

## 2024년 <한국언론학보> 수록 논문 온라인 부록

### 한국 언론은 선거 여론조사 결과를 선택적으로 보도하는가?

#### 제20대 대선에서 조사기관 편향 및 조사결과의 격차와 보도 빈도의 관계에 대한 베이지안 분석

##### 1. 조사기관 편향 추정을 위한 상태-공간 모형의 추가 설명

여론조사 기관의 편향 추정을 위한 상태 공간 모형은 다음과 같은 수식으로 표현할 수 있다(박종희, 2013):

$$x_{ijt} = \alpha_{ij} + \gamma_{jt} + D_{ijt}\beta_j + e_{ijt} \quad e_{ijt} \sim N(0, \sigma_e^2)$$

$$\gamma_{jt} = \gamma_{jt-1} + \epsilon_{jt} \quad \epsilon_{jt} \sim N(0, \sigma_\epsilon^2)$$

여기서  $x_{ijt}$ 는  $i$  조사기관이  $j$  후보에 대해  $t$ 시점의 지지율을 관측한 결과로, 첫번째 수식은 이 관측치가  $i$  조사기관의  $j$  후보에 대해 가지는 편향, 즉 하우스 효과를 나타내는  $\alpha_{ij}$ , 관측 불가능하나 추정하고자 하는  $t$  시점의  $j$  후보에 대한 실제 여론의 추세  $\gamma_{jt}$ ,  $j$  후보를 제외한 나머지 후보 중 하나가 설문에서 빠짐으로써 발생하는 반사이익(메뉴효과)을 표현하는 벡터  $\beta_j$ 와 오차항  $\epsilon_{ijt}$ 의 합으로 구성되는 것을 모형화하고 있다. 이는 관측치가 어떻게 결정되는가를 모형화 한 것이므로 상태공간모형에서는 관측방정식(observation equation)이라고 불린다. 반면, 관측치를 통해 추정하고자 하는 실제 여론 추세의 시계열에 대한 모형은 상태방정식(state equation)이라고 하는데, 위 모형에서는 이 추세가 임의보행(random walk) 과정을 따른다고 본다. 임의보행 모형은 안정시계열(stationary time-series)와 달리, 시계열 데이터가 회귀하는 장기추세가 존재하지 않으며 시계열의 변화가 지속적이고 항구적임을 의미하므로, 여론의 시계열에 적합한 가정이라고 할 수 있다(Box-Steffensmeier, Freeman, Hitt & Pevehouse, 2014). 위의 모형에서 관측 불가능한  $\gamma_{j0}$ 와  $\alpha$ 에 대해 적절한 사전분포(prior distribution)를 부여한 후, 베이지안 업데이트를 통해 해당 모수를 추정할 수 있는데, <여론M>의 주요 관심사가 여론 추세인  $\gamma$ 라면, 본 연구는 각 조사기관의 편향에 해당하는  $\alpha$ 를 분석에 이용하고자 하는 것이다. 단, 하나의 조사를 여러 기관이 함께 수행한 경우에는 해당 조사가 각 조사기관이 별도로 수행한 복수의 조사인 것처럼 처리하여 분석에 포함시켰다.

박종희 교수팀이 깃헙에 공개한 해당 모형의 추정 코드는 베이지안 추정의 방법중 하나인 깃스 샘플링을 자동화한 확률적 프로그래밍 언어 JAGS(Just Another Gibbs Sampler)를 사용하였으나<sup>1</sup>, 본 연구에서는 해당 코드를 Stan (Stan Development Team, 2024)으로 리팩토링하여 이용하였다. Stan은 많은 경우 기존 마르코프체인 몬테카를로(MCMC) 알고리즘에 비해 더욱 빠르고 안정적인 성능을 보이는 것으로 알려진 해밀토니안 몬테카를로(Hamiltonian Monte Carlo) 알고리즘을 구현한 확률적 프로그래밍 언어로 알고리즘의 효율성과 속도라는 장점 이외에도, 몬테카를로 샘플링에 수렴(convergence) 문제에 대한 다양한 측정치를 자동으로 계산하여, 더욱 신뢰할 수 있는 추정이 가능하다는 장점이 있다. 실제로, 기존 JAGS 코드를 그대로 Stan에서 구현하였을 때 샘플링 과정이 수렴하지 않는 문제점이 발견되어, 본 연구에서는 사전분포에 대한 설정을 다소 변경하여 추정하였다.<sup>2</sup> 본 연구를 위해 수행한 베이지안 업데이트는  $\hat{R}$ (Brooks & Gelman, 1998), 실효샘플크기(Effective Sample Size; ESS) 기준으로 보았을 때 신뢰성 있는 결과임을 알 수 있었다(Geyer, 2011).

