

I. 연구 배경 및 목적

유례없는 저출산과 고령화로 인해 생산가능인구가 급감하면서, 외국인 근로자는 이제 대한민국 산업 생태계의 필수 불가결한 구성원이 되었다. 특히 정부는 내국인이 기피하는 경향이 높은 중소기업 및 농어촌 현장의 인력난을 해소하고자 2004 년부터 고용허가제¹(Employment Permit System, EPS)를 도입하여 비전문취업 비자(E-9)를 취득한 외국인 인력을 체계적으로 수급·관리해 왔다. 현재 E-9 비자를 취득하여 체류하고 있는 외국인은 33 만 명이며, 이는 전체 29 개 체류 자격 중 가장 높은 비중을 차지한다. 이들은 태국, 캄보디아, 네팔 등 17 개국 출신이며, 제조업, 건설업, 서비스업, 어업, 농·축산업, 임업, 광업의 7 개 분야에 고용되어 일하고 있다(고용노동부·한국산업인력공단, 2024).

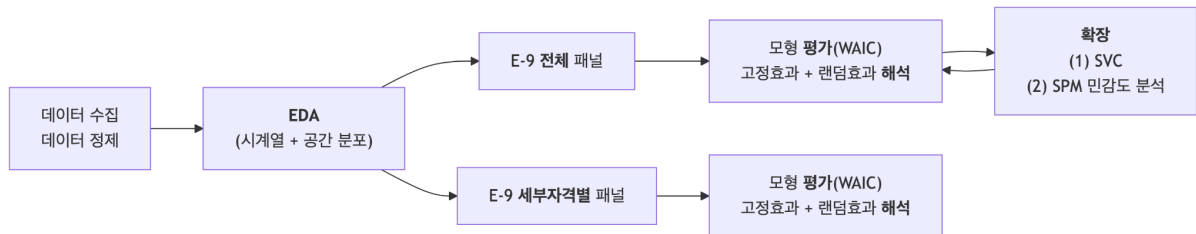
과거의 외국인 근로자 정책이 국가 차원의 총량적 인력 수급에 초점을 맞추었다면, 현재는 '어느 지역에 왜 머무는가'라는 지역적 분포를 파악할 필요성이 커지고 있다. 인구 감소의 충격은 지역, 산업, 연령에 따라 차별적으로 나타나기 때문에(조영태, 2021), 외국인 인력 정책 또한 지역별·산업별 특수성을 고려한 맞춤형 전략으로 전환되어야 한다. 하지만 현장의 요구에도 불구하고, 지자체별 외국인 수급 및 관리를 뒷받침할 실증적 데이터와 연구는 여전히 부족한 실정이다(정기선, 2025).

이러한 배경에서 본 연구는 2013 년부터 2021 년에 걸친 E-9 외국인 근로자 패널 자료와 시군구별 인구·경제 데이터를 결합하여, E-9 외국인 근로자 분포의 결정요인을 시공간적으로 분석하고자 한다. 특히 시공간적 자기상관과 공간적 이질성을 모형에 반영할 수 있으며, 소지역 단위에서의 추정에서 강점이 있는 시공간 베이지안 계층 모형(Spatiotemporal Bayesian Hierarchical Model)을 활용한다. 나아가 E-9 의 업종별 세부 자격에 따른 결정 요인 역시 살펴볼 것이다.

본 연구는 이러한 방법론적 시도를 통해 E-9 외국인 근로자 패널 자료의 분석적 가치를 제고하고, 인구 구조 변화의 지역적 격차 및 다층성이 심화되는 현시점에서 지자체별로 지역 특성에 부합하는 외국인 인력 정책을 수립하는 데 필요한 실증적인 근거를 제공하고자 한다. 이를 위한 구체적인 연구 질문은 다음과 같다:

¹ 고용허가제는 E-9 비자를 대상으로 하는 '일반 고용허가제'와 H-2 비자를 대상으로 하는 '특례 고용허가제'로 나뉜다. H-2 비자는 중국이나 중앙아시아의 '외국 국적 동포'를 대상으로 하며, 고용 허용 업종에서 약간의 차이가 존재한다. 2025 년 기준으로는 약 8 만 명의 인원이 체류 중이다. 이들의 분포는 공간적으로 차이를 보이며(이현욱, 2015) 본 연구에서는 E-9 비자, 즉 일반 고용허가제 등록 인원만을 분석 대상으로 하였다.

- (RQ1) 시군구별 E-9 외국인 근로자 분포를 결정하는 주요 요인은 무엇인가?
- (RQ2) 주요 결정 요인이 외국인 근로자 분포에 미치는 영향은 지역에 따라 어떻게 달라지는가?
- (RQ3) 세부 자격별 고용허가제 외국인 근로자 분포에 유의한 영향을 미치는 요인과 그 지역적 이질성은 차별적으로 나타나는가?



[그림 1] 연구 흐름도

이후 연구 흐름은 다음과 같다: II 장에서는 고용허가제 외국인 근로자의 분포 및 결정 요인을 분석한 선행 연구와 방법론적 연관성을 지닌 선행 연구를 고찰한다. III 장에서는 데이터 구축 과정 및 모형화 과정을 설명한다. IV 장에서는 E-9 전체 패널 데이터에 대한 분석으로, 모형 비교와 전체 모형에서의 고정·랜덤 효과 추정치 해석을 제시한다. 또한 확장된 모형으로 공간 가변계수모형 적합 결과 및 상이한 공간근접성행렬을 이용했을 때의 적합 결과를 제시한다. V 장에서는 E-9 세부 자격별 패널 데이터에 대한 분석을 동일한 흐름으로 진행한 결과를 보고하며, 확장된 모형으로 자격별 통합 모형 적합 결과를 제시한다. VI 장에서는 분석 결과를 요약하고 이를 바탕으로 정책적 시사점을 제안한다. 연구 흐름도는 [그림 1]과 같다.

II. 선행 연구 분석

고용허가제 외국인 근로자의 분포와 관련된 연구로, 이현욱(2015)은 고용허가제 등록 인원의 공간 분포 특성을 체류 자격(E-9, H-2)과 출신 국가별로 나누어 분석한 바 있다. 제도적 틀과 주요 도시별 산업 구조의 특징과 관련하여 지역별로 상이한 분포 패턴이 형성되었음을 밝혔다. 또한, 유희연(2023)은 E-9를 비롯한 저숙련 이주노동자의 지역 분포 현황을 기술적으로 분석하며, 지역별 정책 수립의 중요성을 시사하였다. 이러한 접근은 E-9의 공간적 분포를 지역적 맥락과 연계하여 설명하고자 하였다는 점에서 의의가 있다. 다만 주로 분포의 '현상'을 기술하는 데 초점을 맞추고 있어, 언급한 요인들이 통계적으로 어느 정도의 연관성을 갖는지를 검증하기는 어렵다는 한계점을 노정한다.

외국인 근로자 수 결정 요인을 분석한 국내 연구 사례로, 이찬영(2018)은 특별시·광역시를 제외한 시군 지역의 패널 자료를 이용하여 지역 노동시장 내부에서 내·외국인 고용관계에 초점을 맞춘 실증적 분석을 진행하였다. 통제 변수로 GRDP, 고정투자액, 임금, 산업별 비중 등을 활용하였으며,

표본을 시 지역과 군 지역으로 분리하거나 연령대별로 나누어 추정하였을 때 내·외국인 근로자 수의 보완·대체 관계가 이질적으로 나타남을 밝혔다. 또한 조은지·이찬영(2021)은 유사한 연구 질문에 대해 내생성 통제를 위한 도구변수법을 이용하여 추정 결과가 단순 패널 모형 결과와 달라질 수 있음을 보였다. 이는 외국인 인력의 공간적 분포가 내생적으로 형성된 과정임을 시사하지만, 지역 간 인접성과 시계열 추세를 반영하지 못했다는 점에서 한계가 남는다. 한편 이원규·이재우(2025)는 2020 년 외국인 근로자 실태조사를 바탕으로 구조방정식 모형을 이용하여 주거 만족도와 사회적 관계망이 이주 의향에 영향을 미치는지에 대해 논의한 바 있다. 그러나 이 역시도 시공간적 스케일을 통합하여 설명하지는 못했다는 한계가 존재한다.

본 연구는 E-9 외국인 근로자의 공간적 분포와 그 요인을 설명하고자 하는 점에서 선행 연구의 문제의식을 계승한다. 다만, 구축된 패널 자료가 일부 지역에서 관측치가 낮은 희소성을 지니고, 공간적 인접성과 시간적 연속성을 강하게 나타낼 가능성이 높은 자료라는 점에 주목한다. 이에 본 연구는 지역별 고용·산업·경제 변수를 결합한 패널 자료를 구축한 뒤, 공간 효과와 시간 효과를 동시에 반영하는 시공간 베이지안 계층 모형을 적용하여 희소 지역에서의 추정 불안정성을 완화하고 공간적 군집과 시간적 추세를 분리해 해석할 수 있도록 한다는 점에서 차별점을 지닌다. 특히 최근의 인구학 분야 연구에서 계층적 베이지안 접근이 활용되는 경우가 증가하고 있고(Mazzocchi-Greco, 2011; Bradley et al., 2018; Paglino, 2024), 해당 모형의 설명력이 비공간모형에 비해 모형 적합도 측면에서 개선을 보인다고 보고되고 있다는 점에서, 본 연구의 전략은 E-9 외국인 근로자 분포의 결정요인을 보다 엄밀하게 규명하는 데 기여할 수 있을 것이다.

III. 데이터 및 방법론

1. 데이터

본 연구의 분석 대상 기간은 설명변수의 구축 가능 여부를 고려하여 2013 년부터 2021 년으로 설정하였다. 시간 해상도는 연 단위로 하며, 분기별/반기별 자료는 해당 연도의 마지막 시점의 관측치를 해당 연도의 값으로 지정한다. 연구의 지리적 스케일(공간 범위)은 2024 년 행정구역 구획을 기준으로 세종특별자치시를 제외한 전국으로 하며 측정 스케일(공간 단위)은 비자치구를 고려하지 않은 228 개 시군구로 한다.

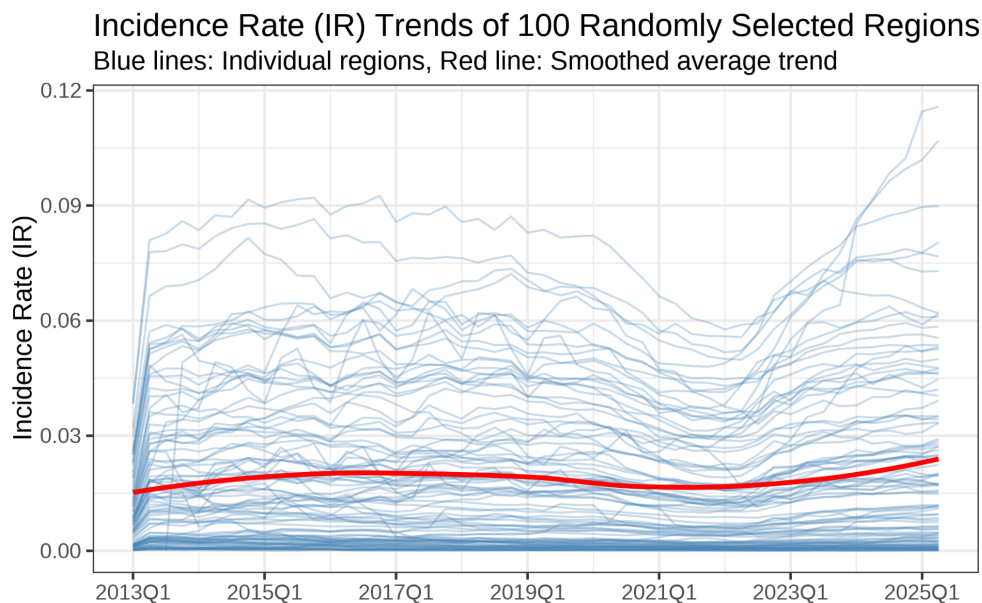
종속변수는 시군구별 E-9 혹은 E-9 세부 자격별 외국인 근로자 수(해당 시점에 거주하는 스톡(stock) 인원)이며, 데이터는 법무부 출입국·외국인정책본부의 통계월보에서 내려받았다. 설명변수는 크게 인구, 경제, 산업, 고용, 코로나-19, 지역별(수도권/비수도권, 시/군/구) 더미 변수로 구성되어 있다. 전체 변수별 데이터 원천 및 시공간 해상도는 [표 1]에 정리하였다. 다중공선성의 가능성을 탐지하는 분산팽창계수(VIF) 계산 결과 모든 변수가 5 이하의 값을 가져 분석에 적합함을 확인하였다.

