새누리당 우세율이 더 높아진다. 특히 다른 조건이 일정할 때 대선에서 새누리당 우세율이 가장 높아진다. 최근 치러진 세 번의 전국단위 선거에서 새누리당이 전체적으로 우세하였고, 또한 대선에서도 승리한 결과를 반영하고 있는 것으로 보인다. 가장 중요한 정책담당자를 뽑는 대선에서 보수후보를 지지하는 성향이 더 두드러지게 나타난다고 해석할 수 있다.

---

14) 최근 치러진 세 번의 전국 단위 선거데이터에서 얻은 결과임에 주의하여 해석하여야 한다. 따라서 추정결과는 가까운 미래에 치러질 선거예측에만 사용할 수 있다는 한계가 있다.

$R^2 = 0.91$로 매우 높은 편이므로 연구모형은 표본 내(in-sample) 적합도가 적절하다고 판단된다. 확률효과 모형의 적절성에 대한 가설검정을 실시하였다. $var(u_i) = \sigma_u^2 = 0$이면 확률효과 모형을 추정하는 것보다 통합(pooled) OLS 추정이 더 효율적이다. 우도비 검정(likelihood ratio test) 결과는 카이제곱 검정통계량이 620이고 p값은 0에 가깝다. 따라서 귀무가설이 기각되어 멀티레벨 확률효과 추정량이 더 적절하다고 결론내릴 수 있다.

관찰되지 않는 읍·면·동의 이질성에 해당하는 u_i의 추정치 $\hat{u}_i$는 추정계수를 이용하여 얻을 수 있다. 직관적으로 $\hat{u}_i > 0$이면 새누리당 우세율이 높은 읍·면·동일 가능성이 크고, $\hat{u}_i < 0$이면 새누리당 우세율이 낮은 읍·면·동일 것으로 예상된다. <표 9>에서는 예시적으로 3개 읍·면·동의 $\hat{u}_i$ 값을 비교하고 있다.

<표 9> u_i 추정치

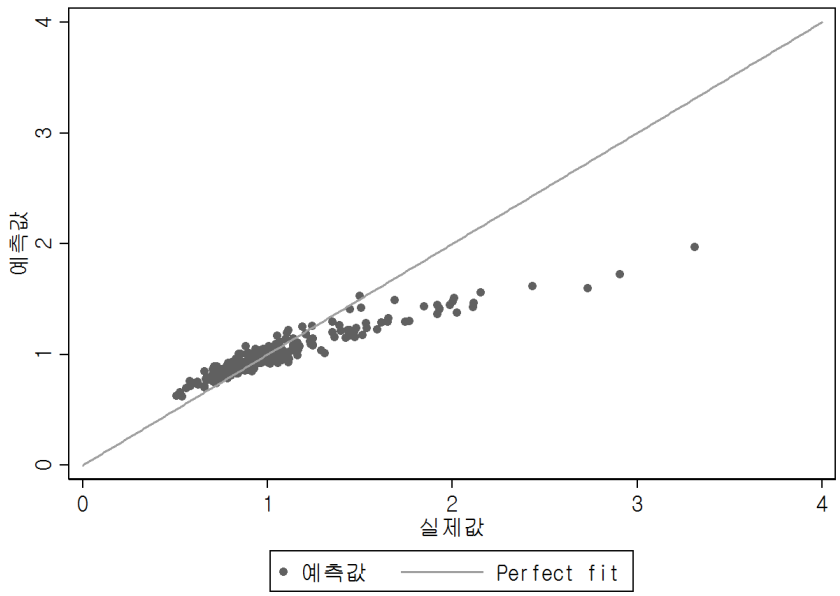
읍·면·동	$\hat{u}_i$
서울특별시 성동구 응봉동	0.081
경상남도 의령군 봉수면	0.106
전라남도 영광군 군남면	-0.033

식(3)의 추정결과를 이용하면 특정 읍·면·동의 새누리당 우세비율(y_{ij}) 예측치를 계산할 수 있으며, 다음과 같이 표현할 수 있다.

$$\hat{y}_{ij} = \exp\left(\hat{\alpha} + X_{ij}\hat{\beta} + \hat{u}_i\right) \quad (4)$$

<그림 1>에서는 18대 대선에서 나타난 서울특별시 읍·면·동의 새누리당 우세비율과 식(4)로 구한 예측치를 비교하고 있다. 모든 점이 45도 선 위에 있다면 완벽한 적합(perfect fit)이 된다. 많은 점들이 1 주위에 겹쳐져 있어 정확한 판별은 어렵지만, 거의 대부분의 점들이 45도선(perfect fit) 근처에 있는 것