## 2. 여론조사의 보도 횟수 예측을 위한 계층적 회귀분석 추가설명

$$y_{os} \sim \text{Poisson}(\lambda_{os})$$

$$\log(\lambda_{os}) = \beta_0 + \beta_{0o} + \beta_1 r_{os} + \beta_{2o} \text{diff}_s + \beta_{3o} \text{house}_{i(s)} + \beta_4 bc_o$$

$$\begin{pmatrix} \beta_{0o} \\ \beta_{2o} \\ \beta_{3o} \end{pmatrix} \sim MVN \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, S \right]$$

$$S = \begin{pmatrix} \sigma_0 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} R \begin{pmatrix} \sigma_0 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

<sup>1</sup> <https://github.com/jongheepark/poll-MBC>

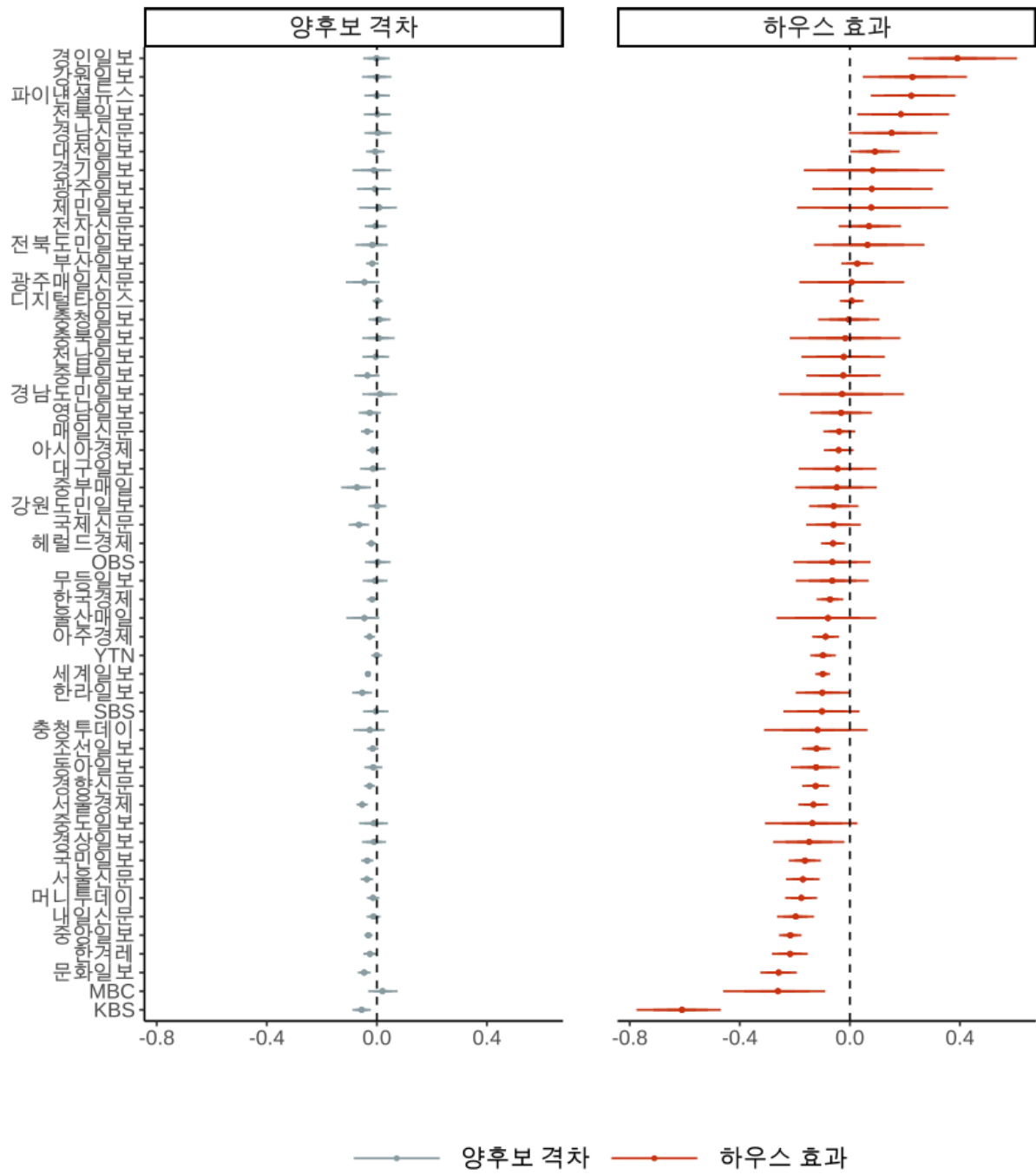
<sup>2</sup> 구체적으로, 종래 사용하던 균등분포(uniform distribution) 기반 사전분포는 신뢰할 수 없는 베이지안 업데이트 과정으로 이어지기 쉽다는 것이 발견되어, 최근에는 그보다 조금 더 정보가 많은(weakly informative)한 사전분포를 이용하는 추세이다(Gelman, Jakulin, Pittau & Su, 2008). 본 연구에서 사용한 자세한 설정은 github을 통해 공개할 예정이다.

