

시공간 베이지안 계층 모형을 이용한 시군구별 외국인 고용허가제 (E-9 비자) 등록 인원 예측

2020-17530 지리학과 박지호

2025년 10월 8일

1 연구 목적

2025년 상반기를 기준으로, 국내에는 약 156만 명의 외국인이 거주하고 있다. 그 중 약 33만 명의 외국인이 E-9 비전문취업 비자를 취득하여 체류하고 있으며, 이는 29개 체류자격 중 가장 높은 비중을 차지한다. E-9 비자는 내국인 고용 수급이 어려운 단순 노동 등의 비전문 분야에서의 인력 수급을 원활히 하고자 하려는 목적으로 도입되었으며, 이를 특별히 ‘외국인 고용허가제’라는 제도로 구분하여 운영하고 있다. 2004년 시행 이후 해당 제도는 불법 체류 외국인 관리 및 내국인 고용권 보장 등의 목적에 부합하여 산업 현장에서 연착륙하였으나, 국내 인구 구조의 변동과 지역 격차의 다층성 등으로 인해 여전히 수정·보완될 필요성이 존재한다. 특히, 현장에서는 지자체별 특성에 맞는 등록외국인 관리의 필요성과 더불어 그 바탕이 될 실증적 데이터가 부족하다는 점이 지적되고 있다. (정기선, 2025) 또한, 인구 감소로 인해 지방에서의 인력 수급 어려움이 더욱 증가함에 따라 2025년에는 고용허가제로 체류한 외국인이 일정 자격을 취득하면 107개 지역에서 숙련기능인력 비자로 전환할 수 있게 하는 ‘지역특화형 비자’ 사업이 시행되는 등 (최종원, 2025) 외국인 고용허가제 등록 인원 예측에 대한 지역별 분석의 필요성은 더욱 증가하고 있다.

본 연구는 약 17개년에 걸친 등록 외국인 수 패널 자료를 시공간적으로 분석하여, 시군구별 외국인 고용허가제(E-9) 등록 인원의 변동 요인을 규명하고 이를 예측할 수 있는 모형을 개발하는 것을 목적으로 한다. 이를 통해 지역별 외국인 노동자 관리 및 정책 수립을 위한 실증적 근거를 제공하고자 한다. 구체적인 연구 질문은 다음과 같다.

- (RQ1) 시공간 베이지안 계층 모형은 고용허가제 등록 외국인 수를 효과적으로 설명할 수 있는가?
- (RQ2) 지역별 및 산업별 고용허가제 등록 외국인 수에 유의한 영향을 미치는 요인은 무엇인가?
- (RQ3) 지역별 인구 구조의 역동성을 고려할 때, 향후 지역별 등록 외국인 수는 어떤 양상으로 변화할 것인가?

시공간 패널 데이터를 분석할 때에는 시공간 두 차원에서의 자기상관을 적절히 반영하고, 상이한 시공간 해상도와 관측 기간을 지닌 설명변수를 편향을 최소화하며 통합할 필요가 있다. (Alahmadi & Moraga, 2025) 따라서 RQ1에서는 이러한 복합적 문제를 효과적으로 처리할 수 있는 시공간 베이지안 계층 모형이 인구통계 자료를 효과적으로 설명하는지를 핵심적으로 검토한다. 이를 기반으로 RQ2와 RQ3에 답하며, 시군구 수준의 공간적 이질성과 다층성을 반영한 예측 모형을 구축하고자 한다.

2 선행 연구 조사

외국인 고용허가제를 통해 체류하는 외국인은 개발도상국에서 출발한 경제적 목적의 국제 이주자라는 점에서 하나의 이동 흐름으로 유형화될 수 있다. 이 흐름은 실제 데이터 상에서는 특정 시점에 E-9 비자를 취득해 체류하고 있는 총 인원 수인 스톡(stock)의 형태로 집계된다. 한편, 이러한 국가 간 이주의 패턴은 도착 국가 내에서 균일하게 나타나지 않는다. (Allen & Turner, 1996) 그에 대한 이유로는 경제적 요인(산업 구조, 노동 시장), 주택 가격, 도시 특성, 위치,