으로 나타났다. 그러나 새누리당 우세비율이 매우 높은 지역(즉, y_{ij}값이 매우 높은 지역)에서는 본 연구모형이 y_{ij}를 과소예측(under-prediction)하고 있음을 알 수 있다.[15]



〈그림 1〉 18대 대선 예측값과 실제값 비교: 서울특별시 읍·면·동

2. 20대 국회의원 선거 시뮬레이션

본 절에서는 <표 8>의 추정결과를 선거 시뮬레이션에서 어떻게 활용할 수 있는지 설명하고자 한다. 미래시점 선거(h)에서 읍·면·동 단위(i)에서 새누리당 우세비율은 다음과 같이 나타낼 수 있다.

15) 서울시 읍·면·동 중 과소예측(under-prediction)이 심한 상위 5곳은 다음과 같다. 강남구 압구정동, 강남구 도곡2동, 강남구 대치1동, 송파구 잠실7동, 서초구 반포2동이다. 서울 내에서 새누리당 지지율이 매우 높은 곳에서 과소예측 결과가 도출되고 있다. 즉 본 연구모형에서는 같은 권역(가령 서울특별시) 내의 타 읍·면·동에 비해 새누리당 지지율이 극단적으로 높은 지역에서는 예측모형으로서 한계점을 지니고 있다.

$$\hat{y}_{ih} = \hat{E}(y_{ih}|u_i, X_{ih}) = \exp\left(\hat{\alpha} + X_{ih}\hat{\beta} + \hat{u}_i\right) \quad (5)$$

여기에서 $\hat{u}_i$는 시간불변인 해당 읍·면·동의 이질성이기 때문에 향후 선거에서도 일정하게 유지된다.

시뮬레이션에서는 예시적으로 서울특별시 종로구에 속한 17개 동의 20대 국선 결과를 예측해보고자 한다. 식(5)를 이용하여 $\hat{y}_{ih}$를 얻기 위해서는 미래 선거시점(h)에서 X_{ih} 값을 가정해야 한다. 관찰된 설명변수인 X_{ih}는 미래 선거시점에서 얻을 수 있다고 가정하자. 가령 2016년 20대 국선을 예측하기 위해서는 2016년 초에 관찰된 읍·면·동 단위의 연령별 주민등록 인구와 주택가격 변수 등이 필요하다. 그런데 2016년 초의 50세 이상 동별 인구비율은 알 수 없기 때문에 2015년 초에 관찰된 값으로 대신한다. 서울시의 경우 동별 50세 이상 인구비율을 서울시청 홈페이지에서 구할 수 있다. 주택 매매가격과 전세가격 변수는 국토교통부 실거래가 홈페이지에서 2015년 1월부터 2015년 6월까지 6개월 간 서울시 종로구의 매매가격과 전세가격 평균으로 대신한다. 기타 정당 후보와 진보정당 후보 출마여부는 각 선거구에서 임의로 가정한다. 서울특별시 종로구 시뮬레이션에서는 기타 정당후보는 1명 출마, 진보정당 후보는 출마하지 않았다고 가정해보자. <표 8>에서 유의한 영향을 미치는 것으로 나타난 투표율에 대한 가정도 중요하다. 본 시뮬레이션에서는 투표율이 55%~75%까지 광범위하게 변한다고 가정하고 각 투표율 값에서 선거결과를 예측한다.

투표율 가정에 따라 각 시뮬레이션에 얻은 $\hat{y}_{ih}$는 새누리당 득표수를 새정연 득표수로 나눈 배수이다. 선거결과를 판단하기 위해서는 각 읍·면·동의 새누리당과 새정연 후보의 득표수를 $\hat{y}_{ih}$을 이용해서 계산해야 한다. 식으로 표현하면 다음과 같다.

$$\text{새누리당 후보 득표수} : n_{ih}^{\text{새누리당}} = \frac{n_{ih}y_{ih}}{1+y_{ih}} \quad (6)$$

$$\text{새정연 후보 득표수 : } n_{ih}^{\text{새정연}} = n_{ih} - n_{ih}^{\text{새누리당}} \tag{7}$$

위 식에서 n_{ih}는 읍·면·동 i의 선거 h에서 새누리당 후보와 새정연 후보가 얻은 총 득표수이다.[16] 해당 선거구에서 새누리당과 새정연 후보가 얻은 총 득표수는 선거구 내의 읍·면·동 득표수를 모두 합하여 구한다.

$$n_h^{\text{새누리당}} = \sum_{i=1} n_{ih}^{\text{새누리당}} \tag{8}$$

위 식에서 i는 종로구에 속한 17개 동이며, $n_{ih}^{\text{새누리당}}$는 식(6)에서 계산한 값이다. <표 10>에서는 서울특별시 종로구 국회의원 선거구에 속한 17개 동을 나열하고 있다.