구분	변수명	자료 출처	시간 해상도	공간 해상도
종속변수	E-9 등록 인원	법무부, 「통계월보」	분기별	시군구-2
인구	성비	행정안전부, 「주민등록인구현황」	월별	시군구-2
	고령인구 수	행정안전부, 「주민등록인구현황」	월별	시군구-2
	인구 이동 유효도 지수	국가데이터처, 「국내인구이동통계」	월별	시군구
경제	무역수지	관세청, 「수출입무역통계」	월별	시군구
	GRDP 총액	서울특별시, 「서울특별시지 역내총생산」 등 시·도별 통계	연도별	시군구
산업	전국 대비 1 차산업 GRDP 비중 편차	서울특별시, 「서울특별시지 역내총생산」 등 시·도별 통계	연도별	시군구
	전국 대비 2 차산업 GRDP 비중 편차	서울특별시, 「서울특별시지 역내총생산」 등 시·도별 통계	연도별	시군구
	산업단지 총면적	한국산업단지공단, 「전국산업단지 현황통계」	연도별	시군구
	제조업 종사자 비율	고용노동부, 「사업체노동력조사」	연도별	시군구
	건설업 종사자 비율	고용노동부, 「사업체노동력조사」	연도별	시군구
	도소매업 종사자 비율	고용노동부, 「사업체노동력조사」	연도별	시군구
고용	고용률 (15-64 세)	국가데이터처, 「지역별고용조사」	반기별	시군구
	실업률	국가데이터처, 「지역별고용조사」	반기별	시군구
시점별 더미	코로나-19 시기	2020-2021 년 = 1	시점별	시군구
지역별 더미	수도권 여부	서울특별시, 인천광역시, 경기도 = 1	시점별	시군구
	시/구 지역 여부	-시/-구 = 1	시점별	시군구
	산업단지 소재 여부	한국산업단지공단, 「전국산업단지 현황통계」, 산업단지 소재 = 1	시점별	시군구

[표 1] 설명변수 자료 출처 및 시공간 해상도

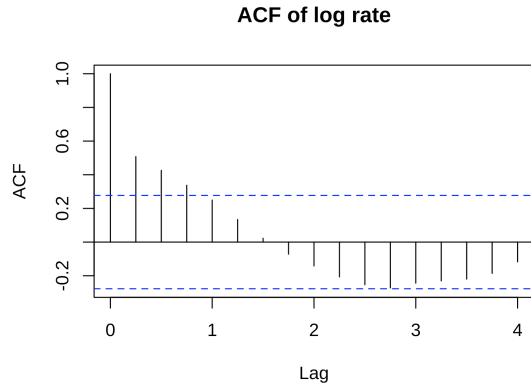
데이터 정제 과정에서 특기할 만한 사항으로, 인구 이동 유효도 지수(migration effectiveness index)는 (순 이동)/(총이동)으로 계산하였으며, 전국 대비 1:2:3 차 산업 GRDP 비중 편차는 세부 산업을 분류한 뒤 각 지역별로 GRDP 가 차지하는 비중을 계산하고 해당 시점 전국 평균 비중을 뺀 값을 이용하였다. 코로나-19 시기를 나타내는 더미는 2020 년과 2021 년에 대해 1 의 값을 부여하는 방식으로 부여하였다. 또한 데이터를 통합하는 과정에서 2024 년 기준 행정구역으로 경계 변동 등을 최신화하였으며, 과거 자료에서 명칭 변동이 발생한 경우는 현재 기준으로 매핑하여 데이터가 연결되도록 하였다.

2. 탐색적 데이터 분석

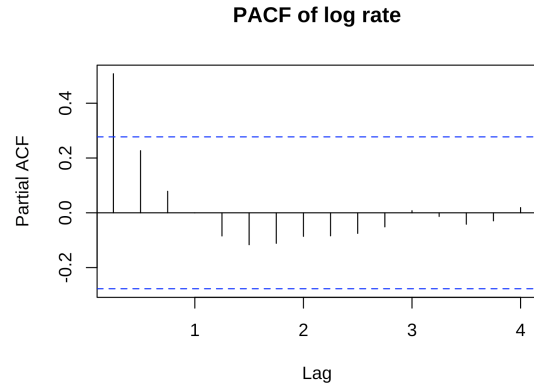


[그림 2] 100 개 지역 및 전국 평균의 E-9 등록 인원 발생률

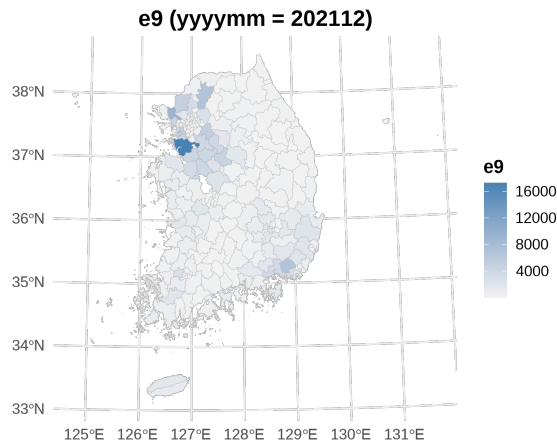
먼저 [그림 2]는 랜덤으로 추출한 100 개 시군구 및 전국 평균의 E-9 발생률(incidence rate)의 추이를 시각화한 것이다. 2013 년 이후로 증가한 뒤 정체기를 가지다가 코로나-19 발생으로 인해 일시적으로 급감한 것을 확인할 수 있었다. 2022 년 이후로는 기존 발생률을 회복한 뒤 지속적으로 증가하고 있는 패턴이 관측되었다. 또한 전국 단위 발생률에 로그를 취한 시계열의 자기상관함수(ACF)와 부분자기상관함수(PACF)를 확인한 결과([그림 3]와 [그림 4]), ACF 가 지속적으로 감소하고 PACF 는 시차 1 이후로 절단(cutoff)되는 AR(1) 과정이 지배적인 패턴으로 나타나 이를 모형 구조에 반영하였다.



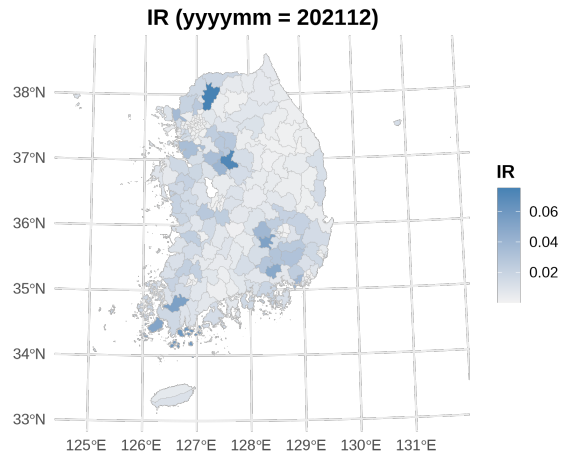
[그림 3] 로그-발생률의 자기상관함수



[그림 4] 로그-발생률의 부분자기상관함수



[그림 5] 2021 년 4 분기 E-9 등록 인원 분포



[그림 6] 2021 년 4 분기 E-9 발생률 분포

다음으로 [그림 5]와 [그림 6]은 각각 2021 년 4 분기 E-9 등록 인원과 발생률의 공간적 분포를 나타낸 것이다. 절대적인 인원수는 수도권 산업도시에서 높게 나타나는 경향을 보였다. 특히 경기도 화성시에서 가장 많은 E-9 외국인 근로자가 분포하고 있었는데 이는 화성시 경제의 큰 축인 대기업을 중심으로 중소기업 및 공장들이 수천 개 자리 잡고 있기 때문으로 파악된다(유희연, 2023). 한편, 같은 분기의 전 산업 종사자 수 대비 비율로 분석한 경우 상이한 결과가 나타났다. 가장 높은 비율이 관측된 지역은 충청북도 음성군(7.55%)이었는데, 이는 대풍지방산업단지, 음성이하이텍산업단지 등 산업단지와 농공단지가 조성되어 있고, 전체 인구수가 수도권 대비 낮기 때문에 1 위를 기록한 것으로 보인다. 추가적으로, 공간적 자기상관을 확인하기 위한 모런 검정을 실시한 결과 유의한 양의 공간적 자기상관이 전 기간에서 관측되었다.

3. 방법론: 시공간 베이지안 계층 모형

공간(시군구) 인덱스를 $i = 1, 2, \dots, N$, 시간(연도) 인덱스를 $t = 1, 2, \dots, T$ 로 둔다. 종속변수 y_{it} 는 시군구 i 의 시점 t 에서 관측된 E-9 등록 인원의 저량(stock)이며, 노출량 E_{it} 는 시군구 i 의 시점 t 에서 관측된 전산업 종사자 수로 설정하였다. 카운트 자료와 노출량을 결합한 계층적 모델링은 데이터, 프로세스, 하이퍼파라미터의 세 가지 계층으로 나뉘어 구성되며(Banerjee et al., 2014; Biangiardo-Cameletti, 2015) 본 연구에서의 구체적인 형태는 다음과 같다.

먼저 데이터 계층에서 E-9 등록 인원은 카운트 자료이므로 포아송 분포를 따른다고 가정한다.

$$y_{it} \mid \lambda_{it} \sim \text{Poisson}(E_{it} \lambda_{it})$$

여기서 λ_{it} 는 전산업 종사자 1 인당 기대되는 E-9 등록 인원의 발생률(intensity)을 의미하게 된다.

프로세스 계층에서 λ_{it} 는 로그 연결 함수를 이용해 공변량과 시공간 랜덤 효과를 이용하여 다음과 같이 모델링한다.

$$\log \lambda_{it} = \alpha + \mathbf{x}_{it}^T \boldsymbol{\beta} + s_i + u_{it} + \eta_t$$

α 는 절편, $\mathbf{x}_{it}$ 는 지역의 특성을 나타내는 설명변수 벡터, $\boldsymbol{\beta}$ 는 전국 공통의 고정 효과 계수이다. s_i 는 시간에 관계없이 지속되는 공간 패턴을 포착하는 공간 랜덤효과이며, u_{it} 는 시점에 따라 변화하는 지역별 변동(즉, 시공간 상호작용)을 나타내는 시공간 랜덤효과, η_t 는 모든 지역에 공통으로 작용하는 시간 추세를 나타내는 시간 랜덤효과이다.

