

## 속성 유사성을 고려한 지리가중회귀분석(SGWR)을 활용한 출산력 영향 요인의 공간적 이질성 분석

사회과학대학 지리학과  
2020-17530 박지호

### I. 서론

Tobler (1970)의 “모든 것은 서로 연관되어 있다. 그러나 가까운 것은 더욱 밀접하게 연관되어 있다”는 언명은, 공간 데이터에서 공간적 의존성(spatial dependence)과 공간적 자기상관(spatial autocorrelation)이 흔히 나타난다는 사실을 간결하게 요약하며 ‘지리학 제1 법칙’으로 불린다. 그러나 어떤 법칙도 예외 없이 통용되기는 어렵고, 이러한 법칙화가 현실을 과도하게 단순화한다는 비판도 꾸준히 제기되어 왔다.(Sui, 2004) 공간 분석가들은 이러한 도전에 직면하였을 때 정량적 도구를 이용하되 실제 공간을 보다 잘 반영할 수 있는 방법을 개발하고자 노력한다. 그 결과물 중 하나가 공간적 의존성을 고려하되 변수 간 관계의 ‘공간적 이질성(spatial heterogeneity)’을 국지적으로 모델링하는 지리가중회귀(Geographically Weighted Regression, GWR) 모형이다. 물론 GWR 역시 여러 한계가 지적되었고, 이를 보완하려는 시도가 계속되고 있다. 한편으로는 일부 연구자들이 부주의하게 공간 분석 기법을 사용하는 경우도 종종 있다. Bailey and Gatrell(1995)는 공간 데이터의 특수성으로 인해 ‘검손함’이 공간 분석가의 미덕이어야 함을 강조한 바 있는데, 실제로는 GWR과 같은 공간 통계 기법의 한계를 고려하지 않고 결과를 정당화하거나 그 구조를 정확히 이해하지 못한 채 분석 결과를 잘못 해석하는 일이 적지 않다.

따라서 본 연구의 목적은 크게 두 가지로 정리할 수 있다. 첫째는 주제적 측면에서 인구 동태의 가장 핵심적인 변수인 출산력에 영향을 미치는 요소들의 공간적 이질성을 분석하는 것이다. 이를 통해 출산력에 영향을 미치는 복합적 인과 메커니즘의 공간적 맥락을 제시하고자 한다. 둘째는 방법론적 측면에서 기존 GWR 기반 연구의 한계를 보완하고 설명력 및 정확성을 높이하고자 하는 것이다. 이를 위해 속성 유사성 및 지리적 근접성을 포괄하는 방법론으로서 개발된 SGWR(Similarity and Geographically Weighted Regression) 모형을 이용한다. 최종 모형 선택 이전에 여러 가지 탐색 방법을 적용했을 때와 다른 GWR 계열 모형을 적합했을 때의 적합 결과들을 비교함으로써 SGWR의 적용 가능성을 점검한다. 최종적으로, SGWR은 공간 단위의 설정 혹은 변수의 공간적 이질성 정도에 대해 GWR에 비해 강건한 특성을 지니므로 본 연구를 통해 출산력의 공간적 이질성 분석에 있어서 기존 연구 대비 편향이 개선된 분석 결과를 제시할 수 있을 것으로 기대된다.

## II. 선행 연구 검토

### 1. SGWR

Brundson et al.(1996)에 의해 최초로 제시된 GWR은 다양한 분야에서 변수 간 관계의 공간적 이질성 혹은 공간적 비정상성을 모델링하는 데 널리 활용되어 왔다. GWR은 개별 관측지에서 다른 관측지까지의 공간적 근접성에 따라 더 높은 가중치를 부여한 뒤 관측치 개수만큼의 회귀 모형을 적합하는 방식이다. 이후 다양한 방법론적 확장이 있었는데, 일부 설명변수 세트의 영향력을 전역적으로 모델링하도록 지정할 수 있는 준 모수적 GWR(Semi-Parametric GWR)이 제시되기도 하였으며, 최근에는 모든 설명변수의 작동 스케일(대역폭)을 상이하게 모델링할 수 있는 MGWR(Fotheringham et al., 2017)이 개발되어 모형의 유연성이 크게 높아졌다.

그러나 GWR의 한계점 중 하나는 지리적으로 근접한 공간 단위 사이에는 변수의 영향력이 동일적이라는 단순한 가정에 기반한다는 것이다. (Anselin and Li, 2020) 즉, 만약 a와 b라는 지점이 i라는 지점으로부터 동일한 거리를 두고 떨어져 있다면 a, b 두 지점의 관측치는 i 지점의 모형을 적합함에 있어서 동일한 크기의 가중치를 부여받는다. 그러나 실제로는 a 지점과 b 지점의 영향력 혹은 연관성은 불균형적일 때가 적지 않다. 예를 들어 a와 b 지점이 각각 도시와 농촌이라거나, 정치적 지지 세력이 다른 상황을 상상해 볼 수 있다. 실제로 지리적 근접성과 속성 유사성이 불일치할 때, 특히 ‘음의 공간적 자기상관’이 존재하는 경우 GWR 모형의 정확도가 다소 낮아질 수 있다 (Griffith, 2019)는 지적이 제기되었다.

최근에는 이러한 가정의 한계를 극복하고자 공간적 의존성과 공간적 이질성에 더하여 속성 유사성<sup>1)</sup>을 공간 분석의 원리 중 하나로 포괄 혹은 확장하고자 하는 시도가 늘어나고 있다. 그 예시로 네트워크 거리 혹은 속성 유사성 거리 기반 커널 함수를 적용하거나 (Ni et al., 2022; Jing et al., 2022), 네이버후드를 인접 공간 단위 개념에서라 해당 지역의 사람들이 하루 동안 방문한 지역(‘Network Neighborhood’)으로 확장한 연구(Brazil, 2023)가 있다. GWR의 맥락에서는 Lessani and Li(2024)에 의해 지리적 근접성 가중치와 속성 유사성 가중치를 결합한 SGWR 모형이 최초로 제시되었으며, 이후 Yu et al.(2025)에 의해 커널 함수의 수정 및 경사 하강 알고리즘 기반 최적해 탐색 방법이 제시되었다.

그러나 아직 SGWR이 다양한 데이터에 적용되지는 않았으며, 모형의 적합 방식 역시 아직 정교화되었다고 볼 수 없다. 따라서 본 연구에서는 SGWR 모형을 활용하기 전 속성 유사성 가중치 커널을 일부 수정하고, 초깃값 및 탐색 방법에 따라 적합 결과가 달라지는 않는지 확인함으로써 GWR 대비 설명력 향상을 이루어냄과 동시에 분석 결과를 잘못 해석하거나 과대 해석하는 일이 없도록 유의하고자 할 것이다.

---

1) Zhu·Turner(2022)는 “두 위치의 지리적 조건(특성)이 비슷하다면, 어떤 변수가 비슷할 확률이 높다 [The more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)]”를 ‘지리학 제3 법칙’으로 명명하기도 하였다.

## 2. 국내 출산력의 공간적 이질성 분석

정명구(2017) 및 Jung et al.(2019)는 GWR을 이용하여 각각 2010년과 2015년 합계출산율에 영향을 미치는 요인의 공간적 변이를 파악하였다. 분석단위는 시군구 수준이며, 설명변수로 정명구(2017)에서는 인구밀도, 가임기 여성 비율, 가임기 여성 미혼율, 아파트 전셋값, 유아 보육시설 수, 사업체 수, 고령인구 비율, 노인복지시설 수, 건강보험료 대비 급여 지급액을 활용하였고, Jung et al.(2019)에서는 아동 1000명당 어린이집 수, 대졸 이상 여성 비율, 인구밀도, 순 이동률, 국제 혼인 비율, 인구성장률을 활용하였다. 분석 결과 변수별 영향력의 공간적 이질성 및 공간적 자기상관이 나타났으며, 공간적 자기상관이 나타남에 따라 출산력 수준에 대한 영향에는 파급효과(spillover effect)가 작용한다고 해석하였다.

장인수·정찬우(2022)는 다중스케일 지리가중회귀모형(MGWR)을 이용하여 출산 지원금이 모의 연령대별 출산율에 미치는 영향을 분석하였다. 분석단위는 시군구 수준이며, 출산 지원금 수혜 금액 외의 공변량으로는 재정자립도, 조혼인율, 고령인구 비율, 주민등록 연앙인구를 활용하였다. 분석 결과 MGWR이 OLS에 비해 높은 조정  $R^2$  값을 나타냈으며, 출산 지원금의 영향력의 공간적 이질성을 파악하여 재정 지원 정책을 시행함에 있어서 지역적 특성을 고려해야 한다는 함의를 도출하였다.