이주자 네트워크의 유무 등이 제기되어 왔다. (Van Der Gaag & Van Wissen, 2001) 또한, 출발지와 도착지의 인구 수에 비례하고 거리에 반비례한다는 전통적인 중력 모형 역시 인구 이동을 잘 설명하는 것으로 알려져 있다. (Poot et al., 2016) 이러한 이론에 기반하여 실제 데이터를 이용해 국가 간 이주자들의 분포를 지역별 인구 수, GDP 등의 설명변수를 포함하는 회귀 모형으로 분석하는 연구가 이루어져 왔다. (Pu et al., 2019; Saa et al., 2020) 그러나 일반 선형 회귀 모형에서는 시공간 자기상관 혹은 비정상성이 고려되지 않고, 이는 회귀 계수 추정량의 편향을 초래한다. (Jeong et al., 2023) 시공간 회귀 모형을 사용한 경우에도, 집계 시작 시점의 차이로 공간 패널 데이터에 결측치가 존재하거나 데이터의 시공간 해상도(주기, 집계구)가 상이한 경우에 추정량의 불확실성이 크게 증가한다는 한계점이 있다. (Antczak & Lewandowska-Gwarda, 2015)

이러한 한계를 극복할 수 있는 접근으로, 회귀계수의 추정량을 베이지안 방식으로 추정하는 베이지안 계층 모형이 주목받고 있다. 베이지안 계층 모형은 모수를 확률변수로 간주하여 데이터 생성 과정 자체를 확률적 계층 구조로 표현한다는 점에서 기존의 빈도론적 회귀 모형과 구분된다. 또한 베이지안 계층 모형은 시공간적 자기상관을 표현하기 위한 시공간 랜덤효과나 잠재 가우시안 필드 (latent Gaussian field)를 유연하게 포함할 수 있다는 장점이 있다. (Rue et al., 2009) 이러한 접근은 시공간 패널 데이터를 주로 이용하는 환경이나 보건 역학 분야의 연구에서 활발히 적용되고 있다. (Bernardinelli et al., 1995; Blangiardo et al., 2013; MacNab, 2022) 그러나 출생·사망이나 인구 이동 등 인구 지표에의 적용은 상대적으로 적은 편이다. (Azose & Raftery, 2015)

또한, 현재까지 국내 고용허가제 등록 외국인 혹은 저숙련 이주노동자의 분포에 대한 연구는 기술적(descriptive)인 연구이거나 수요 예측과 관계된 연구들이 대부분이다. (유희연, 2023; 이규용, 2024) 따라서 본 연구를 통해 실제 어떤 변수들이 등록 외국인의 수에 영향을 미치는지 파악하고, 인구 구조의 역동성 하에서 지역별 외국인 노동자 수가 어떻게 변화할지를 새로운 관점에서 전망해 볼 수 있을 것으로 기대된다. 특별히 본 연구에서는 기업의 요청에 기반해 수급이 이루어지는 고용허가제의 특성을 고려하여 단순 지자체의 특성보다는 경제 지표나 특정 연령층의 인구 지표를 설명변수로 설정하였다. 이를 통해 지방을 중심으로 한 생산가능인구의 감소 및 생산 분야 노동 기피 현상이 실제로 외국 인력 수급으로 연계되어 등록 외국인 수 증가에 영향을 미치고 있는지를 실증적으로 살펴보게 될 것이다.

3 데이터 및 연구 방법론

3.1 데이터

종속변수로는 <법무부 출입국·외국인정책 통계월보>에서 분기별(3, 6, 9, 12월말)로 제공하는 시군구별 체류자격별 등록외국인 현황 자료를 이용한다. 해당 데이터는 2008년부터 2012년까지는 반기별, 2013년부터는 분기별로 제공되며, 시군구별, 성별, 체류자격별 등록외국인 수가 나타나 있다. 또한 업종별로 구분된 E-9 비자의 세부체류자격인 제조업, 건설업, 농업, 서비스업 등에 대한 등록외국인 수 역시 제공되고 있다. (총 59개 시점 × 226개 시군구)

설명변수 목록은 [Table 1]에 나타나 있다.

관련 요인	변수명	자료 출처	시공간 해상도 및 집계 기간
고용 수요	외국인 근로자 구인 인원	MDIS 직종별사업체노동력조사	(2008년~) 시도별 * 반기별
	외국인 근로자 채용 공고	구인구직 사이트 / 지방의회 회의록 등에서 비정형 데이터 수집	시군구별 * 월별
경제	실업/고용률	통계청 지역별 고용조사	(2008년~ / 2010년부터 분기별) 시군구별 * 반기별
	제조업생산지수	통계청 광업제조업동향조사	(2000년~) 시도별 * 월별
	수출 관련 지표 (금리/수출액)	한국무역협회 지자체별 수출액	(2000년~) 시도별 * 월별
정책	EPS 쿼터	외국인고용관리시스템	(2004년~) 연도별
인구 구조	15-64세, 고령화율, 청년 순 이동	통계청 인구동향조사	시군구별 * 월별 (주민등록인구)
시계열 성분	AR(p) 혹은 SAR(p)		
공간 특성	BYM2		