하이퍼파라미터(초모수) 계층에서 랜덤효과를 나타내는 세 가지 항은 잠재 가우시안 필드(latent Gaussian field)를 따르는 것을 가정하여 정의된다. 먼저 전체 공간 추세 s_i 는 공간근접성행렬(spatial proximity matrix, SPM) W 에 기반한 BYM2(Besag-York-Mollié 2) 모형으로 정의하며, 그 형태는 다음과 같다:

$$s_i \sim \text{BYM2}(W, \tau_s, \phi_s),$$

$$s_i = \frac{1}{\sqrt{\tau_s}} \left(\sqrt{\phi_s} u_i^* + \sqrt{1 - \phi_s} v_i^* \right), \quad 0 \leq \phi \leq 1$$

여기서 u_i^* 는 ICAR(intrinsic conditional autoregressive) 성분으로, 인접할수록 비슷한 패턴을 가지도록 하는 구조화된(structured) 공간효과 성분이며, $v_i^* \sim \mathcal{N}(0, 1)$ 은 비구조화 공간효과 성분이다. τ_s 는

전체 변동의 크기를 제어한다. ϕ_s 는 공간 구조화 성분이 차지하는 분산 비중을 의미하므로 ϕ_s 가 클수록 s_i 는 인접성을 반영한 평활화된 공간 패턴을, 작을수록 지역 고유의 노이즈 성분을 더 강하게 반영한다(BYM2 모형에 관한 자세한 설명은 Riebler et al., 2016 참고). 본 연구에서는 W 으로 기본적으로 1 차 퀸(queen) 연접성행렬을 사용하였으며, 다른 SPM 을 사용하였을 때의 결과는 IV 장에서 살펴본다. 다음으로, 시공간 랜덤효과 u_{it} 는 시간에 따라 점진적으로 변화하는 공간 성분을 가정하며, 본 연구의 경우 다음과 같은 AR(1) 시간 구조에 공간 효과가 반영된 오차항을 지니는 형태로 모델링한다.

$$u_{it} = \rho_u u_{it-1} + \epsilon_{it}, \quad |\rho_u| < 1, \\ \epsilon_{it} \sim \text{BYM2}(W, \tau_u, \phi_u)$$

이때, ϵ_{it} 는 시점별 공간적 의존성을, ρ_u 는 지역별 효과의 시간적 의존성을 반영하는 파라미터이다. 마지막으로, 전 지역 공통의 시간 랜덤효과 η_t 는 AR(1) 과정으로 가정하여 시간적 의존성을 모델링한다.

$$\eta_t = \rho_\eta \eta_{t-1} + \xi_t, \quad |\rho_\eta| < 1 \\ \xi_t \sim \mathcal{N}(0, \tau_\eta^{-1})$$

각각의 하이퍼파라미터에는 복잡도-벌점(penalized-complexity) 사전분포를 부여하였다.² 이는 단순한 기준모형으로부터의 이탈을 벌점화하여 과적합을 억제하는 방식으로 사후분포를 INLA 방식으로 계산할 시 표준적으로 사용된다(Simpson et al., 2017). INLA(integrated nested Laplace approximation)는 잠재 가우시안 모형 계열에서 라플라스 근사를 이용하여 빠르게 사후근사를 수행하는 장점이 있으며, 전통적인 마코프 체인 몬테 카를로 방식에 비해 연산 속도가 빨라 복잡한 시공간 모형에서 널리 활용된다. 본 연구에서는 R의 INLA 패키지를 통해 계산을 수행하였다.

추가적으로, 관측지별로 고정 효과가 이질적이라는 공간 가변계수(spatially varying coefficient, SVC) 가정을 추가할 경우, 계수 β_p 가 공간적으로 변하는 함수인 모형으로 확장된다. $\mathcal{P}$ 를 공간 가변계수로 둘 변수의 집합이라고 하면,

$$\log \lambda_{it} = \beta_0 + \sum_{p \notin \mathcal{P}} x_{it,p} \beta_p + \sum_{p \in \mathcal{P}} x_{it,p} \beta_p(i) + (\text{other random effects}) \\ \beta_p(i) = \bar{\beta}_p + w_{pi}$$

와 같은 형태가 되고, 여기서 w_{pi} 는 공간 랜덤 효과로 BYM2 모형을 따른다고 가정한다.

² R 에서 초기값으로 $\tau_u = \text{pc.prec}(1, 0.01)$, $\phi_u = \text{pc}(0.5, 0.5)$, $\tau_\eta = \text{pc.prec}(1, 0.01)$, $\rho_\eta = \text{pc}(0.7, 0.5)$ 로 설정하였다.

IV. E-9 외국인 근로자 패널 자료 분석

1. 분석 결과

1) 모형 비교

먼저 시공간 베이지안 모형의 설명력을 확인하고자 WAIC 를 이용한 모형 비교 절차를 거친다. WAIC(Watanabe-Akaike Information Criterion)는 사후 예측 성능을 근사하는 베이지안 정보 기준으로, 값이 작을수록 더 우수한 모형으로 평가되어 모형 비교에 널리 이용된다(Watanabe, 2010). 모형 비교에는 (1) AR(1) 구조의 시간 랜덤효과 η_t 만을 추가한 모형, (2) 시간 랜덤효과와 설명변수 벡터 x_{it} 를 추가한 모형, (3) 시간 랜덤효과와 BYM2 구조의 공간 랜덤효과 s_i 를 추가한 모형, (4) 시간, 공간 랜덤효과와 더불어 시공간 상호작용 랜덤효과 u_{it} 를 추가한 모형, (5) III 절에서 명시한 바와 같이 랜덤효과와 설명변수 벡터를 포함한 전체 모형(full model)이 이용되었다.

	모형 상세	WAIC
모형 (1)	AR(1)	130211833.73
모형 (2)	AR(1) + 공변량	16071504.90
모형 (3)	BYM2 + AR(1)	98725.18
모형 (4)	BYM2 + AR(1) + 상호작용	19587.32
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	19576.76

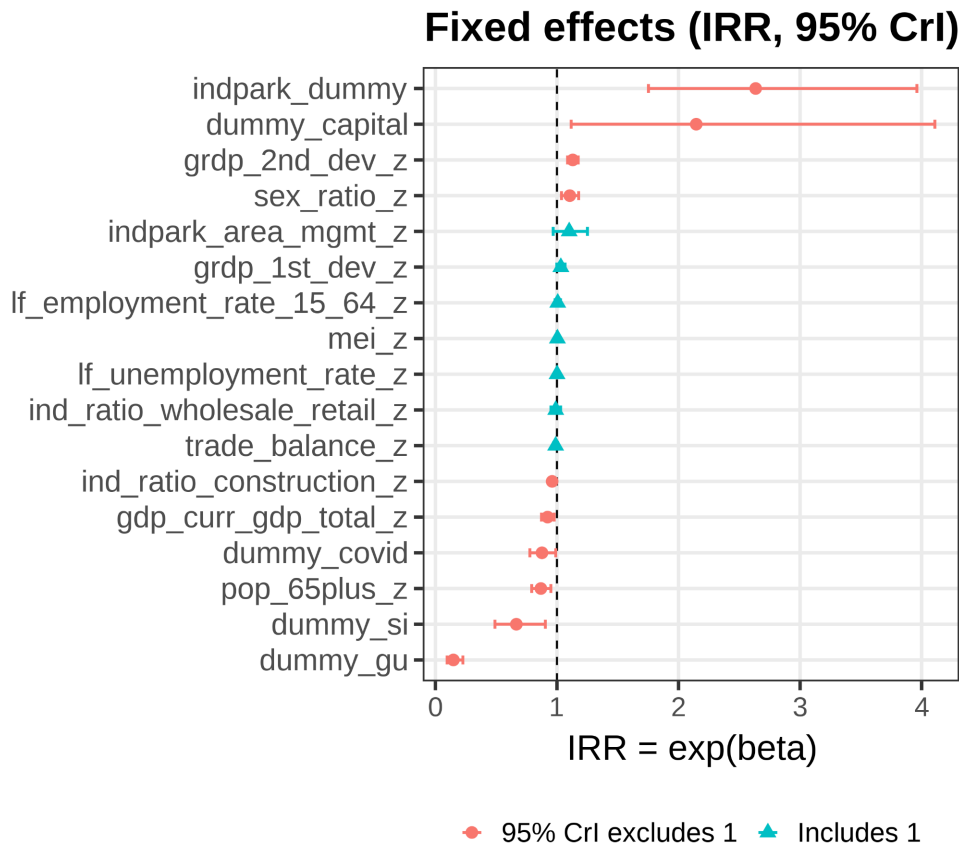
[표 3] WAIC 를 이용한 모형 비교 결과 (E-9 전체, 연도별)

[표 3]은 WAIC 를 이용한 모형 비교 결과를 나타낸다. 우선 단순히 시계열 구조만 넣은 모형 (1)과 이에 공변량을 추가한 모형 (2)에 비해 시간-공간 랜덤효과 항이 가법적(additive)으로 추가된 모형 (3)에서 WAIC 가 비약적으로 감소하는 것을 확인할 수 있었다. 이후 적합도의 증가는 시공간 랜덤효과 항을 추가한 모형 (4)에서 다시 한번 나타났으며 모형 (4)와 이에 공변량을 추가한 모형 (5)의 적합도 차이는 크지 않고 오히려 소폭 증가하였다. 이는 공변량이 설명하는 변동이 이미 랜덤효과에 흡수된 공간 교란(spatial confounding)의 결과일 가능성이 있다. 다만, 본 연구에서는 해석의 측면에서 시공간 요인을 배제한 변수별 설명력 또한 관심의 대상이므로, 최종적으로 적합도와 설명력 면에서 우수한 것으로 나타난 모형 (5)을 분석에 이용하는 것이 타당성을 지닌다고 볼 수 있다.

2) 고정 효과 추정 결과

변수명	변수 코드	평균	IRR	IRR 신용구간	원 변수 표준편차
성비	sex_ratio_z	0.1	1.105	(1.037, 1.178)	5.443
고령인구 수 (명)	pop_65plus_z	-0.142	0.868	(0.792, 0.949)	24438.197
인구 이동 유효도 지수	mei_z	0.004	1.004	(0.997, 1.010)	0.088
무역수지 (천 불)	trade_balance_z	-0.011	0.989	(0.973, 1.005)	496118.00 2
GRDP 총액 (십만 원)	gdp_curr_gdp_total_z	-0.08	0.923	(0.872, 0.976)	10224960. 2
전국 대비 1 차산업 GRDP 비중 편차	grdp_1st_dev_z	0.031	1.031	(0.998, 1.066)	0.084
전국 대비 2 차산업 GRDP 비중 편차	grdp_2nd_dev_z	0.122	1.130	(1.088, 1.175)	0.181
산업단지 총면적 (천 m²)	indpark_area_mgmt_z	0.096	1.101	(0.969, 1.251)	9539.906
건설업 종사자 비율	ind_ratio_construction_z	-0.041	0.960	(0.940, 0.981)	0.039
도소매업 종사자 비율	ind_ratio_wholesale_retail_z	-0.01	0.990	(0.950, 1.030)	0.054
고용률 (15- 64 세)	lf_employment_rate_15_64_z	0.006	1.006	(0.984, 1.029)	0.54
실업률	lf_unemployment_rate_z	0.001	1.001	(0.988, 1.014)	0.014
코로나-19 시기	dummy_covid	-0.132	0.876	(0.777, 0.990)	0.416
수도권 여부	dummy_capital	0.763	2.145	(1.117, 4.108)	0.454
시 지역 여부	dummy_si	-0.407	0.666	(0.489, 0.905)	0.473
구 지역 여부	dummy_gu	-1.912	0.148	(0.097, 0.226)	0.46
산업단지 소재 여부	indpark_dummy	0.969	2.635	(1.752, 3.959)	0.405

[표 4] 고정 효과 계수 추정 결과 (E-9 전체, 연도별)



[그림 7] 고정 효과 IRR 추정 결과 및 95% 신용구간 (E-9 전체, 연도별)

모형 (5)를 이용하여 고정 효과를 추정한 결과는 [표 4]에 나타나 있다. 한편, 본 연구의 모형에서는 모든 연속형 공변량을 표준화하였고, 로그 연결 함수를 이용하였기에 회귀계수 β_p 를 승법적 효과, 즉, 연속형 공변량이 1 표준편차 증가할 때마다 기대 발생률이 $IRR = \exp(\beta_p)$ 배 증가하는 것으로 해석해야 함을 주의해야 한다. 또한, 더미변수의 경우 해당 변수가 1 인 집단의 기대 발생률이 기준집단 대비 $IRR = \exp(\beta_p)$ 배가 되는 것으로 해석해야 한다. [그림 7]을 통해 IRR 의 95% 신용구간이 1 을 포함하지 않는 변수들, 즉 유의미한 영향력이 있는 것으로 나타난 변수들을 확인할 수 있다.