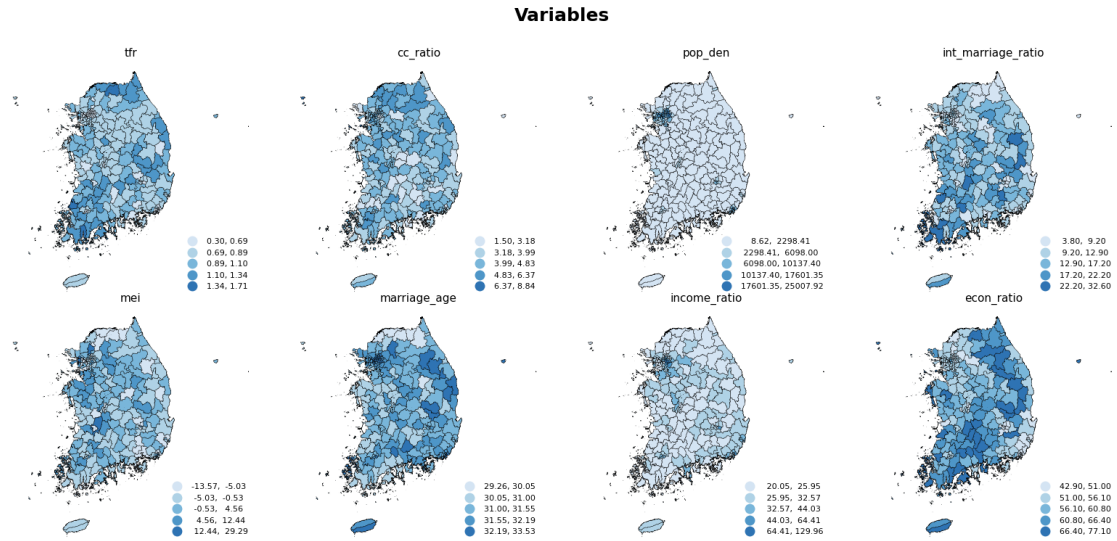
이영서(2024)는 지리가중회귀모형을 이용하여 출산장려금과 출산율의 관계를 탐색하였다. 분석단위는 마찬가지로 시군구 수준이며, 공변량으로는 인구밀도, 아동 1000명당 소아청소년과 전문의 수, 유아 1000명당 보육시설 수, 여성의 평균 초혼 연령, 재정자립도, 주택 소유율, 생활환경 만족도를 투입하였으며, 출산장려금은 지자체별 출산장려금 전체 지출액을 수혜 출생아 수로 나누어 분석을 진행하였다. 그 결과 강원도 일부와 전라남도 일부에서 출산장려금이 유의미한 정(+)의 영향력을 보였다.

위의 선행 연구는 대체로 유사한 분석 틀을 가지고 진행되었으며, 분석단위로 시군구 수준을 사용하며, (M)GWR 모형으로부터 추정된 회귀계수가 공간적으로 군집된 분포를 보이므로 지역별 영향력의 파급효과가 있음을 서술하며 정책적 함의를 강조하고 있다. 그러나 공통적으로 결과 해석에 있어서 오류가 나타나고 있는데, 추정된 회귀계수가 군집된 형태를 보이는 것은 (M)GWR 모형이 주변의 데이터를 ‘빌려 오는’ 과정 및 커널 평활화에 따른 자연스러운 결과이며 그것을 파급효과에 의한 것으로 상정할 때는 보다 면밀한 고려가 필요하기 때문이다. 특히 물리적 거리가 가깝더라도 속성 유사성의 차이가 크거나 가중치의 설정으로 적합에 이용되는 실제 관측지 수가 적어지는 경우, GWR 모형으로 추정된 회귀계수에는 다소간의 편향이 발생할 수 있다.

따라서 본 연구에서는 최근에 개발된 SGWR 모형이 GWR 모형의 일부 한계점을 개선하고 더 높은 설명력을 지닌다는 점에서 SGWR 모형을 활용하고자 한다. 특히 시군구 수준 데이터의 경우 관측지의 절대적 숫자가 크지 않을뿐더러 인접 지자체라고 하더라도 지리적 특성이 다른 지자체가 많이 포함될 가능성, 즉 공간 단위 임의성 문제(MAUP)가 발생할 가능성이 커 SGWR 모형을 활용하는 것의 합리성이 더 높다고 여겨진다.

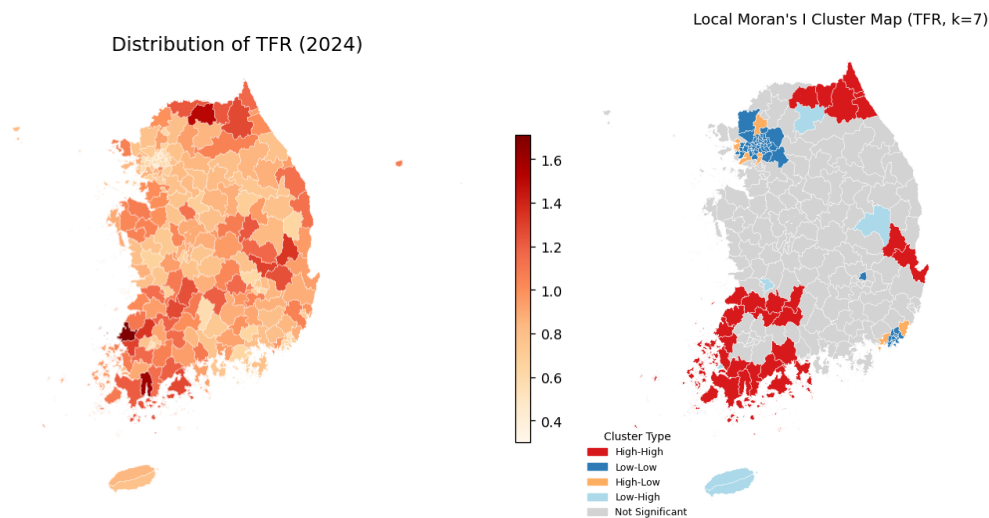
### III. 데이터 및 연구 방법론

#### 1. 출산력 및 설명 변수의 설정



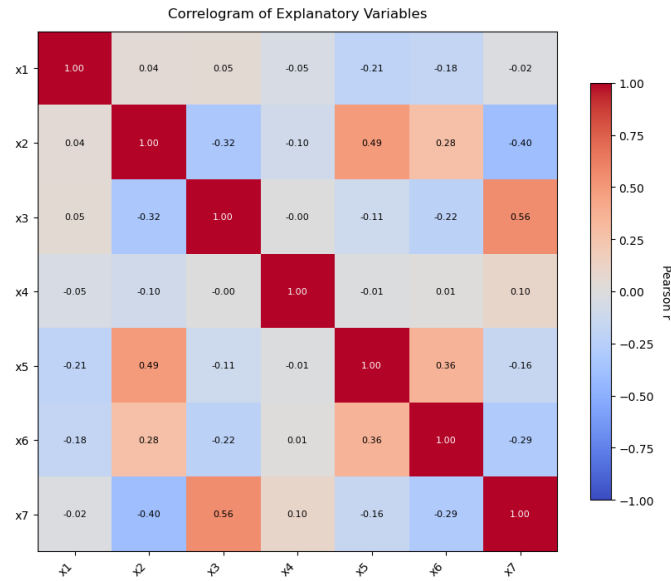
[그림 1] 분석에 활용된 종속변수 및 설명변수 단계구분도 (첫째 줄 좌측부터 합계출산율, 어린이 1000명당 어린이집 수, 인구밀도, 다문화 혼인 비율. 둘째 줄 좌측부터 인구 이동 유효도 지수, 여성 평균 초혼 연령, 평균 종합 소득 신고액, 여성 경제활동참가율)

분석에 앞서 모든 변수는 시군구 수준(일반구 제외)으로 전처리를 진행하였으며, 총  $n=229$ 개의 관측지에 대해 종속변수와 설명변수를 구성하였다. 시군구 수준을 선택한 까닭은 앞서 살펴본 것과 같이 대부분의 관련 선행 연구에서 시군구 수준을 채택하고 있기 때문에 분석 결과를 비교하기 위함이다. 먼저 종속변수로는 2024년 인구 동향 조사에서 집계된 시군구별 합계출산율(TFR)을 이용하였다. TFR의 분포 및 공간적 클러스터는 [그림 2]와 [그림 3]의 지도로 표현하였다.



[그림 2] 2024년 시군구별 TFR

[그림 3] TFR 국지 모런 클러스터 지도



[그림 4] 상관관계 Plot (변수 순서는 그림 1과 동일)

모든 설명변수는 출산력에 영향을 미치는 시차를 고려하여 전년도인 2023년에 수집된 데이터로 구성하였다. 다중공선성 진단을 위해 피어슨 상관계수([그림 4])와 VIF를 계산하였으며, 일반적으로 다중공선성이 있는 것으로 판단되는 기준인 상관계수 0.8 이상 혹은 VIF 5 이상을 나타내는 변수들은 존재하지 않았다.<sup>2)</sup>

변수명	Moran's I	VIF	mean	std
합계출산율	0.4503 (0.001)	—	0.855	0.220
아동 1000명당 어린이집 수	0.1635 (0.001)	1.104	4.091	0.836
인구밀도(명/km <sup>2</sup> )	0.8127 (0.001)	1.615	3752.649	5873.445
다문화 혼인 비율(%)	0.3413 (0.001)	1.502	12.470	4.885
인구 이동 유효도 지수	0.0088 (0.37)	1.025	-0.231	5.709
여성 평균 초혼 연령(세)	0.4307 (0.001)	1.510	31.536	0.716
평균 종합 소득 신고액(백만 원)	0.3906 (0.001)	1.254	29.259	12.567
여성 경제활동참가율(%)	0.4905 (0.001)	1.628	57.313	6.978

[표 1] 종속변수 및 설명변수 요약 통계량

2) 인구성장률, 대졸 이상 학력의 여성 비율 등 기존 문헌에서 출산력에 영향을 미치는 변수로 보고된 변수 중 다중공선성이 높은 것으로 판단되는 변수들은 확인 후 분석에 포함하지 않았다.