<표 10> 서울특별시 종로구 국회의원 선거구

선거구	읍·면·동
서울특별시 종로구	청운효자동, 사직동, 삼청동, 부암동, 평창동, 무악동, 교남동, 가회동, 종로 1·2·3·4 가동, 종로 5·6 가동, 이화동, 혜화동, 창신 제 1 동, 창신 제 2 동, 창신 제 3 동, 숭인 제 1 동, 숭인 제 2 동

<표 11>에서는 투표율에 따른 종로구 각 동의 우세율 시뮬레이션 결과를 보여준다. 먼저 투표율이 55%로 낮은 경우에는 평창동과 종로1.2.3.4가동에서만 새누리당이 앞서는 것으로 나온다. 그러나 투표율이 75%로 높아지면 평창동, 종로1.2.3.4가동뿐만 아니라 사직동과 삼청동에서도 새누리당이 새정연보다 앞서는 결과를 얻는 것으로 나타난다. <표 8>의 추정결과에서 예상할 수 있듯이 합리적 범위 내 투표율을 가정하면 투표율이 높아질수록 새누리당 우

16) n_{ih} = 선거인수 × 투표율 × $(1-other_{ih})$로 계산할 수 있다. $other_{ih}$는 해당 읍·면·동 선거에서 새누리당과 새정연 이외 후보가 얻은 득표율이다. 종로구의 경우 $other_{ih}=5\%$로 가정한다.

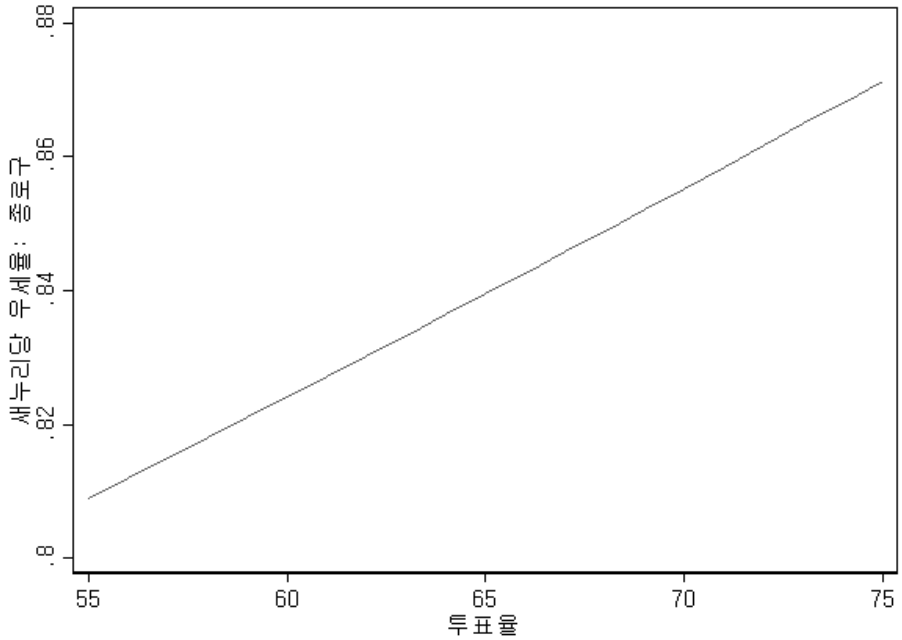
세율이 증가한다. <표 11>에서는 식(8)의 공식에 따라 종로구 국선 결과를 제시한다. 모든 투표율 가정에서 새누리당이 열세임을 알 수 있다.[17]