모형 추정 결과, 수도권 여부, 전국 대비 2 차산업 GRDP 비중 편차, 성비, 산업단지 소재 여부가 전산업 종사자 규모 대비 E-9 비율을 유의하게 증가시키는 것으로 나타났다. 구체적으로, 수도권은 비수도권 대비 E-9 비율이 약 2.13 배 높은 것으로 해석된다. 또한 전국 대비 2 차산업 비중 편차는 1 표준편차(=18%p) 증가할 때마다 E-9 비율이 약 13% 증가하는 경향이 확인되었으며, 성비는 1 표준편차(=5.442) 증가할 때마다 E-9 비율이 약 11% 증가하는 경향이 확인되었다. 가장 큰 효과를 드러낸 산업단지 소재 여부는 비존재 지역 대비 E-9 비율이 약 2.64 배 높은 것으로 추정되었다.

양(+)의 영향을 미치는 요인들을 살펴보면, E-9 고용이 제조업·공업 중심의 지역 산업구조와 결부되어 있음을 뚜렷하게 보여준다. 산업단지 소재 여부 역시 지역의 산업구조가 유의미한 설명변수임을 확인시켜 주는 경향이 나타났다. 한편, 성비의 경우 비숙련 노동 분야에서 남성 비중이 높기 때문에 양(+)의 영향력이 나타난 것으로 파악된다. 또한, 수도권 여부 더미의 효과는 다른 변수들이 통제되었을 경우, 수도권에서의 거주를 선호하는 경향이 뚜렷하다는 사실을 보여준다.

반면 고령인구, GRDP 총액, 건설업 종사자 비율, 코로나-19, 그리고 시/구 지역 여부는 E-9 비율을 유의하게 감소시키는 경향이 있는 것으로 추정되었다. 먼저 고령인구는 1 표준편차(=24,438 명) 증가할수록 E-9 비율이 13% 감소하는 것으로 나타났으며, GRDP 총액 역시 1 표준편차(=1 조 224 억 원) 증가할수록 E-9 비율이 8%가량 감소하는 결과를 보였다. 산업구성 측면에서는 건설업 종사자 비율이 음(-)의 효과를 보여 1 표준편차(=3.9%p) 증가할 때 E-9 비율이 4% 감소하는 것으로 나타났으며, 코로나-19 시기에는 기존에 비해 12%가량 E-9 비율이 감소하는 것으로 나타났다. 또한 시·구 지역의 경우 기준 지역인 군 지역에 비해 E-9 비율이 각각 33%, 85%가량 감소하는 패턴이 관찰되었다.

음(-)의 영향을 미치는 요인을 세부적으로 살펴보면, 고령인구가 많이 분포한 지역은 대도시 혹은 농촌으로 복합적인 요인으로 E-9 노동력에 대한 수요가 상대적으로 적을 수 있다는 점을 시사한다. 또한, GRDP 총액을 기준으로 하였을 경우 음(-)의 영향이 나타난 것은 경제 규모가 큰 지역은 서비스업·고숙련 인력의 비중이 높을 가능성이 높고, 그에 따라 E-9의 고용 비중이 상대적으로 적은 것으로 해석할 수 있다. 건설업 종사자 비중의 경우 E-9가 아니라 H-2 등 다른 체류 자격과 연관되어 있을 가능성을 시사하며, 코로나-19 더미의 경우 2020 년과 2021 년 당시 외국 인력 입국이 제한적이었던 상황을 반영한다. 지역적 측면에서는 산업단지 및 농공단지가 주로 군 지역에 위치하여 시나 구 지역은 발생률에 대해 음(-)의 관계가 나타난 것으로 볼 수 있다.

이 밖의 변수들은 95% 사후 신용구간이 1 을 포함하여 유의한 영향력을 갖지 못하는 것으로 나타났다. 특히 선행 연구에서 내국인의 고용과 관련되어 언급되었던 고용률이나 실업률 관련 지표는 유의한 영향력을 드러내지 못했다.

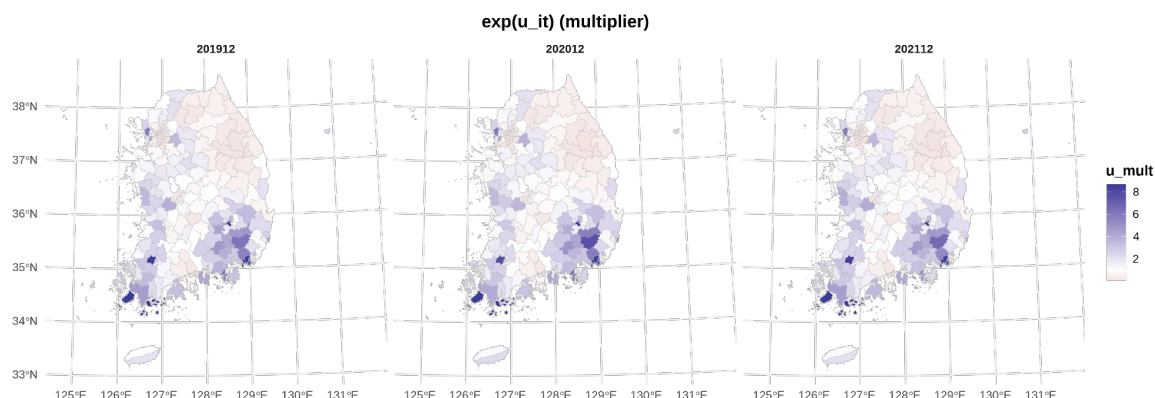
3) 랜덤효과 추정 결과

[표 4]는 랜덤효과의 하이퍼파라미터 추정 결과를 나타낸다. τ 는 랜덤효과 변동성(분산)의 역수에 이며, ϕ 는 공간 구조화 성분의 비율, ρ 는 AR(1) 자기상관계수를 나타낸다. 이로부터 알 수 있는 주요 사실은 다음과 같다: 첫째, 변동성의 정도를 살펴보았을 때 시공간 랜덤효과의 표준편차(= $1/\sqrt{\tau_u} \simeq 0.953$)가 가장 큰 것으로 나타났다. 이는 특정 지역과 특정 분기에 나타나는 추가 변동이 상당하다는 의미이며, 이는 앞선 모형 비교에서 시공간 상호작용 항을 추가하였을 때 WAIC 값이 크게 감소한 결과와도 정합한다. 둘째, 시공간 랜덤효과의 AR(1) 상관계수가 1 에 근접한 값으로 공통 추세로

설명되지 않는 지역적 상호작용의 효과가 경로(trajecotory)를 따라 지속적으로 이어지는 것으로 나타났다. 이는 앞서 공변량을 포함하더라도 WAIC 개선이 제한적이었던 이유로 해석된다. 셋째, 시공간 랜덤효과에서 공간적으로 구조화된 공분산의 비율은 0.475, 즉 거의 절반으로 나타나 구조화-비구조화 성분의 비중이 거의 같게 추정되었다. 반면 공간 랜덤효과에서는 비구조화 성분의 비중이 더 크게 추정되었다. 넷째, 시간 랜덤효과에서 AR(1) 상관계수는 1에 근접한 값으로 강한 상관성을 보였다.

		표준편차	2.5 백분위수	중앙값	97.5 백분위수
공간 랜덤효과	τ_s	3748.038	1.233	34.968	2588.191
	ϕ_s	0.274	0.011	0.255	0.922
시간 랜덤효과	τ_η	279.268	10.729	99.472	870.532
	ρ_η	0.229	0.117	0.859	0.986
시공간 랜덤효과	τ_u	0.179	0.775	1.099	1.477
	ϕ_u	0.061	0.359	0.475	0.598
	ρ_u	0.001	0.990	0.993	0.995

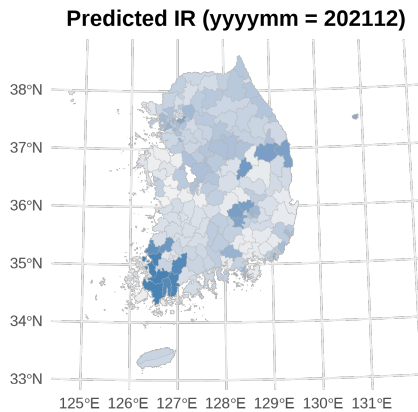
[표 4] 하이퍼파라미터 사후분포 추정 결과



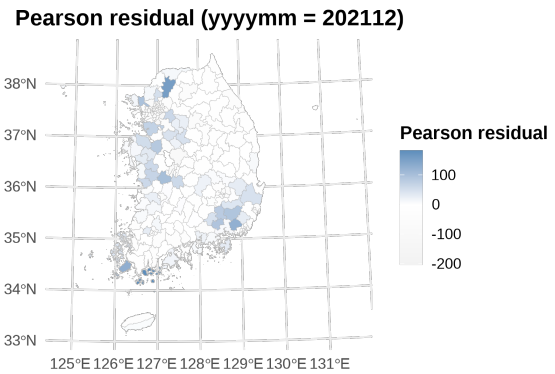
[그림 8] 마지막 3개 시점의 시공간 랜덤효과 분포

[그림 8]은 연속된 3개 시점을 슬라이싱하여 시공간 랜덤효과의 분포를 지도화한 것이다. 하이퍼파라미터 추정 결과에서 확인된 바와 같이 지역별 수준의 추세가 유지되는 흐름이 확인된다. 이는 E-9 분포가 경로 의존적 축적의 성격이 강함을 보여주며, 지역에 대한 정보를 바탕으로 한 지역 맞춤형 정책 설계가 필요함을 뒷받침한다. 또한, 공변량보다도 시공간 랜덤효과의 영향이 크다는 점은 관측 가능한 지역 특성 외에 제도 운영, 이주자 네트워크 등 비관측 요인의 시공간 구조가 더욱 중요할 수 있다는 점을 함의한다.³

³ AR(1)의 상관계수가 1에 가깝기도 하고, '저량'이라는 변수의 특성을 고려하면 다소 자명한 결과일 수 있겠습니다. 차분을 진행하여 유입 요인에 대한 분석을 진행해볼 필요성이 있다고 생각합니다.