최종적으로 채택된 설명변수는 7개로, 크게 경제, 복지, 인구학적 프로파일 관련 세 부분으로 나눌 수 있다. 먼저 경제 관련 변수로는 평균 종합 소득 신고액과 여성 경제활동참가율을 설정하였다. 평균 종합 소득 신고액(income\_ratio)은 국세청의 <국세통계>에서 취득하였으며, 이자, 배당, 사업, 부동산임대, 근로, 연금 소득금액을 합한 뒤 신고 인원으로 나누어 계산하였다. 1인당 국내총생산 혹은 1인당 국민총소득에 비해 모집단이 편향되어 있을 수 있다는 단점이 있으나, 법인의 미분배 이익이나 외국인 소득 등을 포함하지 않으므로 실질적 생활 수준의 척도로서는 더 나은 점이 있다고 평가하였다. 평균 종합 소득 신고액은 대체로 출산율에 음의 영향력을 미칠 것으로 예상되나, 최근 연구에서는 인간 개발 지수가 높은 선진국에서는 관계가 역전되는 J-자 형태의 커브가 나타나는 사실(Myrskylä et al., 2009)이 알려져 있어 그러한 관계가 미시적 스케일인 시군구 수준에서도 나타날지 주목해 볼만한 부분이다. 여성 경제활동참가율(econ\_ratio)은 통계청의 <경제활동인구조사>에서 취득하였고, 15세 이상 인구 중 경제활동인구의 비율이다. 전국의 여성 경제활동참가율은 57.2%로 남성의 73.9%에 비해 낮은 수치이다. 여성 경제활동참가율은 단순히 가구당 소득 증가 측면의 경제적 요인일 뿐만 아니라 정책적 요소와도 연관성을 지닌다고 볼 수 있는데, OECD(2024)에서는 일·가정 양립 및 성평등 제도의 확충이 이루어지면 여성 경제활동참가율이 출산력과 양의 상관관계를 보인다고 보고한 바 있다.

복지 혹은 정책 관련 변수로는 아동 1000명당 어린이집 수(cc\_ratio)를 활용하였다. 출처는 교육부의 <보육통계>이며, 모든 종류(국공립, 민간, 직장 등)의 어린이집 수를 합산한 뒤 시군구별 7세 이하 아동 수로 나누어 계산하였다. 이는 지자체의 보육 서비스의 질적 수준을 정량적으로 보여줄 수 있는 지표이다. 그 밖에도 II-3에서 언급된 선행 연구들에서 다루고 있는 출산장려금 혹은 보육장려금을 정책 변수로 살펴볼 필요성이 있겠으나, 상이한 지급 기간, 대상, 금액을 일관적으로 정량화하는 데 어려움이 있어 본 연구에서는 활용되지 못했다.

마지막으로 인구학적 프로파일 관련 변수로 인구 밀도, 다문화 혼인 비율, 인구 이동 유효도 지수, 여성 평균 초혼 연령을 활용하였다. 인구 밀도와 다문화 혼인 비율, 여성 평균 초혼 연령은 통계청의 <인구동향조사>에서 취득하였고, 인구 이동 유효도 지수는 통계청의 <국내인구이동통계>에서 순 이동이 총이동에서 차지하는 비율로 계산하였다. 인구 밀도(pop\_den)와 인구 이동 유효도 지수(mei)는 인구 분포 및 인구 이동이 출산력에 미치는 영향력을 살펴보는 데 사용되었다. 특히 인구 이동 유효도 지수는 순이동률에 비해 개념적 타당성, 소인구단위에서 추정의 강건성, 역동성의 측정 측면에서 더 나은 점이 있어 (이상일·이소영, 2023) 사용되었고, 이를 출산력 연구에서 활용하는 것은 최초의 시도인 것으로 보인다. 한편 [표 1]에서 확인할 수 있듯이 인구 이동 유효도 지수는 유일하게 전역적 Moran's I가 유의하지 않은 변수로, 높은 공간적 이질성을 지닌 변수가 투입되었을 때 SGWR이 어떻게 모델링하는지 주목해 볼 가치가 있다.

다문화 혼인 비율(int\_marriage\_rate)과 여성 평균 초혼 연령(marriage\_age)은 출산력과 밀접한 관련을 지닌 인구학적 변수이다. 한편, 두 변수의 영향력이 공간적으로 어떤 패턴을 보이는지 분석한 연구는 많지 않다. 다문화 혼인 비율이 출산력에 미치는 영향력은 도시와 농촌 지역에서 상이할 가능성이 있다. 국내 다문화 혼인 중 농촌 지역에서는 주로

개발도상국 출신 여성의 ‘결혼 이주’ 형태가 큰 비중을 차지하고, 그 경우 노동력 수요(Cho, 2018)와 본국의 높은 출산율 규범(고원식, 2019)이 더 높은 출산력으로 이어진다는 연구 결과가 있는 한편 도시 지역에서는 다문화 혼인의 양상이 다소 다를 수 있기 때문이다. 여성 평균 초혼 연령은 지연된 혼인이 ‘혼인-후-출산’ 관념이 강하게 자리 잡은 동아시아 국가에서 낮은 출산율의 원인으로 분석되고 있는 한편 (Jung and Lee, 2023; Fukuda, 2020), 인간 개발 지수가 0.9 이상인 유럽 일부 선진국에서는 초혼 연령과 출산율이 반드시 부(-)의 관계를 보이지 않는 것으로 나타나 있어(OECD, 2024) 국내 시군구 수준에서 어떤 공간적 이질성을 보이는지 확인해 보고자 선택되었다.

## 2. GWR 및 파생 모형

본 연구에서 사용된 SGWR 모형을 살펴보기에 앞서 간략히 GWR 및 파생 모형의 구조를 서술한다. 이번 장에서 언급된 모형은 IV-3에서 모형 적합 결과 비교를 위해 사용될 것이다. 먼저 GWR 모형의 관측지( $s_i$ )별 데이터 생성 프로세스(DGP)는 아래 식과 같다.

$$y_i = \sum_{j=0}^k \beta_{ij}(u_i, v_i) x_{ij} + \epsilon_i$$

이때 회귀계수  $\beta = (\beta_{ij})(n \times k)$ 의 추정에는 가중 최소제곱법(weighted least squares)을 통해 이루어진다. 추정된  $\mathbf{y}(n \times 1)$ 를 투영 행렬  $\mathbf{S}(n \times n)$ 를 이용하여 표현할 경우 다음과 같다.

$$\hat{\mathbf{y}} = (\hat{\beta} \otimes \mathbf{X}) \mathbf{1} = \mathbf{S} \mathbf{y}$$

$$\text{where } \mathbf{S} = \begin{bmatrix} \mathbf{x}_1 (\mathbf{X}^\top \mathbf{W}_{s_1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{s_1} \\ \mathbf{x}_2 (\mathbf{X}^\top \mathbf{W}_{s_2} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{s_2} \\ \vdots \\ \mathbf{x}_n (\mathbf{X}^\top \mathbf{W}_{s_n} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{s_n} \end{bmatrix}$$

관측지별 가중치 매트릭스  $\mathbf{W}_{s_i}$ 는 대각행렬로, 각 성분은 커널 함수를 통해 계산된 관측지별 가중치가 된다. 이때 커널 함수에 사용되는 대역폭은 AICc를 기준으로 Golden Section 방식 등을 이용하여 반복적으로 적합 후 최적값을 도출하는 방식을 취하는 것이 일반적이다.

MGWR 모형은 GWR의 가정을 완화하여 설명변수 별로 다른 크기의 대역폭을 추정한다. 따라서 회귀계수 추정은 변수별 대역폭  $bw_j$ 에 의존적이며 관측지별 DGP는 아래와 같다.

$$y_i = \sum_{j=0}^k bw_j (\beta_{ij}(u_i, v_i) x_{ij}) + \epsilon_i$$

MGWR은 가능한 변수별 대역폭 조합의 개수가 많아 GWR과 같은 탐색 방법을 적용하기는 어렵다. 따라서 모형을 GAM(Generalized Additive Model)의 일종으로 생각하여 GAM 적합에 이루어지는 Back-Fitting 알고리즘을 이용하여 대역폭 탐색과 회귀계수 추정을 동시에 진행한다. (Fotheringham et al., 2022)

### 3. SGWR로의 확장

#### 1) SGWR 모형 식별 및 추정 절차

SGWR 모형의 추정을 위한 투영 행렬 구조는 GWR 모형과 동일하다. 다만 관측지별 가중치 매트릭스  $\mathbf{W}_{s_i}$ 를 다음과 같이 재구성한다. 아래 식은 가중치 커널로 가우시안 커널을 채택했을 때의 경우이다.

$$\begin{aligned}\mathbf{W}_{s_i} &= \alpha \times \mathbf{W}_{s_i}^G + (1 - \alpha) \times \mathbf{W}_{s_i}^S \\ \mathbf{W}_{s_i}^G &= \text{diag}(w_{i1}^G, \dots, w_{in}^G) \text{ where } w_{ij}^G = \exp\left(-\frac{1}{2}\left(\frac{d_G(i,j)}{b_G}\right)^2\right) \\ \mathbf{W}_{s_i}^S &= \text{diag}(w_{i1}^S, \dots, w_{in}^S) \text{ where } w_{ij}^S = \exp\left(-\frac{1}{2}\left(\frac{d_S(i,j)}{b_S}\right)^2\right)^3\end{aligned}$$

$\alpha$ 는 영향 비(impact ratio)로, 지리적 근접성 기반 가중치와 속성 유사성 기반 가중치의 비중을 조정하는 파라미터이며 0과 1 사이의 값을 지닌다.  $\alpha=1$ 인 경우 GWR 모형이 된다.