있으며, 해당 두 모수는 언론사의 여론조사 보도 성향에 해당하는  $\beta_{0o}$ 와 함께 다변수정규분포를 따르고, 해당 다변수정규분포의 공분산 행렬( $S$ )을 분해하여 상관계수행렬( $R$ )이  $LKJ(3)$  사전 분포(Lewandowski, Kurowicka & Joe, 2009)를 따르도록 하는 일반적인 설정을 따랐다.

Table A.1: Results of Population-Level Estimation (with one bias each)

<i>Predictors</i>	하우스 효과		양후보 격차	
	<i>Estimate</i>	<i>CI (95%)</i>	<i>Estimate</i>	<i>CI (95%)</i>
Intercept	-1.92	[-2.33,-1.51]	-1.93	[-2.46,-1.37]
	(0.21)		(0.27)	
Requested	0.25	[0.16,0.34]	0.44	[0.36,0.54]
	(0.05)		(0.05)	
Broadcaster	-0.89	[-2.25,0.51]	-0.03	[-1.36,1.27]
	(0.70)		(0.67)	
Random Effect				
$\sigma_0$	1.44	[1.17,1.78]	1.47	[1.20,1.82]
$\sigma_2$			0.04	[0.02,0.05]
$\sigma_3$	0.18	[0.14,0.23]		
$\rho_{02}$			-0.24	[-0.62,0.22]
$\rho_{03}$	-0.35	[-0.61,-0.01]		]
N		52		52
Observations		16,172		16,172
Bayesian $R^2$		0.241		0.221

Figure A.1 Separate bias estimations



## 부록

- Box-Steffensmeier, J. M., Freeman, J. R., Hitt, M. P., & Pevehouse, J. C. (2014). *Time Series Analysis for the Social Sciences*. Cambridge University Press.
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455.
- Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (pp. 3-48). CRC Press.
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989-2001.
- Park, J. H. (2013). House bias in the 2012 presidential campaign polls in South Korea. *Survey Research*, 14(1), 1-30. [박종희 (2013). 제18대 대선 여론조사에서 나타난 조사기관 편향. <조사연구>, 14권 1호, 1-30.]
- Stan Development Team (2024) *Stan modeling language users guide and reference manual*. Version 2.35. Retrieved from <https://mc-stan.org>