Table 1: 등록외국인 수 관련 설명변수: 자료 출처 및 시공간 해상도 정리

3.2 모형

시군구 i , 시점 t 에서의 등록 외국인 수를 포아송 분포를 따르는 확률변수로 가정한다.

$$\begin{aligned}
y_{i,t} &| \lambda_{i,t} \sim \text{Poisson}(E_{i,t}\lambda_{i,t}) \\
\log \lambda_{i,t} &= \alpha + \mathbf{x}'_{i,t}\boldsymbol{\beta} + \mathbf{z}'_{j(i)}\boldsymbol{\gamma} + u_{i,t} + v_{j(i)} + \eta_t \\
u_{i,t} &= \rho_u u_{i,t-1} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \text{BYM2}(W, \tau_u, \phi) \\
v_{j(i)} &\sim N(0, \tau_v^{-1}), \quad \eta_t \sim \text{AR}(1)(\rho_\eta, \tau_\eta)
\end{aligned}$$

- $E_{i,t}$ 는 노출량(exposure, 예: 해당 지역 인구수),
- $\mathbf{x}_{i,t}$ 는 시군구 수준의 공변량,
- $\mathbf{z}_{j(i)}$ 는 시도 수준의 공변량,
- $u_{i,t}$ 는 시군구 단위의 시공간 랜덤효과,
- $v_{j(i)}$ 는 시도 단위의 상위 수준 랜덤효과,
- η_t 는 전체 시점별 공통 추세(AR(1) 또는 SAR(1) 과정)이다.

위 모형은 시공간 자기상관을 가우시안 마르코프 랜덤 필드를 이용하여 모델링하게 된다. 먼저, 시군구 단위 시공간 랜덤효과 $u_{i,t}$ 는 공간 랜덤효과를 BYM2(Besag-York-Mollié 2) (Simpson et al., 2017)로, 시간 랜덤효과를 AR(1)으로 모델링하여 결합한다. (공간 랜덤효과를 모델링하는 다른 방법으로는 Leroux 기반 방법이 있다. (Napier et al., 2016; Sun et al., 2023) 한편, 시도 단위 랜덤효과 $v_{j(i)}$ 는 관측치의 수가 17개로 적고, 자기상관을 이중 모델링하는 것을 방지하고자 일반적인 IID 정규분포로 가정한다. 또한, 추세항 η_t 을 통해 전역적 추세를 모델에 반영하도록 하였다.

위의 모형은 결측치 처리 및 상이한 공간 해상도의 통합이 가능하다. 먼저 종속변수의 결측치 (예: 2009년의 1분기 등록외국인 수)는 '예측 분포'로 표현될 수 있으며, 이는 종속변수의 사후분포를 계산하는 과정에서 자동적으로 계산된다. 한편, 공변량에서의 결측치는 해당 공변량 자체를 확률변수로 간주하여 사전분포를 부여하여 내삽을 진행하는 방식을 택할 수 있다. (Gómez-Rubio, 2020) 또한, 위의 [Table 1]과 같이 시군구와 시도 수준의 다중 스케일로 공변량이 나뉘어진 경우 다층 모형의 구조와 유사하게 각각의 공간 단위에 대해 랜덤효과를 부여하여 서로 다른 공간 단위의 공변량을 통합하여 분석할 수 있다. (Lawson, 2018) - *이 부분의 타당성에 대한 추가적 조사 필요

사후분포는 R의 INLA 패키지(Rue et al., 2024)를 이용하여 라플라스 근사를 이용해 추정하게 된다. 이 방법은 전통적인 MCMC에 비해 계산량이 훨씬 적고, 특히 시공간 랜덤효과를 포함하는 복잡한 베이저안 계층모형에서도 안정적인 근사 사후분포를 제공하는 것으로 알려져 있다. (Moraga, 2019)

4 연구 타임라인

현재 전산망 화재로 KOSIS 접속이 불가능하여 일부 데이터의 구득이 지연되고 있는 상황입니다. 현재까지 계획된 설명변수는 모두 온라인으로 다운로드 받을 수 있어 이외의 큰 어려움은 없을 것 같습니다. 우선은 종속변수인 등락 외국인 수 자료에 대해 데이터 전처리를 진행하여 패널 데이터 형태로 구축하였습니다. INLA 패키지 사용법을 확인하여 일부 데이터를 이용해 위의 모형을 실제로 적합할 수 있는지를 10월 중으로 확인해 볼 예정이며, 이후 예정된 11월 18일까지 중간 분석 결과를 제출할 수 있도록 하겠습니다.