기존 문헌에서는 별도의 정당화 없이  $\mathbf{W}_{s_i}^S$ 의 거리 척도로 맨해튼 거리를 사용하였으나, 본 연구에서는 마할라노비스 거리를 채택하였다. 그 이유는 두 가지로, 첫째는 마할라노비스 거리를 채택한 이유는 변수의 단위와 무관하므로 변수 간 스케일 차이가 큰 경우 보다 안정적이기 때문이다. 둘째는 마할라노비스 거리는 변수 간의 상관 구조를 반영하므로 중복 정보가 많은 (즉, 공분산이 큰) 변수에 대해서는 상대적으로 가중치가 낮아 다중공선성을 완화할 수 있다는 장점이 있기 때문이다.

SGWR 모형에서 추정되어야 할 파라미터는  $\mathbf{b}=(\alpha, b_G, b_S)^\top$ 로 세 가지이다. 따라서 기존 GWR 모형에 비해 계산 복잡도가 높다. Yu et al.(2025)에서는 경사 하강 알고리즘 (gradient descent algorithm)을 이용하여 최적해를 찾는 방법을 제시하고 있다. 평가 척도로는 GWR 모형에서 사용되는 AICc를 이용한다. (Fotheringham et al., 2022)

$$\text{AICc}(\mathbf{b}) = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left(\frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})}\right)$$

이고  $\hat{\sigma}^2 = \frac{\text{RSS}}{n}$  이므로 그래디언트 벡터는 다음과 같이 계산된다.

$$\begin{aligned}\frac{\partial \text{AICc}(\mathbf{b})}{\partial \mathbf{b}} &= \frac{1}{\hat{\sigma}^2} \frac{\partial \text{RSS}(\mathbf{b})}{\partial \mathbf{b}} + \frac{2n(n-1)}{(n-2-\text{tr}(\mathbf{S}))^2} + \frac{\partial \text{tr}(\mathbf{S})}{\partial \mathbf{b}} \\ \text{where } \frac{\partial \text{RSS}(\mathbf{b})}{\partial \mathbf{b}} &= -2 \sum_i (y_i - \hat{y}_i) \frac{\partial \hat{y}_i}{\partial \mathbf{b}} = -2 \begin{bmatrix} \left(\frac{\partial \hat{\mathbf{y}}}{\partial b_G}\right)^\top \\ \left(\frac{\partial \hat{\mathbf{y}}}{\partial b_S}\right)^\top \\ \left(\frac{\partial \hat{\mathbf{y}}}{\partial \alpha}\right)^\top \end{bmatrix} (\mathbf{y} - \hat{\mathbf{y}}),\end{aligned}$$

3) Yu et al.(2025)의 식(7)에는 속성 유사성 기반 가중치 커널이 지리적 근접성 기반 가중치 커널과 다르게 정의되어 있다. 그러나 뒷부분의 AICc 그래디언트 계산식을 확인하였을 때 이는 오키로 추정된다. 따라서 본고에서는 두 가중치의 커널을 동일하게 가우시안 커널로 설정하였다.



$$\frac{\partial \text{tr}(\mathbf{S})}{\partial \mathbf{b}} = \frac{\partial}{\partial \mathbf{b}} \sum_i \mathbf{x}_i [(\mathbf{X}^\top \mathbf{W}_{s_i} \mathbf{X})^{-1} \mathbf{x}_i] = \begin{bmatrix} (\frac{\partial \text{tr}(\mathbf{S})}{\partial b_G})^\top \\ (\frac{\partial \text{tr}(\mathbf{S})}{\partial b_s})^\top \\ (\frac{\partial \text{tr}(\mathbf{S})}{\partial \alpha})^\top \end{bmatrix}$$

한편, 지리적 근접성 대역폭( $b_G$ )을 적응형으로 설정하고자 할 경우 이산형이 되므로 위 식을 그대로 이용할 수 없다. 따라서 관측지별로  $r$ 번째로 가까운 관측지와와의 거리를 5차 다항식  $V_{(r)} \simeq a_{i0} + a_{i1}r + a_{i2}r^2 + a_{i3}r^3 + a_{i4}r^4 + a_{i5}r^5$  으로 적합하고, 연쇄 법칙(chain rule)을 이용하여 위 그래디언트 벡터 중  $b_G$ 에 대한 그래디언트를 다음과 같이 변형한다. 계산을 위한 자세한 수식은 Yu et al.(2025)에서 참고하였다.

$$\begin{aligned} \frac{\partial \text{AICc}(b_G)}{\partial b_G} &= \frac{1}{\hat{\sigma}^2} \frac{\partial \text{RSS}(b_G)}{\partial b_G} + \frac{2n(n-1)}{(n-2-\text{tr}(\mathbf{S}))^2} + \frac{\partial \text{tr}(\mathbf{S})}{\partial b_G} \\ \text{where } \frac{\partial \text{RSS}(b_g)}{\partial b_g} &= -2 \sum_i (y_i - \hat{y}_i) \frac{\partial \hat{y}_i}{\partial V_{(b_g)}} \frac{dV_{(b_g)}}{db_G} = -2 \left( \frac{\partial \hat{y}_i}{\partial V_{(b_g)}} \right)^\top (\mathbf{y} - \hat{\mathbf{y}}) \frac{dV_{(b_g)}}{db_G}, \\ \frac{\partial \text{tr}(\mathbf{S})}{\partial b_G} &= \frac{dV_{(b_g)}}{db_g} \frac{\partial}{\partial V_{(b_g)}} \sum_i \mathbf{x}_i [(\mathbf{X}^\top \mathbf{W}_{s_i} \mathbf{X})^{-1} \mathbf{x}_i] \end{aligned}$$

본 연구에서는 가우시안 커널을 채택하였고, 지리적 근접성 대역폭은 적응형으로, 속성 유사성 대역폭은 고정형으로 설정하여 분석을 진행하였다.

경사 하강 알고리즘의 수렴 속도를 높이기 위해 현재 여러 가지 방법론이 개발되어 있으며, Zhou et al.(2023)은 MGWR의 경우 L-BFGS 등의 준-뉴턴(quasi-Newton) 방법 대비 신뢰 영역(trust-region) 방법이 가장 좋은 최적해에 도달했다고 보고한 바 있다. Yu et al.(2025)는 이를 인용하여 SGWR에서도 신뢰 영역 방법을 채택하였으나 본 연구에서는 그 밖에도 Golden Section 방법(Kiefer, 1953)을 결합하는 접근과 Differential Evolution 방법(Storn and Price, 1997)을 활용한 접근을 추가로 시도하였다.

## 2) 통계적 검정

GWR의 경우  $k$ 개의 파라미터에 대해  $n$ 개의 회귀 모형을 적합함에 따라 다중 검정 문제가 발생하며, 이는 FWER(Family-Wise Type I Error Rate)을 많이 증가시킨다. 따라서 일반적인 경우에 비해 보수적인 유의수준을 설정해야 한다. 그뿐만 아니라 관측지별로  $n$ 개의 회귀 모형을 적합함에 따라 다중 검정이 비독립적으로 이루어진다는 점 또한 고려해야 한다. 그러나 대부분의 문헌에서는 이러한 다중 검정 문제를 간과하는 경우가 많고, 이는 국지 회귀계수 추정 결과의 신뢰도를 저하한다. da Silva and Fotheringham(2016)은 이러한 문제를 해소하고자 조정된 유의수준을 도출하는 방안을 제시하였으며, 이는 아래 식과 같다.

$$\alpha^* = \frac{\xi}{\text{ENP}/k}$$

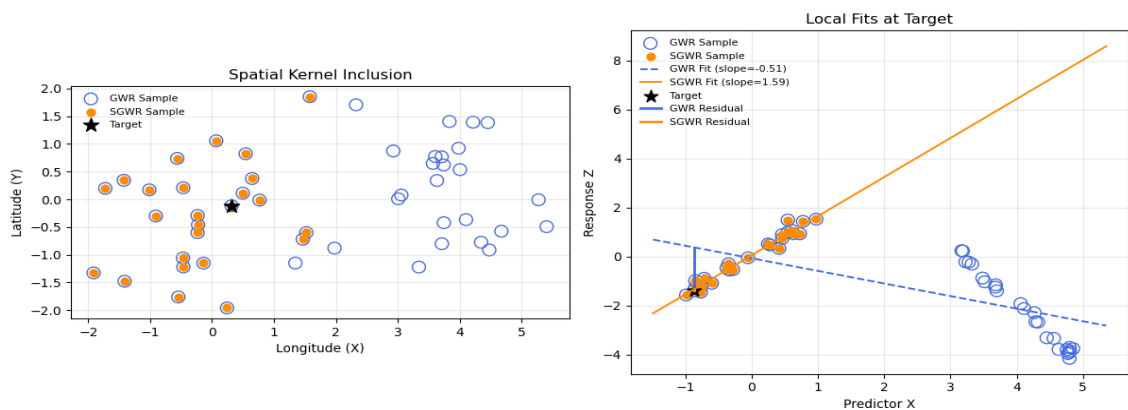
ENP는 유효 파라미터 수(effective number of parameters)로 투영 행렬 대각성분의 합,  $\text{tr}(\mathbf{S})$ 이며  $\xi$ 는 기대되는 제1종 오류율,  $k$ 는 실제 파라미터의 개수이다. SGWR은 GWR과 회귀 모형의 구조가 같으므로, 동일한 방법을 통해 다중 검정 문제를 해소할 수 있다.