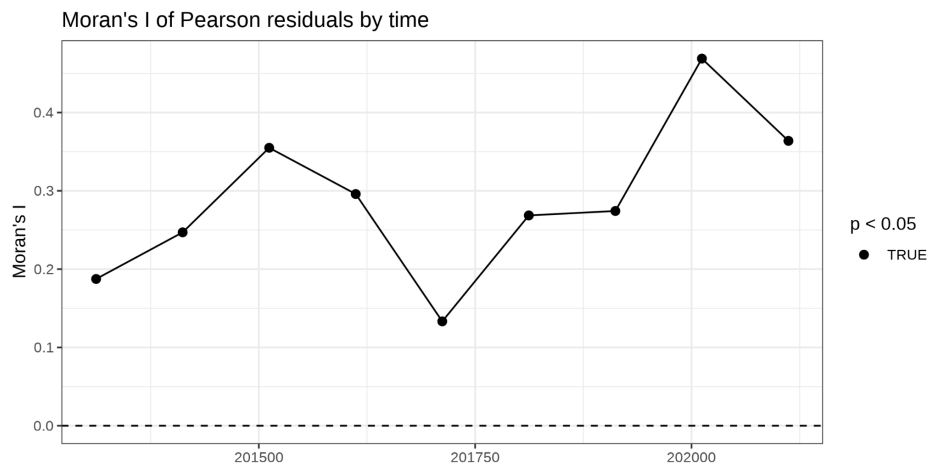
[그림 9] E-9 등록 인원 발생률 추정 결과



[그림 10] 2021 년 4 분기의 피어슨 잔차

4) 종속변수 추정 결과 및 잔차 검정

[그림 9]는 최종적으로 추정된 2021 년 4 분기의 E-9 등록 인원 발생률을 나타낸 지도이며, [그림 10]은 잔차의 값을 표준화한 피어슨 잔차를 나타낸 지도이다. 잔차 지도에서 특정 권역에 양(+) 또는 음(-)의 잔차가 지속적으로 군집하는 뚜렷한 공간 패턴은 관찰되지 않으며, 이는 모형이 주요한 공간 구조를 적절히 설명하고 있음을 시사한다. 또한 시점별로 피어슨 잔차의 전역적 모란 지수(global Moran's I)를 계산하고 통계적 검정을 실시한 결과([그림 11]), 모두 유의하지 않게 나와 시공간 베이지안 모형을 적합함으로써 검출 가능한 잔차의 공간적 자기상관을 제거하였음을 확인할 수 있었다.



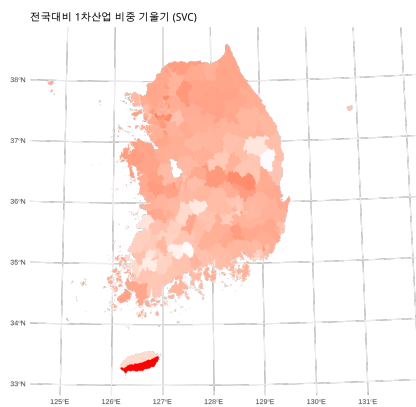
[그림 11] 시점별 피어슨 잔차의 전역적 모란 검정 결과

2. 확장된 모형

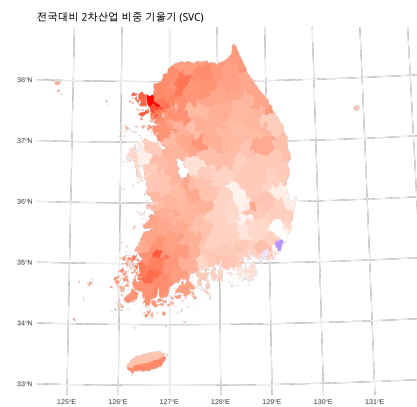
1) 공간 가변계수모형

다음으로, III 장에서 언급하였던 확장된 모형인 공간 가변계수모형을 적합하였다. 공간가변으로 두고자 하는 계수의 선택은 얻고자 하는 함의에 따라 달라질 수 있겠으나, 본 연구에서는 지역별 산업 요인이 핵심적인 변수라는 점을 IV-1 절에서 확인한 바 있으므로 해당 변수 영향력의 공간적 이질성을 추가적으로 탐구해보고자 전국 대비 1·2 차산업 GRDP 비중 편차 변수에 대한 계수를 공간가변으로 설정하였다.

먼저, WAIC 측면에서는 19419.26 으로 적합도의 개선이 일부 이루어진 사실을 확인하였다. IV-1 절에서 두 변수의 고정 효과(전역적) 회귀계수가 각각 0.031 과 0.122 였는데, 본 모형에서는 0.032 와 0.168 로 소폭 변동이 있었다. [그림 12]와 [그림 13]은 전국 대비 1·2 차산업 GRDP 비중 편차 회귀계수의 공간적 분포를 나타낸다. SVC 모형에서의 지역별 회귀계수는 “해당 지역에서 산업구조 편차가 E-9 발생률과 어떻게 결합되는가”를 보여주며, 특히 GRDP 비중 편차는 생산 중심 지표이므로 고용 구조와의 결합 방식이 지역별로 달라질 수 있다.



[그림 12] 전국 대비 1 차산업 GRDP 비중 편차 회귀계수의 공간적 이질성



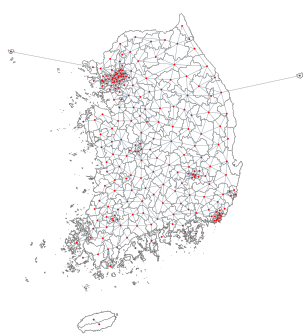
[그림 13] 전국 대비 2 차산업 GRDP 비중 편차 회귀계수의 공간적 이질성

전국 대비 1 차산업 GRDP 비중 편차는 전국적으로 크게 유의한 값을 보이지는 않으나, 경상북도와 전라남도 일부, 그리고 제주특별자치도 서귀포시에서 상대적으로 높은 양(+)의 영향력을 나타냈다. 이는 1 차산업 비중이 높은 지역이라 할지라도 E-9 발생률이 전국적으로 일관되게 증가하는 것이 아니라 특정 지역에서 노동집약적 농·어업이 결합될 때에 E-9 발생률과의 양(+)의 연관이 강화될 가능성을 시사한다.

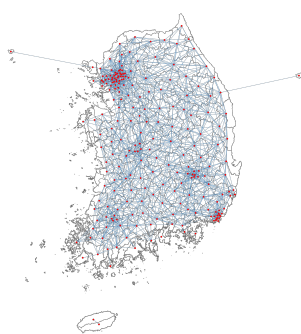
전국 대비 2 차산업 GRDP 비중 편차는 더욱 큰 공간적 이질성을 드러냈는데, 수도권과 전라남도 지역에서는 상대적으로 높은 양(+)의 영향력을, 경상남도와 부산 지역에서는 약한 음(-)의 영향력 혹은 무상관의 경향성을 보였다. 이는 2 차산업 GRDP 비중이 상대적으로 높은 지역이라도 산업

구성 혹은 규모와 E-9 외국인 근로자가 주로 투입되는 직종·사업체가 연관되는 메커니즘이 지역별로 상이할 경우 E-9 발생률에 영향력이 다르게 나타나기 때문이라는 가설을 설정할 수 있다. 예를 들어, 수도권에서 양(+)의 영향력이 나타난 것은 산업단지나 이주민 네트워크의 집적 혹은 수도권 선호 경향으로 E-9 외국인 근로자가 축적되기 상대적으로 쉬운 구조인 반면 경상남도나 부산권에서는 상대적으로 자본 집약적 제조업의 비중이 높아 GRDP 비중의 증가가 고용 증가로 이어지는 정도가 약하기 때문일 것으로 풀이해볼 수 있다.

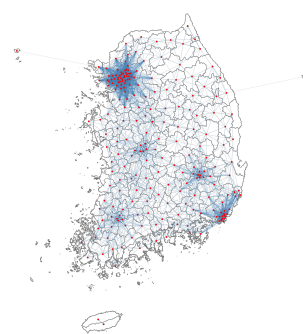
2) 상이한 공간근접성행렬(SPM)의 경우



[그림 14] 1 차 권 연결성 기반



[그림 15] 2 차 권 연결성 기반



[그림 16] 거리 기반

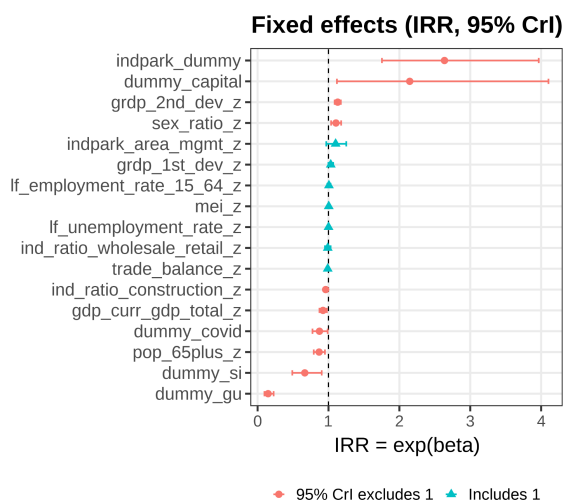
다음으로는 공간근접성행렬 W 의 설정에 따라 결과가 달라질 수 있는지에 대한 민감도 분석을 진행하였다. 현재는 [그림 14]과 같이 1 차 권 연결성 기반 공간근접성행렬을 이용하고 있으며, [그림 15], [그림 16]와 같은 (누적) 2 차 권 연결성 기반 혹은 거리 기반 공간근접성행렬을 이용한 경우에 대해 모형 적합 및 계수 추정에 미치는 영향력을 살펴보았다. (누적) 2 차 권 연결성 방식은 연결한 이웃의 이웃까지도 자신의 이웃으로 상정하는 것을 말한다. 거리 기반 방식은 임계거리를 설정하고 해당 범위 내에 중심점이 위치한 지역을 이웃으로 상정하는 것을 의미하며, 이를 통해 이격성을 반영하여 조밀한 지역일수록 많은 이웃을 가지게 된다. 본 연구에서는 임계거리로 60km 를 설정하였으며, 역거리 가중치를 계산할 때 스케일 파라미터로 30km 를 설정하였다.

	모형 상세	WAIC	
		2 차 권 연결성 기반	거리 기반 (60km)
모형 (1)	AR(1)	130211837	130211837
모형 (2)	AR(1) + 공변량	16071492.8	16071502.8
모형 (3)	BYM2 + AR(1)	98725.1857	98725.2823
모형 (4)	BYM2 + AR(1) + 상호작용	19584.0298	19540.1644
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	19576.6535	19556.0445

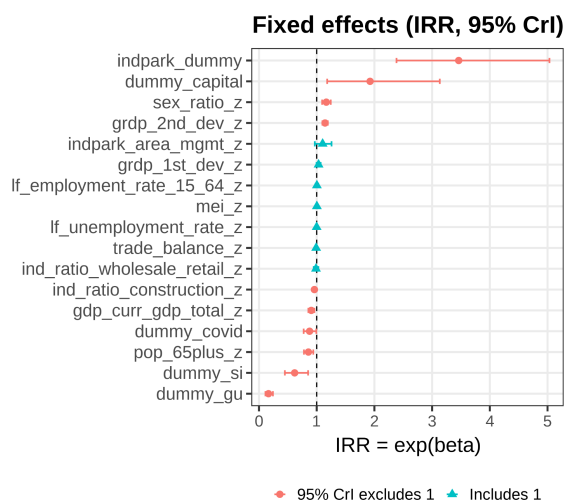
[표 5] SPM 을 달리하였을 때 WAIC 비교 결과

먼저 모형 적합도([표 5])를 비교하면 WAIC 기준으로는 큰 차이를 보이지 않았다. 공간 랜덤효과와 시공간 랜덤효과를 추가하였을 때 WAIC 가 크게 감소하는 경향성 또한 동일하게 관찰되었다. 다만 거리 기반 공간근접성행렬을 이용하였을 때 전체 모형의 WAIC 가 20 가량 감소하는 것을 확인할 수 있었다. 이는 거리 감쇠 및 이격성을 반영한 거리 기반 공간근접성행렬이 E-9 분포의 실제 공간적 상호작용 등을 학습한 결과일 수 있다. 다만 감소폭이 크지 않으므로 1 차 쿼 연접성 기반 공간근접성행렬을 사용하여도 무방할 것으로 판단된다.