<표 11> $\hat{y}_{ih}$ 시뮬레이션 결과

읍·면·동	투표율		
	55%	65%	75%
청운효자동	0.784	0.813	0.845
사직동	0.977	1.015	1.054
삼청동	0.989	1.027	1.066
부암동	0.860	0.892	0.927
평창동	1.041	1.081	1.122
무악동	0.818	0.849	0.881
교남동	0.764	0.793	0.824
가회동	0.844	0.876	0.909
종로 1·2·3·4 가동	1.025	1.064	1.104
종로 5·6 가동	0.851	0.884	0.917
이화동	0.655	0.680	0.706
혜화동	0.689	0.716	0.743
창신제 1 동	0.774	0.804	0.834
창신제 2 동	0.663	0.688	0.715
창신제 3 동	0.713	0.741	0.769
숭인제 1 동	0.730	0.758	0.786
숭인제 2 동	0.726	0.753	0.782
합계 (득표수)	새누리당 30,299	새정연 37,460	새누리당 36,547	새정연 43,532	새누리당 43,024	새정연 49,375
	비율: 0.808		비율: 0.839		비율: 0.871	

17) <그림 1>에서 확인할 수 있듯이 서울지역에서는 새누리당의 득표수가 상대적으로 과소평가되는 경향이 있다. 실제 선거에서는 같은 투표율에서 새누리당의 득표수가 더 나올 것으로 예상할 수 있다.

<그림 2>에서는 투표율이 55%~75%일 때, 종로구에서 새누리당 후보와 새정연 후보의 득표수 비율을 그래프로 보여준다.[18] 투표율이 증가할수록 새누리당 우세비율이 꾸준히 증가하기는 하지만 모든 투표율 범위에서 여전히 1보다 작은 값을 가지기 때문에 열세임을 확인할 수 있다. 따라서 종로구 국회의원 선거에서는 투표율보다는 경제환경 변수(주택가격과 전세가격), 모형에서 고려되지 않은 요인들(입후보자 변수 등)이 새누리당 후보와 새정연 후보의 격차를 줄이는 데 중요한 역할을 한다고 볼 수 있다.



〈그림 2〉 국회의원 선거에서 새누리당 후보 우세비율 예측: 서울시 종로구

V. 요약 및 결론

우리나라에서 선거예측에 대한 연구는 주로 여론조사에 근거한 예측이 이

18) <그림 2>는 20대 국회의원 선거시점의 x값을 가정하고 계산된 결과이기 때문에 과거 세 번 선거에서의 종로구 선거결과와 직접적으로 비교하기 어렵다.

루어져 왔다. 여론조사는 후보자가 결정된 상태에서 직접 유권자를 조사하여 발표하기 때문에 정확성 측면에서는 우수하다고 볼 수 있지만, 비용(경제성)이 많이 들고, 선거에 출마할 후보들이 결정된 시점에서 이루어져야 하는 제약이 있다(사전성).

본 연구에서는 경제성과 사전성 측면에서 장점이 있는 모델링에 기초한 정당 우세율 분석모형을 제시하였다. 모형의 범용성을 높이기 위한 방법으로 읍·면·동을 단위로 한 모형을 개발하였다. 이에 따라 어느 특정 선거가 아니라 대선, 국선 및 지선 등 다양한 선거에 적용될 수 있다. 또한 모형의 정확도를 높이기 위해 멀티레벨 데이터를 구축함으로써 관찰되지 않는 이질성을 포착하고자 했으며, 선거에 영향을 미치는 경제환경 변수로서 주택가격을 포함시킨 점도 특징이다. 이러한 방법들을 통해 선거예측모형이 갖춰야 할 네 가지 조건인 정확성, 경제성, 사전성, 범용성을 최대한 높이고자 하였다.

최근 치러진 세 번의 선거(2012년 총선, 2012년 대선, 2014년 지선)에서 나타난 정당지지 결과를 활용하여 멀티레벨 확률효과 모형을 추정한 결과는 다음과 같다. 첫째, 고령화 요인이 중요하게 작용한다. 50세 이상 인구비율이 1%p 증가하면 새누리당 우세율이 1.7%p 높아진다. 둘째, 투표율 역시 고령화 변수와 동일한 방향으로 작용하여 투표율이 1%p 높아지면 새누리당 우세율이 0.37%p 높아진다. 셋째, 주택 매매가격이 높아질수록 새누리당 우세율이 높아지지만 전세가격이 높아지면 오히려 새누리당 우세율이 낮아진다. 매매가격이 1% 높아지면 새누리당 우세율이 0.31%p 증가하고, 전세가격이 1% 높아지면 오히려 우세율이 0.32%p 감소한다. 넷째, 도시에 위치한 읍·면·동일수록 새누리당 후보 우세율이 낮아지지만, 근소하게 10% 수준에서 유의하지 않는 것으로 나타났다. 다섯째, 진보정당 출마여부와 기타 후보자 출마자 수 변수는 통계적으로 유의하지 않다. 여섯째, 다른 조건이 일정할 때 대선에서 새누리당 우세율이 가장 높아진다. 가장 중요한 정책담당자를 뽑는 대선에서 보수후보를 지지하는 성향이 더 두드러지게 나타난다고 해석할 수 있다.