국지 회귀계수의 검정 통계량은  $t_{ki} = \frac{\hat{\beta}_{ki} - \beta_{ki}}{se(\beta_{ki})}$ 이며 귀무가설  $\beta_{ki} = 0$  하에서 자유도가  $n - \text{ENP} - 1$ 인  $t$ -분포를 따른다.  $P(|t_{ki}| > c_{\alpha^*}) = \alpha^*$ 인 임계값  $c_{\alpha^*}$ 에 대해  $|t_{ki}| > c_{\alpha^*}$ 인 경우 귀무가설을 기각하고 해당 국지 회귀계수의 값이 통계적으로 유의한 것으로 판단한다.

### 3) (S)GWR과 심슨의 역설

SGWR이 GWR에 비해 지니는 장점은 잘 알려진 심슨의 역설(Simpson's Paradox) 상황에서 두드러진다. 심슨의 역설은 하위 집단에서 나타나는 변수 간 관계가 전체 데이터에서 사라지거나 역전되는 현상을 말한다. GWR에서 심슨의 역설이 나타날 수 있는 이유는, 지리적 근접성 가중치를 통해 데이터를 빌려 오는 집단의 설정이 기본적으로 임의적이기 때문이다. 따라서 이는 MAUP의 일종이기도 하다. (Sachdeva and Fotheringham, 2023) 예를 들어 해당 집단의 특성이 가정과 달리 이질적인 경우, 가중치로 빌려 온 정보는 오히려 관계를 왜곡해 전체 혹은 다른 집단과 역전된 추정치 및 편향을 초래한다. 한편 SGWR의 경우 속성 유사성 가중치를 포괄함으로써 속성 공간에서 가까이 있는 정보를 활용할 수 있다. 이 경우 일반적으로 추정량의 변동성은 증가할 수 있지만, Bias-Variance Trade-Off에 따라 편향을 낮춤으로써 모형의 전체적인 적합도는 증가할 수 있는 것이다. 물론 이러한 개선이 모든 데이터에 적용된다고 일반화하기는 아직 어렵다. 그러나 이론적으로는 영향 비 및 대역폭 파라미터의 조정을 통해 설명력이 가장 높은 모형으로 자연스럽게 수렴하게 되기 때문에 (예를 들어  $\alpha=1$ 인 경우 GWR), SGWR은 기존 GWR의 유연성(flexibility)을 향상시킨 확장된 모형이면서도 심슨의 역설에 빠지지 않게 하는 장점이 있게 되는 것이다.

아래 [그림 5]는 심슨의 역설이 발생할 수 있는 상황을 모식적으로 나타낸 것이다. 물론 극단적인 경우이기는 하지만, GWR 가중치(파란색)를 부여받은 관측치들로 이용해 회귀직선을 적합했을 때와 비교해 SGWR 가중치(주황색)를 부여받은 관측치들로 이용해 회귀직선을 적합했을 때 직선의 기울기가 달라지는 것은 물론 타겟 지점에서의 잔차의 편향이 줄어드는 것을 확인할 수 있다.



[그림 5] SGWR 가중치 적합에 따른 국지 회귀 모형의 변화

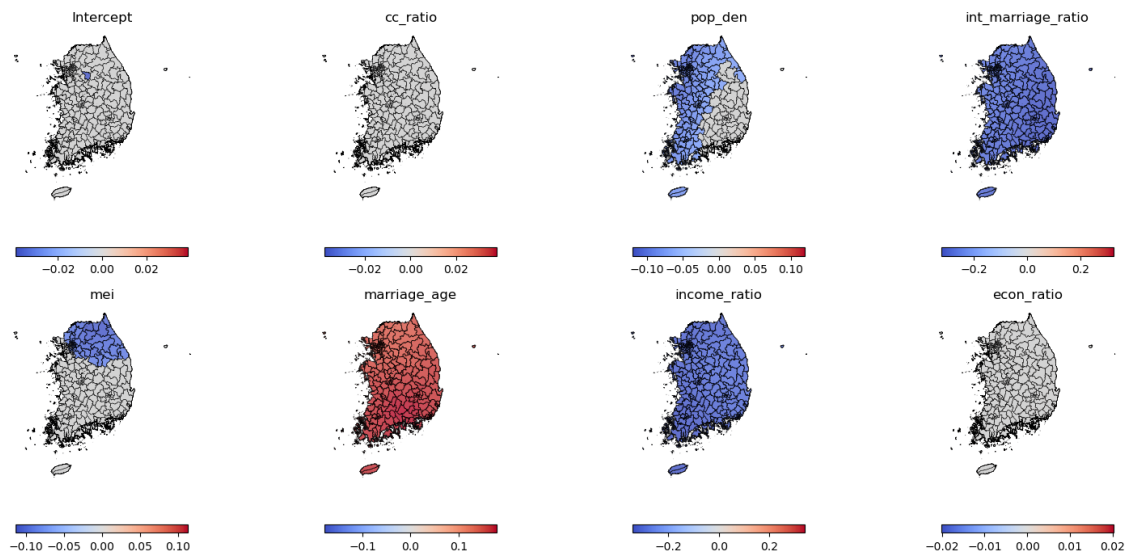
다른 관점에서 보면 심슨의 역설은 어떤 교란 변수(편의를 위해 범주형이라고 가정)가 존재할 때 교란 변수로 조건화하지 않고 (집단을 섞어서) 분석할 때 종속 변수와 설명 변수 간의 관계가 약화되거나 역전되는 것이다. GWR의 맥락에서 교란 변수가 제대로 통제되지 않는 상황은 교란 변수가 공간적으로 군집되어 형성된 것이 아니라 흩어져 있을 때 발생하게 된다. 예를 들어 읍면동 수준으로 분석할 때 대학이 위치한 공간 단위는 주변 공간 단위에 비해 높은 청년층 인구 비중이 높아 불연속적인 ‘섬’으로 나타나는 경우가 있다. GWR은 지리적 거리만으로 이웃을 정의하므로, 이런 대학 지역이 주변(청년 비중이 낮은) 구역과 함께 커널에 포함돼 청년층 인구 비중과 종속 변수의 관계가 희석되기 쉽다. 반면 SGWR은 공간 거리와 속성 거리를 동시에 고려해 가중치를 계산하므로, 지리적으로 멀리 떨어져 있어도 대학이 있는 구역들끼리는 속성상 유사한 집단으로 묶인다. 결과적으로 이러한 ‘공간적 이상치(outlier)’의 영향이 과소-또는 과대평가되지 않고, 패턴이 더 정확히 추정된다.

#### IV. SGWR 적합 결과

##### 1. 모형 적합도 진단

##### 1) 탐색 방법에 따른 민감도 분석

우선 Yu et al.(2025)을 따라 Trust-Region 방법으로 최적해를 탐색하였고, 해당 파라미터를 이용해 SGWR 모형을 적합하였다.



[그림 6] Trust Region 방법으로 추정된 국지 회귀계수 (임계값 = 1,881)

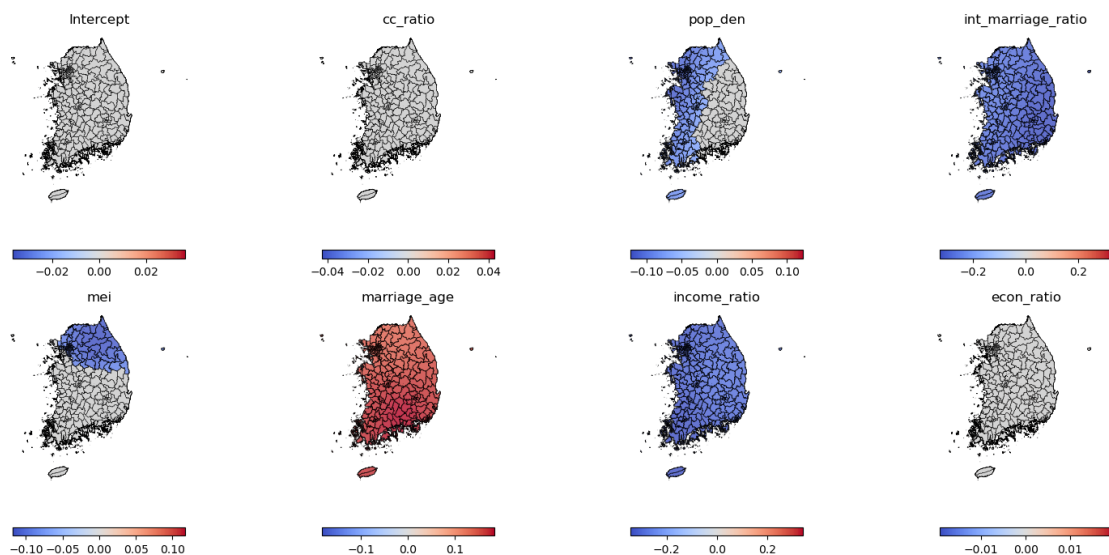
그러나 AICc를 계산한 결과 480.020으로 전통적 GWR로 적합했을 때의 AICc에조차 미치지 못하는 결과였다. 즉, Trust Region 방법 역시도 초깃값의 설정에 따라 국소 최적(local optima)에 갇히는 문제를 피할 수 없음이 드러난 것이다. 이를 보다 구체적으로 확인하고자  $\alpha$ 의 범위를 고정한 뒤 Trust Region 방법으로 탐색을 진행하면서 AICc의 변화를 추적하였고, 그 결과는 [표 2]와 같다.