고정 효과 추정 결과([그림 17], [그림 18]) 역시 큰 차이를 보이지 않았다. 회귀계수별로 추정량은 달라졌으나, 유의한 양(+) 또는 음(-)의 영향력을 지니는 변수의 목록은 IV-1 절에서 규명한 바와 동일하였다. 다만, 거리 기반 공간근접성행렬을 이용하였을 때 수도권 여부에 해당하는 더미의 95% 신용구간의 길이가 감소하는 사실을 확인할 수 있었다. 이는 이격성을 반영한 거리 기반 공간근접성행렬이 시군구가 조밀하게 위치한 수도권 지역에서의 공간 의존 구조를 잘 포착함으로써 수도권이라는 범주형 더미가 나타내던 공간적 차이를 공간 랜덤효과가 보다 효과적으로 설명하게 되었기 때문으로 해석할 수 있다.



[그림 17] 2 차 쿼 연접성 기반 SPM 을 이용한 고정 효과 추정



[그림 18] 거리 기반 SPM 을 이용한 고정 효과 추정

정리하면 IV-1 절의 분석 결과는 공간근접성행렬의 선택과 독립적이라는 사실을 확인할 수 있었다. 그러나 본 연구에서 적용해 본 공간근접성행렬의 종류가 제한적이라는 점에서, 추가적인 민감도 분석을 통해 공간근접성행렬 선택의 영향을 확인하고 더 나은 설명력을 도출하는 공간근접성행렬을 탐색할 필요성이 존재한다.

V. E-9 세부 자격별 외국인 근로자 패널 자료 분석

1. E-9-1(제조업)

1) 모형 비교

	모형 상세	WAIC
모형 (1)	AR(1)	90228845.8
모형 (2)	AR(1) + 공변량	7068782.98
모형 (3)	BYM2 + AR(1)	96277.1358
모형 (4)	BYM2 + AR(1) + 상호작용	16780.573
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	16773.4617

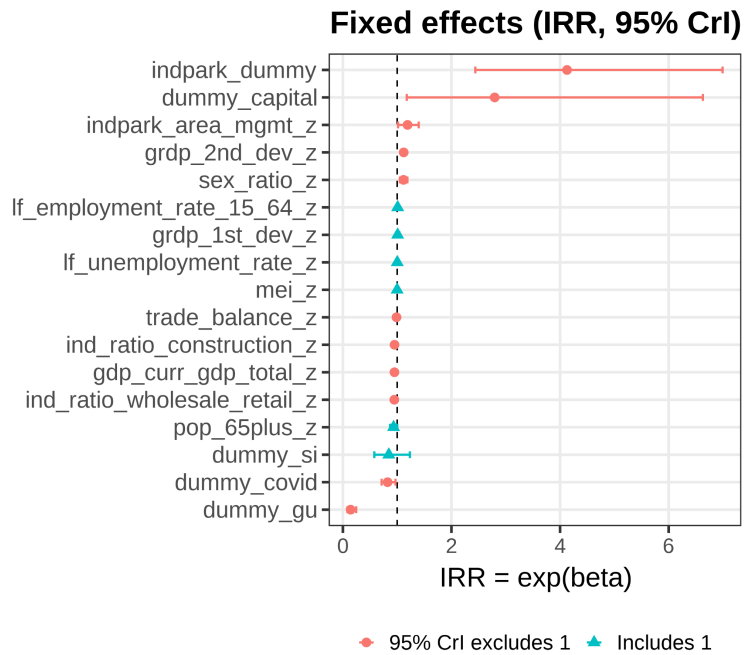
[표 6] WAIC 를 이용한 모형 비교 결과 (E-9-1, 연도별)

모형선택기준 WAIC 를 이용해 5 가지 모형을 비교한 결과, IV-1 절과 마찬가지로 공간 랜덤효과와 시공간 랜덤효과를 추가하면서 WAIC 의 급격한 감소, 즉 적합도의 증가가 나타났다. 또한 공변량까지 추가한 전체 모형의 WAIC 가 가장 낮은 것으로 나타나 이를 분석에 사용함에 있어서 합리성이 충족되었다.

2) 고정 효과 추정 결과

고정 효과 IRR 을 추정한 결과([그림 19]), IV-1 절에서 양(+)의 영향력을 나타내는 변수들이 E-9-1 제조업 분야 외국인 근로자 발생률에 양(+)의 영향력을 나타내는 것으로 드러났으며, 그 밖에도 산업단지 면적이 1 표준편차(=950 만 m²) 증가할수록 E-9-1 발생률이 19% 증가하는 유의한 양(+)의 영향력을 나타냈다. 제조업 인력 수요의 경우 산업단지와 관련된 영향력이 더 직접적이라는 사실을 확인할 수 있었다.

음(-)의 영향력을 나타내는 변수로는 무역수지와 도소매업 종사자 비율이 새롭게 추가되었으며, 고령인구 수와 시 지역 더미는 유의하지 않게 바뀌었다. 구체적으로 무역수지는 1 표준편차(=4 억 9 천만 불) 증가할수록 E-9-1 발생률이 2%가량 감소하고 도소매업 종사자 비율은 1 표준편차(=5.4%p) 증가할수록 E-9-1 발생률이 5%가량 감소하는 수준의 경향성을 보였다. 무역수지는 수출 중심의 제조업 기업의 경기를 반영한다는 점에서, 도소매업 종사자 비율은 도시 중심지에서 높게 나타난다는 점에서 제조업 분야 외국인 근로자 발생률과는 음(-)의 관계를 맺는 것으로 유추할 수 있다.



[그림 19] 고정 효과 IRR 추정 결과 및 95% 신용구간 (E-9-1, 연도별)

2. E-9-2(건설업)

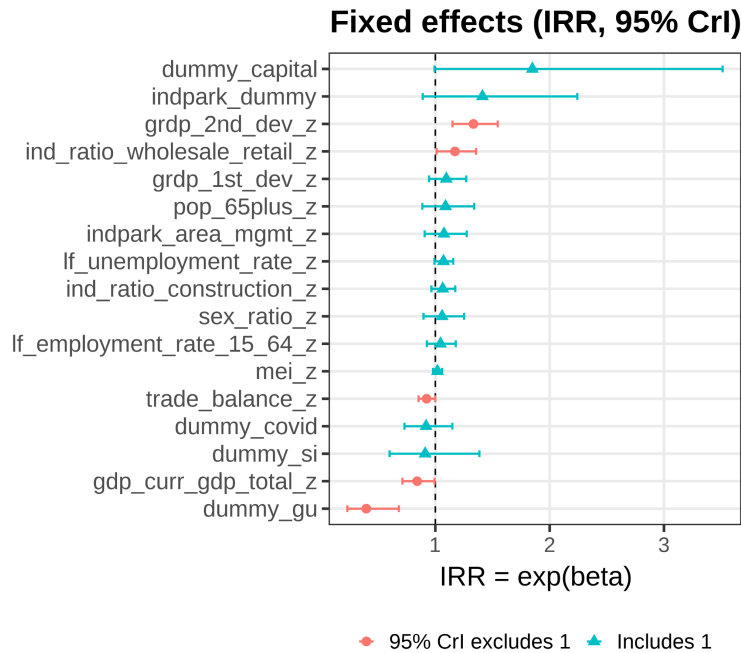
1) 모형 비교

	모형 상세	WAIC
모형 (1)	AR(1)	755885.052
모형 (2)	AR(1) + 공변량	317198.111
모형 (3)	BYM2 + AR(1)	95739.9475
모형 (4)	BYM2 + AR(1) + 상호작용	13004.8334
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	13004.1179

[표 7] WAIC 를 이용한 모형 비교 결과 (E-9-2, 연도별)

모형선택기준 WAIC 를 이용해 5 가지 모형을 비교한 결과([표 7]), IV-1 절과 마찬가지로 공간 랜덤효과와 시공간 랜덤효과를 추가하면서 WAIC 의 급격한 감소, 즉 적합도의 증가가 나타났다. 또한 공변량까지 추가한 전체 모형의 WAIC 가 가장 낮은 것으로 나타나 이를 분석에 사용함에 있어서 합리성이 충족되었다.

2) 고정 효과 추정 결과



[그림 20] 고정 효과 IRR 추정 결과 및 95% 신용구간 (E-9-2, 연도별)

건설업 분야 E-9-2 외국인 노동자 발생률에 대한 고정 효과 추정 결과([그림 20])는 E-9 혹은 E-9-1 발생률에 대한 고정 효과 추정 결과와 다소 상이한 모습이었다. 먼저 유의한 양(+)의 영향력을 나타낸 변수로, 전국 대비 2 차산업 GRDP 비중 편차와 도소매업 종사자 비율이 각각 1 표준편차 증가 시 33%, 17% 증가하는 경향성을 나타냈다. 건설업에 대한 수요는 제조업 중심의 산업구조와 함께 움직이는 경향이 강하다는 측면과 도소매업 종사자 비중이 높은 지역은 상권·생활권 확장 등의 수요 상승을 불러일으켜 건설업 수요를 동반할 가능성이 있다는 측면에서 두 변수의 영향력을 설명할 수 있을 것이다.

반면 유의한 음(-)의 영향력을 나타낸 변수로는 무역수지, GRDP 총액, 구 지역 여부 더미가 있었다. 구체적으로 무역수지는 1 표준편차 증가함에 따라 E-9-2 발생률이 8%가량 감소하는 경향을, GRDP 총액은 1 표준편차 증가함에 따라 16% 감소하는 경향을, 구 지역의 경우 군 지역에 비해 발생률이 60%가량 감소하는 경향을 나타냈다. 무역수지나 GRDP 총액의 경우 구체적인 해석을 제시하기는 어렵지만 반드시 무역 지표나 생산액으로 대표되는 경제 상황이 좋아짐에 따라 건설업의 신규 인력 수요가 증가하지는 않는다는 사실을 확인할 수 있었다. 또한 건설업 인력 역시 구 지역과 같은 도심보다는 외곽에 체류하는 경향성이 높다는 사실을 확인할 수 있었다.

추가적으로, V-1 절의 제조업 발생률 고정 효과와 비교하였을 때, 산업단지 더미 변수와 코로나 시기 더미의 영향력이 약해진 사실을 확인할 수 있었다. 이는 산업별 외부 충격으로 인한 인력 수요 요인의 변동이 상이함을 함의한다.

3. E-9-3(농업)

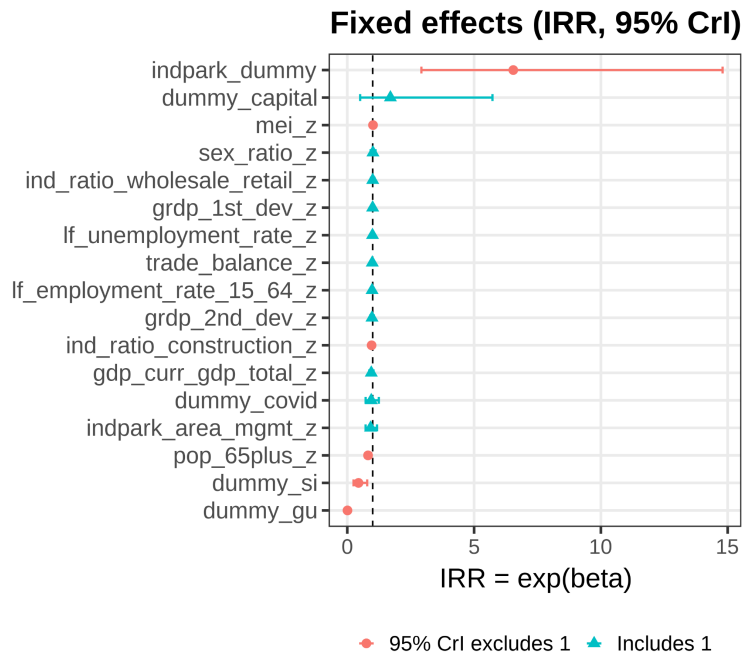
1) 모형 비교

	모형 상세	WAIC
모형 (1)	AR(1)	23439955.7
모형 (2)	AR(1) + 공변량	2532374.96
모형 (3)	BYM2 + AR(1)	30555.5394
모형 (4)	BYM2 + AR(1) + 상호작용	11673.0881
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	11674.9491

[표 8] WAIC 를 이용한 모형 비교 결과 (E-9-3, 연도별)

모형선택기준 WAIC 를 이용해 5 가지 모형을 비교한 결과([표 8]), IV-1 절과 마찬가지로 공간 랜덤효과와 시공간 랜덤효과를 추가하면서 WAIC 의 급격한 감소, 즉 적합도의 증가가 나타났다. 또한 공변량까지 추가한 전체 모형의 WAIC 가 가장 낮은 것으로 나타나 이를 분석에 사용함에 있어서 합리성이 충족되었다.