본 연구에서 제시한 읍·면·동-선거 단위의 데이터를 통해 어떤 설명변수들이 득표율에 영향을 미치는지 파악할 수 있으며, 이를 통해 후보자들은 선거대책을 마련할 수 있다. 특히 추정결과에 기초하여 읍·면·동 단위의 선거예측 시

뮬레이션을 해볼 수 있다는 점에서 활용도가 매우 높을 것으로 기대한다.

참고문헌

김길수. 2013. “선거결과 예측에 관한 연구: 19대 총선 ‘종로구’를 중심으로.” ≪정치·정보연구≫ 16(1): 137-161.

김석우. 2006. “17대 총선과 정칙적 충원.” ≪한국정치외교사논총≫ 27(2): 287-315.

김재한. 1993. “제14대 대선과 한국경제.” ≪한국정치학회보≫ 27(1): 99-120.

김재한. 1995a. “여론조사식 선거예측의 허와 실.”≪성곡논총≫ 26: 1247-1269.

김재한. 1995b. “한국유권자의 이념분포와 정계구도.” 김재한 외. ≪한국정치외교의 이념과 논쟁≫ 11-59, 서울: 소화출판사.

김재한. 1995c. “한국선거예측의 방법론적 모색.”≪한국정치학회보≫ 29(1): 221-241.

민인식·최필선. 2012. STATA 패널데이터 분석, 서울: 지필출판사.

박명호·김민선. 2008. “한국 총선에서 나타난 현직자의 재선 추이에 관한 분석.” ≪동국대학교 사회과학연구≫ 15(1): 161-176.

송근원. 2011. “후보자 득표율 예측모형과 지표의 구성.” ≪조사연구≫ 12(1): 31-63.

송근원·정봉성. 2007. “16대 대선에서 나타난 유권자들의 정책성향과 투표 행태.” ≪21세기 정치학회보≫ 17(1): 45-70.

신계균. 2012. “미국 대선 예측의 연구의 과거와 현재.” ≪미래정치연구≫ 2(1): 5-36.

조진만·최준영·가상준. 2006. “한국 재·보궐선거의 결정요인 분석.” ≪한국정치학회보≫.20(2), 75-100.

최필선·민인식. 2015. “고령화가 투표의 보수지지 성향에 미치는 영향: 최근 전국단위 선거를 중심으로.” ≪조사연구≫ 16(1): 89-115.

한정택. 2007. “한국 현직 국회의원 재당선 요인분석.” ≪한국정치학회보≫ 40(2): 73-99.

Campbell, J.E. 1992. “Forecasting the Presidential Vote in the States.” *American Journal of Political Science* 15: 386-407.

Greene, J. 1993. “Forewarned Before Forecast: Presidential Election Forecasting Models and the 1992 Election.” *Political Science* 26: 17-21.

Holbrook, T.M. 1991. “Presidential Elections in Time and Space.” *American Journal of Political Science* 35: 91-109.

Krammer, G.H. 1971. “Short Term Fluctuations in U.S. Voting Behavior, 1896-1964.” *American Political Science Review* 65: 131-143.

Lewis-Beck, M.S. and T.W. Rice. 1982. “Presidential Popularity and Presidential Vote.” *Public Opinion Quarterly* 46: 534-537.

Lewis-Beck, M.S. and T.W. Rice. 1984. “Forecasting Presidential Elections: A Comparison of Naive Models.” *Political Behavior* 6: 9-21.

Norpoth, H. 2004. “From Primary to General Election: A Forecast of the Presidential Vote.” *Political Science and Politics* 37: 737-740.

Rosenstone, S.J. 1983. *Forecasting Presidential Elections*. New Haven, CT: Yale University Press.

Singleman, L. 1979. “Presidential Popularity and Presidential Elections.” *Public Opinion Quarterly* 43: 532-534.

Tufte, E.R. 1978. *Political Control of the Economy*. Princeton, NJ: Princeton University Press.

<접수 2015/9/18, 수정 2015/11/11, 게재확정 2016/1/21>