$\alpha$	지리적 근접성 대역폭( $b_G$ )	속성 유사성 대역폭( $b_S$ )	AICc
0.99	134.785149	2.76887687	470.027
0.70	111.942659	2.75724589	471.346
0.50	95.1822142	2.70449481	476.936
0.30	79	2.64575131	489.411
0.10	228	0.1	492.773
0.05	228	0.1	493.108

[표 2]  $\alpha$ 를 고정한 뒤 Trust Region 방법으로 탐색한 결과

Trust Region 방법의 최적해와 달리  $\alpha$ 가 0.99일 때 AICc가 더 낮은 것을 확인할 수 있었으며, 단순히 Trust Region 방법만을 이용할 경우 실제 최적해가 아닌 곳으로 수렴하는 결과가 나타날 수 있었다. 이는 변수별 대역폭을 추정하는 MGWR과 달리 SGWR의 경우 추정되는 파라미터 간의 종속성이 존재하여 탐색 공간이 훨씬 더 복잡하다는 점에서 기인한다. 특히  $\alpha$ 와  $b_S$  간에는 구조적 관계가 관찰되었는데, 예를 들어  $\alpha$ 의 초깃값이 0에 가깝게 설정된 경우 속성 유사성 가중치의 비중이 커지므로 해당 대역폭  $b_S$ 가 극단적으로 줄어들면서 과적합이 발생하고 알고리즘이 수렴하는 것으로 판단을 내릴 수 있는 것이다.

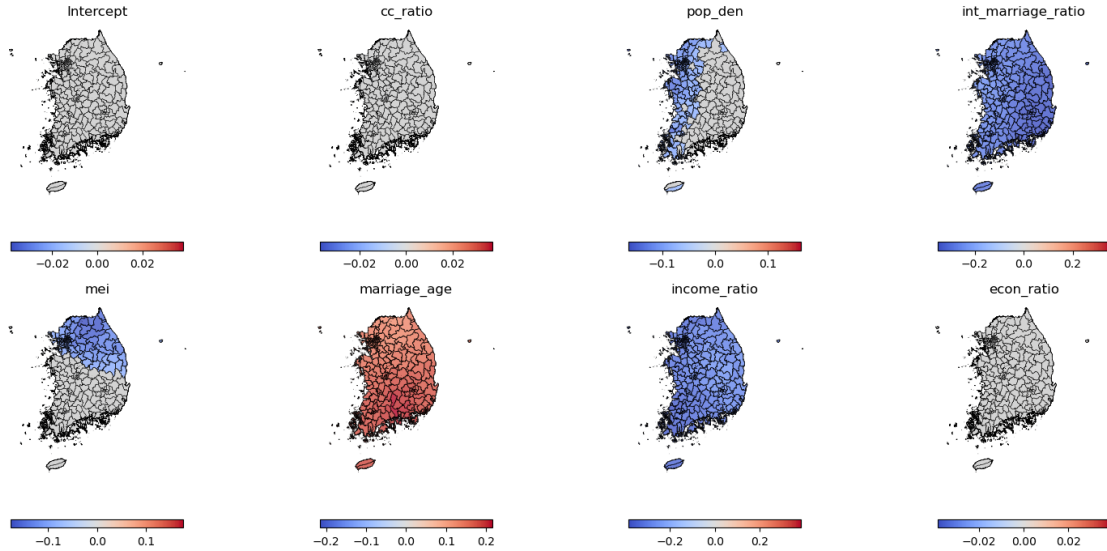
국소 최적에 갇히는 문제를 방지하고자 두 가지 탐색 방법을 추가로 시도하였다. 첫째는  $\alpha$ 를 Golden Section 방법으로 분할하면서 해당 구간 중 하나로 고정한 뒤, 나머지 대역폭 파라미터를 L-BFGS<sup>4)</sup> 방법으로 탐색하는 것이다. Golden Section 방법을 이용한 탐색은 비교적 파라미터 공간이 단순한 1차원 혹은 2차원에서 진행되는데, 실제로 SGWR의 초기 버전을 고안한 Lessani and Li(2024)에서도 채택한 방법이기도 하다.



[그림 7]  $\alpha$ 에 대한 Golden Section 방법으로 추정된 국지 회귀계수 (임꺽값 = 1.779)

4) Trust-Region 방식 또한 이용할 수 있으나 연산량이 가중되어 우선 L-BFGS 방식을 적용하였다.

둘째는 유전 알고리즘의 구조를 차용한 Differential Evolution 방법을 활용하는 것이다. 먼저 Differential Evolution 알고리즘을 통해 넓은 파라미터 공간을 군집 단위로 탐색한 뒤 초깃값 후보를 확보하고, 각각에 대해 L-BFGS 방법으로 2차 탐색을 진행한다.



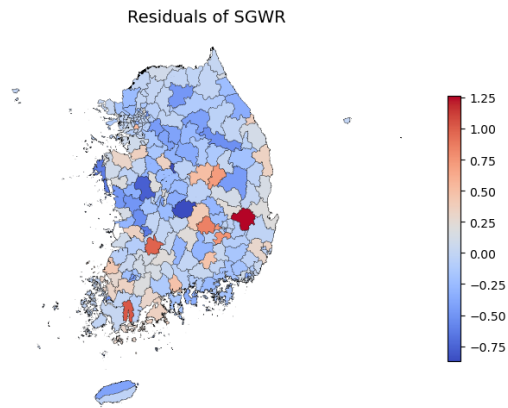
[그림 8] Differential Evolution 방법으로 추정된 국지 회귀계수 (임궛값 = 1.804)

탐색 방법	$\alpha$	$b_G$	$b_S$	AICc	RSS	조정 $R^2$	연산 시간
L-BFGS	0.500	95.18	2.70	545.363	53.496	0.6828	50s
Trust Region	0.498	111.53	2.83	480.020	45.912	0.7110	1m 28s
Golden Section( $\alpha$ ) + L-BFGS	0.759	118.09	2.78	470.730	54.178	0.6835	6m 41s
Differential Evolution + L-BFGS	0.500	111.37	1.14	469.537	18.517	0.894	2m 47s

[표 3] 탐색 방법에 따른 최적해 및 SGWR 모형 적합 결과

[표 3]은 최종적으로 탐색 방법에 따른 최적해 및 AICc, RSS(residual sum of squares), 조정  $R^2$  값을 정리한 것이다. 연산 시간은 5회 반복 시행 후 평균값으로 기재하였다. 탐색 방법에 따라 얻어진 최적해가 어느 정도 유사한 지점으로 수렴하기는 하나, 일부 파라미터는 탐색 방법에 따른 차이가 큰 것을 확인할 수 있다. 따라서 여전히 국소 최적에 갇히는 문제가 해결되었는지에 대한 의심을 지우기는 어려우며 최적의 탐색 방법을 정하기 위해서는 추가적인 방법론적 정련화가 필요해 보인다. 본 연구에서는 AICc 및 기타 적합도 지표에서 우수한 성능을 보인 Differential Evolution 방법으로 찾은 최적해로 SGWR 모형을 적합하여 아래 분석 및 결과 해석을 진행하였다.