References

- Alahmadi, H., & Moraga, P. (2025). Bayesian modelling for the integration of spatially misaligned health and environmental data. *Stochastic Environmental Research and Risk Assessment*, 39(4), 1485–1499. <https://doi.org/10.1007/s00477-025-02927-z>
- Allen, J. P., & Turner, E. (1996). Spatial patterns of immigrant assimilation *The Professional Geographer*, 48(2), 140–155. <https://doi.org/10.1111/j.0033-0124.1996.00140.x>
- Antczak, E., & Lewandowska-Gwarda, K. (2015). Spatial durbin panel model in analysis of migration processes in european cities. 41-42, 25–36.
- Azose, J. J., & Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, 52(5), 1627–1650. <https://doi.org/10.1007/s13524-015-0415-0>
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space–time variation in disease risk. *Statistics in Medicine*, 14(21–22), 2433–2443. <https://doi.org/10.1002/sim.4780142112>
- Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with r-inla. *Spatial and Spatio-Temporal Epidemiology*, 4, 33–49. <https://doi.org/10.1016/j.sste.2012.12.001>
- Gómez-Rubio, V. (2020). *Bayesian inference with inla*. Chapman; Hall/CRC. <https://doi.org/10.1201/9781315175584>
- Jeong, H., Lin, Y., & Lee, L.-f. (2023). Estimation of spatial autoregressive models for origin–destination flows: A partial likelihood approach. *Economics Letters*, 229, 111202. <https://doi.org/https://doi.org/10.1016/j.econlet.2023.111202>
- Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* (3rd). CRC Press. <https://doi.org/10.1201/9780429444876>
- MacNab, Y.-C. (2022). Bayesian disease mapping: Past, present, and future. *Spatial and Spatio-Temporal Epidemiology*, 43, 100526. <https://doi.org/10.1016/j.sste.2022.100526>
- Moraga, P. (2019). *Geospatial health data: Modeling and visualization with r - inla and shiny*. Chapman & Hall / CRC.

- Napier, G., Lee, D., Robertson, C., Lawson, A., & Pollock, K. G. (2016). A model to estimate the impact of changes in mmr vaccine uptake on inequalities in measles susceptibility in scotland. *Statistical Methods in Medical Research*, 25(4), 1185–1200. <https://doi.org/10.1177/0962280216660420>
- Poot, J., Alimi, O., Cameron, M. P., & Maré, D. C. (2016). The gravity model of migration: The successful comeback of an ageing superstar in regional science. *Investigaciones Regionales-Journal of Regional Research*, (36), 63–86.
- Pu, Y., Chi, G., Zhao, X., Zhao, J., & Kong, F. (2019). A spatial dynamic panel approach to modelling the space-time dynamics of interprovincial migration flows in China. *Demographic Research*, 41(31), 913–948. <https://doi.org/10.4054/DemRes.2019.41.31>
- Rue, H., Lindgren, F., & Krainski, E. T. (2024). *Inla: Full bayesian analysis of latent gaussian models using integrated nested laplace approximations*. Chapman; Hall/CRC. <https://doi.org/10.1201/9781003353361>
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations (inla). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Saa, I., Novak, M., Morales, A., & Pentland, A. (2020). Looking for a better future: Modeling migrant mobility. *Applied Network Science*, 5. <https://doi.org/10.1007/s41109-020-00308-9>
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1), 1–28. <https://doi.org/10.1214/16-STS576>
- Sun, N., Bursac, Z., Dryden, I., Lucchini, R., Dabo-Niang, S., & Ibrahimou, B. (2023). Bayesian spatiotemporal modelling for disease mapping: An application to preeclampsia and gestational diabetes in florida, united states. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-023-29953-0>
- Van Der Gaag, N., & Van Wissen, L. (2001). Determinants of the subnational distribution of immigration. *Tijdschrift voor Economische en Sociale Geografie*, 92(1), 27–41. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9663.00137>
- 유희연. (2023). 우리나라 저숙련 이주노동자 지역분포 현황과 특성 (Working Paper). 국토연구원. https://www.krihs.re.kr/gallery.es?mid=a10103090000&bid=0025&list_no=29993&act=view
- 이규용. (2024). 산업 및 직종별 인력수급전망과 외국인력 수요연구 (협동연구보고서). 경제·인문사회연구회, 한국노동연구원. https://www.nrc.re.kr/board.es?mid=a10301000000&bid=0008&act=view&otp_id=OTP_00000000000014619
- 정기선. (2025). 이민정책, 감(感) 아닌 데이터로...지역사회 통계가 부족하다 [Accessed: 2025-10-08]. 교수신문. <https://www.kyosu.net/news/articleView.html?idxno=142499>
- 최종원. (2025). 지역특화형 비자로 이주노동자 늘리면 지역 경기 살아날까? [Accessed: 2025-10-08]. 뉴스포스트. <https://www.newspost.kr/news/articleView.html?idxno=217277>