2) 고정 효과 추정 결과



[그림 21] 고정 효과 IRR 추정 결과 및 95% 신용구간 (E-9-3, 연도별)

고정 효과 IRR 추정 결과([그림 21]), 농업 분야 E-9-3 발생률에 양(+)의 영향력을 미친 변수는 MEI 와 산업단지 소재 여부 더미로 나타났다. MEI 는 1 표준편차(=8.7%p) 증가함에 따라 1.4%가량

발생률을 증가시키는 것으로 나타났는데, 이는 농촌의 경우 외부로의 유출 인구가 높음에 따라 MEI와는 음(-)의 영향력을 가질 것이라는 가설과 반대되는 결과였다. 한편, 산업단지가 소재한 지역은 미 소재 지역보다 6 배 가량 발생률이 높은 것으로 나타났는데, 이는 산업단지 자료에 농공단지가 포함되어 있으며 산업단지 소재 여부가 농업 관련 특성을 어느 정도 대리하고 있는 것으로 해석할 수 있다.

반면, 고령인구 수, 건설업 종사자 비율, 시/구 지역 여부는 음(-)의 영향력을 나타내었다. 고령인구는 1 표준편차 증가함에 따라 18% 발생률이 감소하는 패턴을 보였는데, 이는 고령화가 심한 지역에서는 오히려 농업 관련 산업 구조가 악화되어 있음을 방증한다. 건설업 종사자 비율은 1 표준편차 증가함에 따라 E-9-3 발생률이 4%가량 감소하는 경향을 보이며, 시나 구 지역에서는 발생률이 각각 60%, 93%가량 감소하는 경향을 보여 농업 부문이 군 지역에 편중되어 있음을 드러내고 있다.

VI. 결론 및 제언

본 연구는 2013 년부터 2021 년까지의 E-9 외국인 근로자 분포를 시공간 베이지안 계층 모형으로 분석하였으며, 전체 E-9 등록 인원 및 세부 자격별 등록 인원 수의 결정 요인을 파악하였다. 주요 분석 결과는 다음과 같다: 첫째, 전체 E-9 발생률에 대한 고정 효과를 분석한 결과 산업단지 소재 여부, 수도권 여부, 전국 대비 2 차산업 GRDP 비중 편차, 성비는 양(+)의 영향력을, 고령인구, GRDP 총액, 건설업 종사자 비율, 코로나-19 및 시/구 지역 여부는 음(-)의 영향력을 가졌다. 둘째, 전체 E-9 발생률에 대한 랜덤효과를 분석한 결과, 시공간 상호작용 랜덤효과의 변동성이 가장 크게 나타났으며, 상호작용의 시간적 지속성 역시 높게 나타나 단기적 충격보다는 장기적이고 경로의존적인 추세에 따라 E-9 발생률이 변동하는 것으로 파악되었다. 셋째, 전국 대비 산업별 GRDP 비중 편차를 나타내는 변수를 SVC 모형으로 확장함으로써, 산업별 요인이 E-9 발생률에 영향을 미치는 메커니즘이 지역별로 상이할 가능성을 확인하였다. 넷째, 세부 자격별 패널 자료에 대해 시공간 베이지안 모형을 적합하여, 각 세부 자격별로 서로 다른 공변량이 영향을 미치고 있음을 확인하였다.

본 연구는 시공간 의존성을 반영하는 모형이 E-9 외국인 근로자 등록 인원과 같은 시공간 패널 자료를 통계적으로 안정적으로 모델링할 수 있다는 사실을 확인하였다는 점에서 의의가 있다. 특히 모형 비교를 통해 시공간 구조의 영향을 크게 받는 경로 의존적 축적의 성격이 강함을 실증적으로 확인함과 동시에, 시공간 의존성의 영향을 배제한 공변량들의 고정 효과까지 통합적으로 추정하였다는 점에서 차후 지역별 맞춤형 정책 설계에 있어서 참고 자료로 활용될 수 있을 것으로 기대된다. 또한 상이한 공간근접성행렬이나 공간 가변계수모형으로의 확장, 혹은 사전 정보에 기반한 계층 구조의 상이한 모델링 등 모형 설계의 유연성이 높다는 점에서 향후 E-9 패널을 비롯한 인구 구조 변동의 다층성 연구에 활용될 가능성을 함의한다.

다만 본 연구는 다음의 한계점을 노정한다: 첫째, 공변량을 추가하였을 때 WAIC 개선 효과가 크지 않은 등 공변량의 설명력이 다소 제한적으로 드러났다. 이는 공변량이 설명하는 변동이 이미 랜덤효과에 흡수되었을 수 있는 공간 교란(spatial confounding)의 가능성(Paciorek, 2010; Donegan, 2025)을 시사하며, 이를 올바르게 다루기 위한 방법론적 정련화가 요구된다. 혹은 차분을 취하여 저장량이 아닌 유량(flow) 데이터의 분석을 시도해볼 필요가 있다. 둘째, 분석 기간 및 단위의 제약이 존재한다. 본 연구는 변수의 구득 가능 여부를 고려하여 시군구-연도별 자료를 구축하였는데, 그로 인해 시도별로 제공되는 사업체별 노동력 부족 인원 등의 변수를 사용하지 못하거나 더 세분화된 시간 단위에서의 분석이 이루어지지 못했다는 아쉬움이 있다. 셋째, 세부 자격별 모형 적합이 개별적으로 이루어져 분석의 효율성이 떨어지고 카테고리 간 분석 결과에 대한 직접 비교가 불가능하다는 점이다. 이에 대해 범주별 공동 요인을 모형에 통합하는 공유 요소 모형(shared component model) 혹은 다변량 계층 모형의 적합을 시도해 볼 필요가 있다(Vicente et al., 2020 등). 향후 자료 및 방법론을 보완함으로써, 본 연구의 실증적 분석 시도가 정책 수립의 토대가 되어 이주 노동자들이 한국 사회의 일원으로 정착하고 지역 산업의 지속가능성에 기여하도록 하는 데 도움이 되기를 희망한다.

참고 문헌

- 고용노동부·한국산업인력공단. (2024). 외국인 고용허가제 20 주년 백서.
https://www.moel.go.kr/policy/policydata/view.do?bbs_seq=20241101417
- 이원규, 이재우. (2025). 고용허가제 비전문취업(E-9) 외국인 근로자의 주거 만족도와 이주 의향에 미치는 영향 요인에 관한 연구. *한국지역개발학회지*, 37(3), 171-194.
- 이찬영. (2018). 시군 지역 노동시장의 내외국인 고용관계. *산업경제연구*, 31(6), 2293-2317.
- 이현욱. (2015). 한국 고용허가제 정책과 이주노동자의 공간분포 특성. *한국도시지리학회지*, 18(3), 57-74.
- 유희연. (2023). 우리나라 저숙련 이주노동자 지역분포 현황과 특성 (KRIHS Working Paper 23-05). 국토연구원.
- 정기선. (2025 년 9 월 4 일). 이민정책, 감(感) 아닌 데이터로...지역사회 통계가 부족하다. *교수신문*.
<https://www.kyosu.net/news/articleView.html?idxno=142499>
- 조영태. (2021). 인구 미래 공존: 인구학의 눈으로 기획하는 미래. 북스톤.
- 조은지, 이찬영. (2021). 도구변수법을 이용한 지역 노동시장의 내외국인근로자 고용관계 분석. *노동경제논집*, 44(2), 33-69.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). Hierarchical modeling and analysis for spatial data (2nd ed.). Chapman & Hall/CRC. doi:10.1201/b17115
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20. doi:10.1007/BF00116466
- Blangiardo, M., & Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. Wiley. doi:10.1002/9781118950203
- Bradley, J. R., Holan, S. H., & Wikle, C. K. (2018). Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Analysis*, 13(1), 253–310. doi:10.1214/17-BA1069
- Donegan, C. (2025). Plausible reasoning and spatial-statistical theory: A critique of recent writings on “spatial confounding.” *Geographical Analysis*, 57(1), 152–172. doi:10.1111/gean.12408

- Mazzocchi, A., & Greco, L. (2011). Bayesian hierarchical modelling of internal migration rates: An application to Italian provinces. *Statistica*, 71(2), 165–182. doi:10.6092/issn.1973-2201/3607
- Paciorek, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1), 107–125. doi:10.1214/10-STS326
- Paglino, E. (2024). Estimating excess migration associated with tropical storms in the USA, 1990–2010. *Population and Environment*, 46, Article 11. doi:10.1007/s11111-024-00452-9
- Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4), 1145–1165. doi:10.1177/0962280216660421
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319–392. doi:10.1111/j.1467-9868.2008.00700.x
- 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. doi:10.1214/16-STS576
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- Vicente, G., Goicoa, T., & Ugarte, M. D. (2020). Bayesian inference in multivariate spatio-temporal areal models using INLA: Analysis of gender-based violence in small areas. *Stochastic Environmental Research and Risk Assessment*, 34, 1421–1440.

코드 및 데이터

분석에 활용된 코드 및 데이터는 <https://github.com/jihography/e9> 에 업로드되어 있습니다.

A. E-9 외국인 근로자 패널 자료 분석 결과 (분기별)

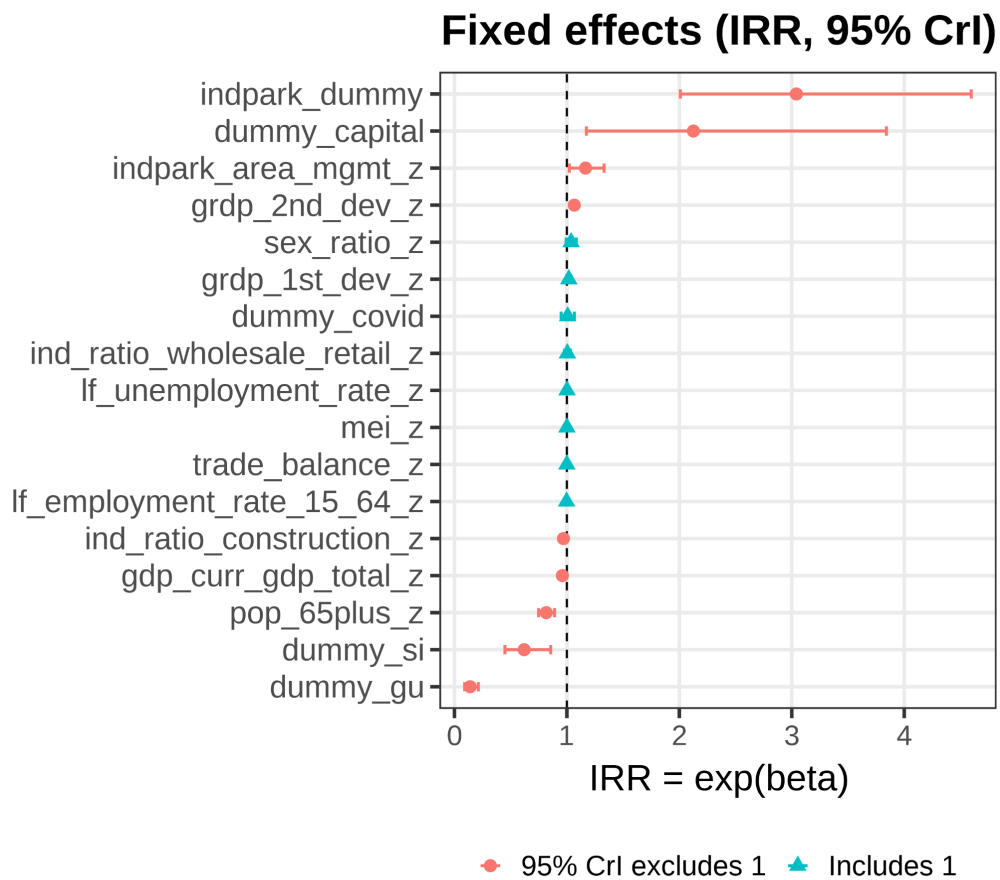
	모형 상세	WAIC
모형 (1)	AR(1)	500637134.26
모형 (2)	AR(1) + 공변량	75579318.75
모형 (3)	BYM2 + AR(1)	721960.11
모형 (4)	BYM2 + AR(1) + 상호작용	73174.89
모형 (5)	BYM2 + AR(1) + 상호작용 + 공변량	73249.93