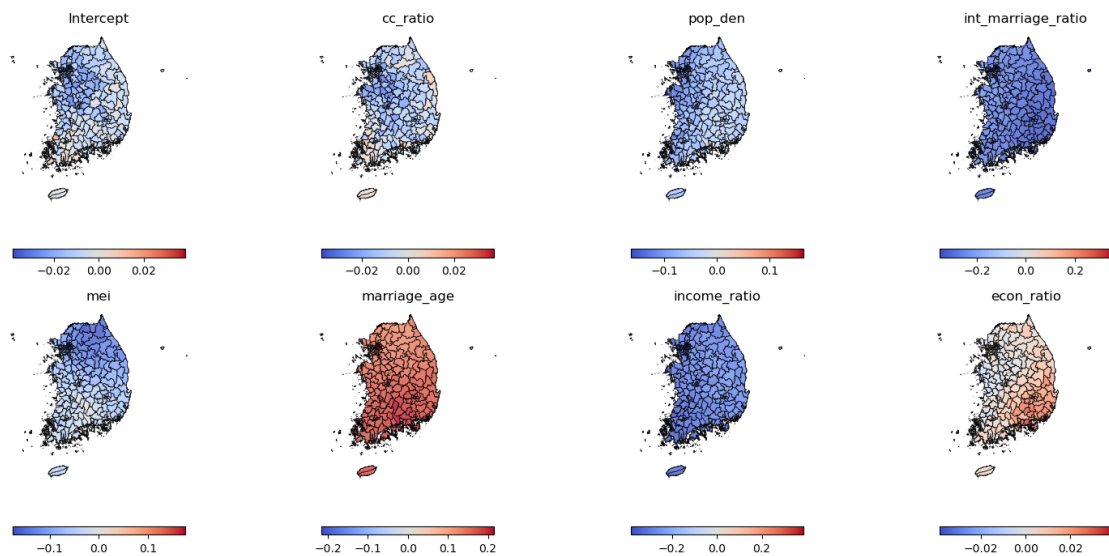
## 2) 잔차 진단



[그림 9] SGWR 모형의 잔차 지도

[그림 9]는 적합 된 SGWR 모형의 잔차를 지도로 나타낸 것이다. 전역적 Moran's I는 0.0511로 999회의 순열 기반 검정에 따르면 Pseudo p-value가 0.0930으로 유의수준 0.05에서 귀무가설을 기각할 수 없고, 따라서 잔차의 공간적 자기상관이 존재하지 않음이 확인된다.

## 2. 지역별 회귀계수 추정치 해석



[그림 10] SGWR 모형으로 추정된 회귀계수 (전체 표시)

[그림 10]은 선택된 SGWR 모형으로 추정된 회귀계수를 지도로 나타낸 것이다. [표 4]에서도 확인할 수 있듯이 인구 밀도, 다문화 혼인 비율, 평균 종합 소득 신고액은 전 지역에서 회귀계수가 음수로 나타났으며, 평균 초혼 연령은 전 지역에서 회귀계수가 양수로 나타났다. 이외 변수들은 음수와 양수 회귀계수가 혼재하는 경향을 보였다.

변수	Mean	Std,	Min.	Max.	Sig. Rate
Intercept	-0.0113	0.0089	-0.0383	0.0170	0%
cc_ratio	-0.0118	0.0088	-0.0372	0.0116	0%
pop_den	-0.0788	0.0295	-0.1638	-0.0201	51.5%
int_marriage_ratio	-0.2938	0.0327	-0.3527	-0.2433	100%
mei	-0.0762	0.0413	-0.1758	0.0050	42.4%
marriage_age	0.1632	0.0217	0.1241	0.2156	100%
income_ratio	-0.3083	0.0250	-0.3838	-0.2545	100%
econ_ratio	0.0086	0.0104	-0.0095	0.0387	0%

[표 4] 지역별 회귀계수 추정치 요약 통계량

한편 [표 4]의 Sig. Rate 열은 임계값 이상  $t$ -값을 나타낸, 즉 추정된 회귀계수가 통계적으로 유의미한 것으로 나타난 관측지의 비율이다. 다문화 혼인 비중, 초혼 평균 연령, 평균 종합 소득 신고액은 모든 지역에서 유의미한 영향력이 있는 것으로 나타났다. 인구 밀도와 인구 이동 유효도 지수는 각각 51.5%, 42.4%의 관측지에서만 유의한 영향력을 보였다. 이외 변수는 전 지역에서 유의한 회귀계수가 도출되지 않았다. [그림 7]을 통해 공간적 분포를 살펴보면 인구 밀도는 경기도, 충청남도, 전라북도, 전라남도에서, 인구 이동 유효도 지수는 경기도와 강원도를 중심으로 영향력의 크기가 컸다. 회귀계수 추정치의 분포에 대한 해석은 IV-3에서 GWR 적합 결과와의 비교를 통해 자세하게 기술하였다.

### 3. 기타 모형과의 비교

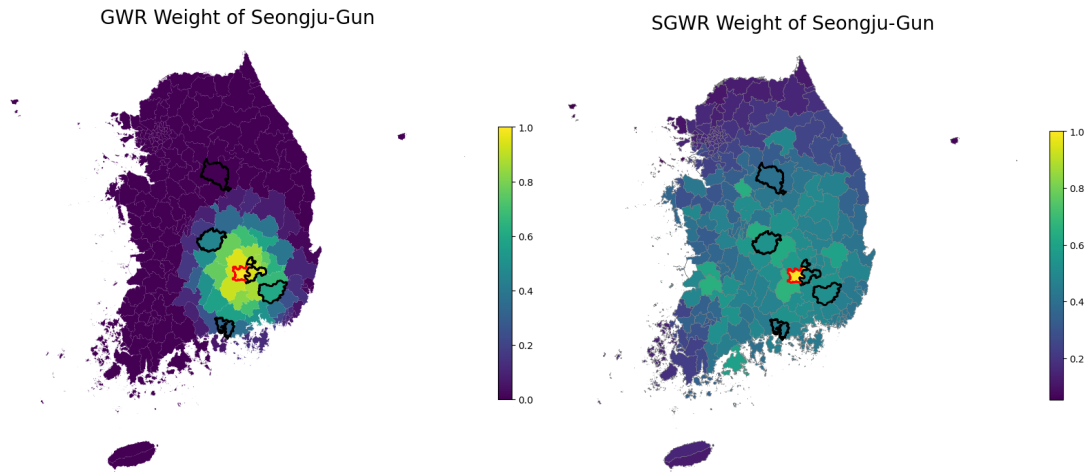
SGWR 모형의 적합도 및 회귀계수 추정치를 평가 혹은 해석하기 위해 다른 모형과의 비교를 시행하였다. 적합도 평가 지표로는 AICc, RSS, 조정  $R^2$ ,  $\hat{y}_i$ 에 대한 RMSE를 이용하였으며 그 결과는 아래 [표 5]에 나타나 있다. 기존 SGWR 관련 선행 연구에서 경험적으로 나타난 경향성과 동일하게 OLS, GWR, MGWR 모형에 비해 SGWR 모형이 모든 지표에서 적합도가 우수한 것으로 확인되었다. 특히 RMSE 측면에서는 상당한 수준의 개선을 보였다.

모형	AICc	RSS	조정 $R^2$	RMSE
OLS	495.325	107.919	0.512	0.688
GWR ( $b_G = 79$ )	488.013	97.357	0.542	0.676
MGWR	579.123	92.410	0.475	0.724
SGWR (Lessani and Li, 2024) <sup>5)</sup>	478.073	77.646	0.604	0.629
SGWR	466.242	17.614	0.898	0.278

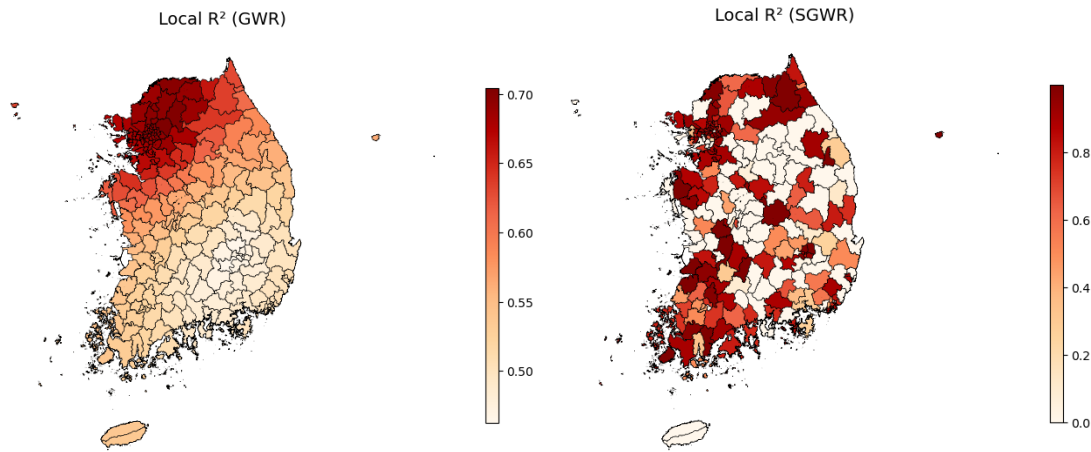
[표 5] OLS, GWR, MGWR, SGWR 모형의 적합도 비교

5) 해당 모형은 본 연구에서 사용된 SGWR 모형과 달리 지리적 근접성 대역폭과 영향 비만을 파라미터로 설정한다. 모형의 적합은 Lessani(2025)의 Python sgwr 0.2 패키지를 이용하였으며, 최종적으로 선택된 지리적 근접성 대역폭은 122, 영향 비는 1에 가까운 값이었다.