[표 A1] WAIC 를 이용한 모형 비교 결과 (E-9 전체, 분기별)

변수명	변수 코드	평균	표준편차	신용구간	원변수 표준편차
성비 (z)	sex_ratio_z	0.037	0.022	(-0.007, 0.080)	5.397
고령인구비율 (z)	pop_65plus_z	-0.203	0.044	(-0.290, -0.116)	23966.801
무역수지 (z)	trade_balance_z	0	0.001	(-0.002, 0.003)	487751.486
MEI (z)	mei_z	0	0.003	(-0.006, 0.005)	0.080
전국 대비 1 차산업 GRDP 비중 편차 (z)	grdp_1st_dev_z	-0.041	0.017	(-0.074, -0.008)	0.084
전국 대비 2 차산업 GRDP 비중 편차 (z)	grdp_2nd_dev_z	0.017	0.01	(-0.003, 0.037)	0.181
GRDP 총액 (z)	gdp_curr_gdp_total_z	0.064	0.012	(0.040, 0.087)	10224090.881
건설업 종사자 비율 (z)	ind_ratio_construction_z	0.153	0.068	(0.021, 0.286)	0.039
도소매업 종사자 비율 (z)	ind_ratio_wholesale_retail_z	-0.031	0.008	(-0.047, -0.016)	0.054
실업률 (z)	lf_unemployment_rate_z	0.005	0.013	(-0.021, 0.031)	0.151
고용률(15- 64 세) (z)	lf_employment_rate_15_64_z	-0.003	0.006	(-0.014, 0.008)	0.543
코로나-19 시기	dummy_covid	0.001	0.003	(-0.004, 0.006)	0.416

수도권 여부	dummy_capital	0.007	0.03	(-0.052, 0.065)	0.453
시 지역 여부	dummy_si	0.754	0.302	(0.160, 1.346)	0.473
구 지역 여부	dummy_gu	-0.478	0.165	(-0.801, -0.155)	0.459
산업단지 소재 여부	indpark_dummy	-1.971	0.216	(-2.395, -1.548)	0.405
산업단지 총면적 (z)	indpark_area_mg mt_z	1.112	0.211	(0.698, 1.525)	9539.033

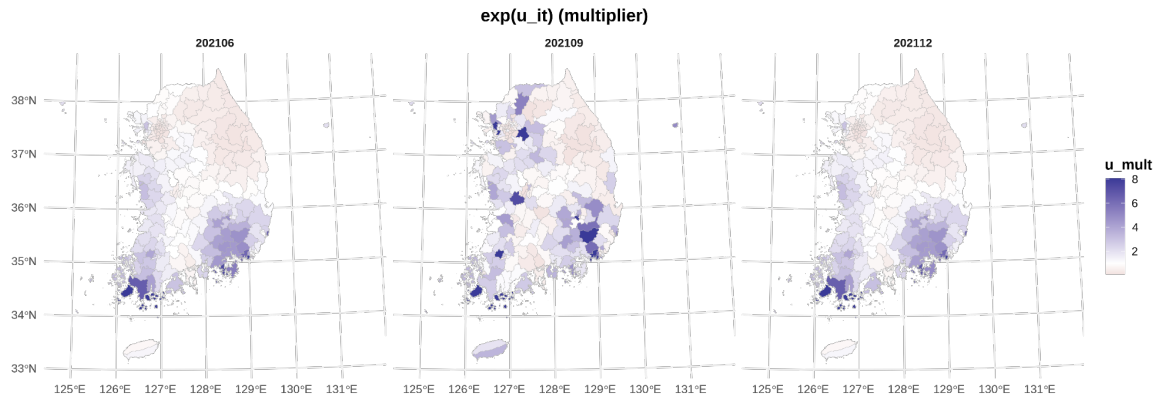
[표 A2] 고정 효과 계수 추정 결과



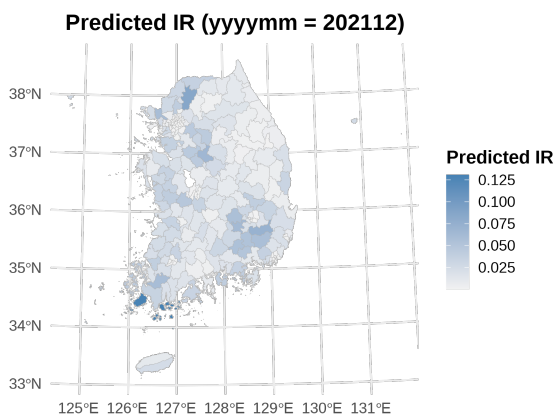
[그림 A1] 고정 효과 IRR 추정 결과 및 95% 신용구간 (E-9 전체)

		표준편차	2.5 백분위수	중앙값	97.5 백분위수
공간 랜덤효과	τ_s	35610.661	5.504	156.534	17308.823
	ϕ_s	0.315	0.073	0.526	0.995
시간 랜덤효과	τ_η	130.762	2.872	48.600	429.106
	ρ_η	0.043	0.851	0.983	0.999
시공간 랜덤효과	τ_u	0.138	1.071	1.329	1.614
	ϕ_u	0.030	0.223	0.279	0.342
	ρ_u	0.000	0.997	0.998	0.998

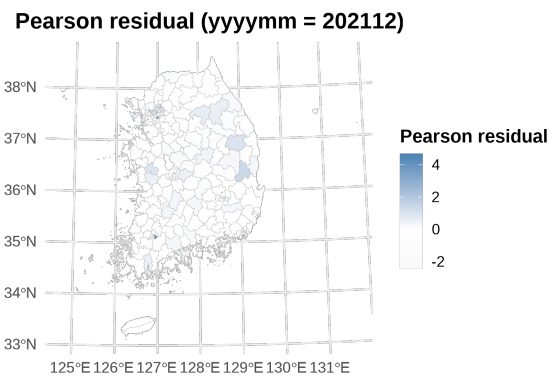
[표 A3] 하이퍼파라미터 사후분포 추정 결과



[그림 A2] 마지막 3 개 시점의 시공간 랜덤효과 분포



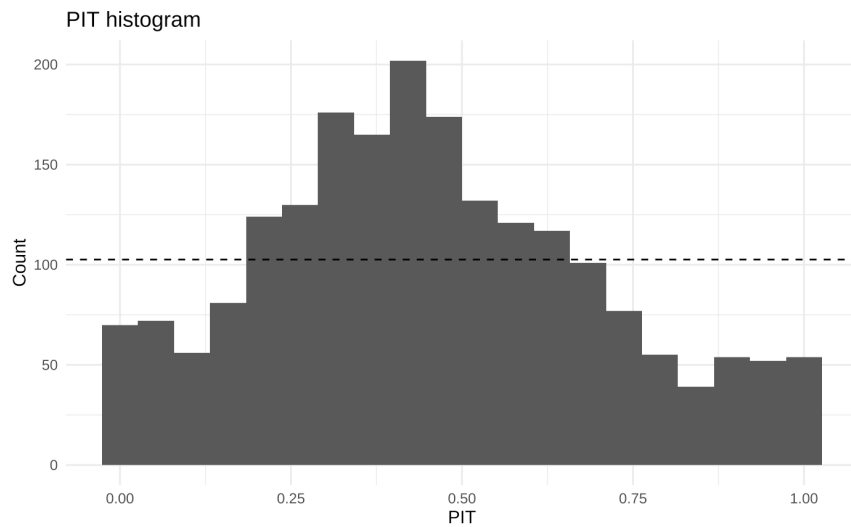
[그림 A3] E-9 등록 인원 발생률 추정 결과



[그림 A4] 2021 년 4 분기의 피어슨 잔차

- 전체적으로 연도별 자료를 이용하였을 때와 유사한 분석 결과가 나타났다.
- 그러나, [그림 A2]에서와 같이 시공간 랜덤효과가 반기(2 분기) 간격으로 상이하게 나타나는 현상이 확인되었고, 이는 상이한 시간 해상도의 자료가 포함된 것이 원인 중 하나로 추측된다.
- 따라서 계절 추세를 추가적으로 모델링하는 등의 수정 사항이 필요할 것으로 사료된다.

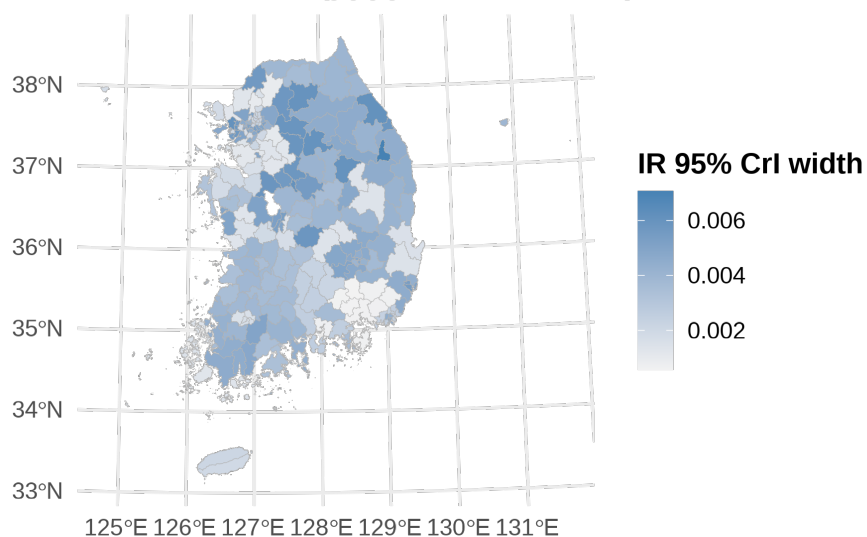
B. E-9 전체인원 종속변수 추정 결과: 추가 통계량



[그림 B1] PIT 히스토그램: 분산이 0.06 으로 균등분포의 분산 0.083 에 비해 낮다.

- 사후예측분포 F_{it} 에서 $PIT = F_{it}(y_{it})$ 를 히스토그램으로 나타낸 것이다. 모형이 잘 맞을 경우 이상적으로 PIT는 균등분포의 형태를 보여야 하나, 현재는 가운데에 몰려 있고 양 끝이 낮아 불확실성을 과대추정하고 있을 가능성이 관측된다.

IR 95% CrI width (yyyyymm = 202112)



[그림 B2] 2021년 4 분기의 사후 추정된 IR 신용구간의 길이