[그림 11] GWR 모형의 가중치 (성주군) [그림 12] SGWR 모형의 가중치 (성주군)



[그림 13] GWR 모형의 국지  $R^2$

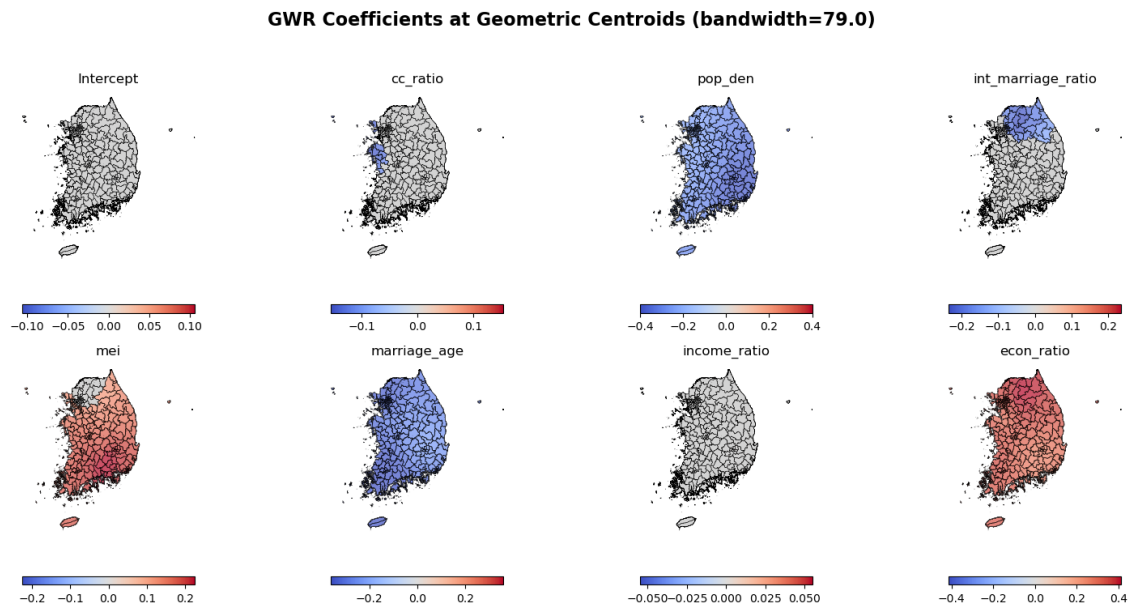
[그림 14] SGWR 모형의 국지  $R^2$

[그림 11-14]는 GWR 모형에서 속성 가중치를 반영하는 것이 어떠한 변화를 불러 오는지를 시각적으로 잘 보여준다. [그림 11]은 경상북도 성주군에서 적합 되는 GWR 모형의 가중치로 성주군을 중심으로 동심원 모양으로 가중치가 단계적으로 감소하는 것을 확인할 수 있다. 한편, 검은색 테두리로 표시된 지역은 본 연구에서 활용된 7개의 설명변수를 기반으로 속성-마할라노비스 거리를 계산하였을 때 가장 거리가 가까운 상위 5개의 지역(경남 거제, 충북 충주, 충북 영동, 경남 김해, 대구 달성)이다. 그러나 GWR 모형의 가중행렬은 오로지 물리적 거리에만 의존하기 때문에 해당 지역의 가중치는 상대적으로 낮으며, 심지어 충북 충주시의 가중치는 0이다. 대신 지리적으로 근접한 대구광역시의 지자체가 대부분 높은 가중치를 부여받는데, ‘군’과 ‘구’의 차이를 떠올려 보면 성주군과는 크게 다른 특성의 지역이라 할 수 있다. 극단적인 경우 심슨의 역설과 같이 해당 반경을 일종의 이웃 집단으로 고려하는 것이 교란 변수로 작용하여 출산력에 영향을 미치는 변수들 혹은 영향력의 방향이 실제와 달라질 수 있는 것이다. 한편 [그림 12]과 같이 SGWR 모형의 가중행렬은 영향 비 0.5로 속성 유사성을 고려하기 때문에 검은색으로 표시된 지자체들의 가중치가 상대적으로 높아진 것을 확인할 수 있다. 단순히 속성 유사성을 고려하였다고 해서 교란 변수



의 개입을 방지하였다고 정당화할 수는 없지만, Bias-Variance Trade-Off에 의해 SGWR 모형의 추정치는 일반적으로 편향이 감소하게 되고, 파라미터의 변동성보다 편향의 최소화를 목표로 하는 연구에 있어서는 SGWR 모형을 이용하는 편이 더 적절할 것이다.

[그림 13]와 [그림 14]에서는 지역별 가중회귀모형을 적합했을 때의  $R^2$  값이 나타나 있다. 국지  $R^2$ 는  $1 - \frac{RSS_i}{TSS}$ 로 계산되므로, 잔차의 크기와 무관하게 일부 지역에서는 국지  $R^2$ 이 낮게 나타날 수도 있음을 유의해야 한다. GWR과 SGWR 모형의 국지  $R^2$ 는 상당히 다른 공간적 패턴을 보인다. 이는 GWR 모형의 적합 과정을 생각해 보면 예측 가능한 결과이다. 앞서도 언급했듯이 GWR 모형의 경우 가중치가 물리적 거리에 기반해 계산된다. 따라서 인접한 지역은 실제로 회귀 모형에 사용되는 관측치의 조합이 거의 유사할 것이며 전체적으로 국지  $R^2$  패턴이 평활하게 나타난다. 특히 선택되는 관측치의 조합이 전체 자료를 잘 대표한다면  $RSS_i$  값이 낮아지고  $R^2$ 는 높아진다. 반면 SGWR 모형에서는 속성 유사성을 동시에 고려하므로  $R^2$ 의 공간적 이질성이 크고 파라미터의 변동성이 높아진 만큼 국지  $R^2$ 의 최솟값은 GWR 모형에서의 최솟값과 비교했을 때 더 낮은 것을 확인할 수 있었다.



[그림 15] GWR 모형으로 추정한 국지 회귀계수

마지막으로 [그림 7]과 [그림 15]를 통해 GWR 모형과 SGWR 모형의 회귀계수 추정치를 비교 및 해석해 보고자 한다. 먼저 인구 이동 유효도 지수와 평균 초혼 연령 변수의 경우 합계출산율에 대한 영향력의 방향이 역전되었다는 점이 특기할 만하다. GWR 모형에서 인구 이동 유효도 지수는 전체적으로 정(+)의 관계를, 평균 초혼 연령은 전체적으로 부(-)의 관계를 드러냈지만 SGWR 모형에서는 반대로 나타났다. 즉, SGWR 모형에 따르면 인구 이동 유효도 지수가 높아질수록 출산율이 감소하고, 평균 초혼 연령이 많아질수록 출산율이 증가하는 경향성을 보인다. 이는 심슨의 역설 상황과 같이 발생 가능한 경우이기는 하나, 방향성이 완전히 역전되었다는 점에서 추정 절차에 문제가 없었는지 살펴보거나 어떤

경우 이런 상황이 나타나는지 확인해 볼 필요가 있다. 둘째, GWR에서 유의하지 않은 변수가 SGWR에서 유의한 것으로 나타나는 경우, 혹은 그 반대 경우이다. GWR 모형에서는 전 지역에서 평균 종합 소득 신고액이 유의하지 않은 변수로, 여성 경제활동참가율이 정(+)의 관계를 갖는 유의한 변수로 나타났으나 SGWR 모형에서는 평균 종합 소득 신고액이 부(-)의 관계를 갖는 유의한 변수로, 여성 경제활동참가율은 유의하지 않은 변수로 나타났다. 셋째, GWR과 SGWR 모형 모두에서 유사한 영향력을 보이는 변수들로는 인구 밀도와 다문화 결혼 비율이 있었으며 유의한 회귀계수의 분포는 달라지기는 했지만, 전체적인 경향성은 부(-)의 관계로 유지되었다. 마지막으로 인구 1000명당 어린이집 수는 두 모형 모두에서 유의한 영향력을 드러내지 못했다.

## V. 결론

본 연구는 SGWR 모형을 이용하여 국내 출산력에 영향을 미치는 요인들의 공간적 이질성을 분석하고 기존 모형보다 적합도 및 정확성이 제고되었음을 확인하였다. 우선 모형 적합 절차로, 4종류의 서로 다른 탐색 방법을 적용하여 적합 결과를 비교하였다. 기존 문헌에서 보고된 것과 달리, 단일한 경사 하강 알고리즘을 이용하는 것보다 Differential Evolution 방법과 L-BFGS 등의 준-뉴턴 방법을 동시에 활용할 때 국소 최적에 빠지는 것을 방지하고 가장 높은 적합도를 담지하는 파라미터를 찾을 수 있었다. 적응형 가우시안 커널을 이용하는 경우, 최종적으로 선택된 모형의 최적 파라미터는  $b_G = 111.37$ ,  $b_S = 1.14$ ,  $\alpha = 0.50$ 였으며, 이를 이용하여 SGWR 모형을 적합한 결과 AICc, RSS, RMSE 등의 적합도 지표에서 OLS, GWR, MGWR 모형에 비해 우수한 성능을 보였다. 이는 Bias-Variance Trade-Off 관계에서 속성 유사성을 고려함으로써 편향을 낮아지고, 전체적으로 모형의 설명력을 높아지는 것으로 해석할 수 있다.

최종적으로 선택된 모형으로 출산력에 영향을 미치는 요소들의 공간적 이질성을 분석한 결과는 다음 네 가지로 요약할 수 있다. 첫째, 아동 1000명당 어린이집 수와 여성 경제활동참가율은 모든 지역에서 유의미한 영향력을 보이지 않았다. 둘째, 다문화 혼인 비율과 평균 소득 신고액은 모든 지역에서 유의미한 부(-)의 영향력을 보였으며, 인구 밀도와 인구 이동 유효도 지수는 절반에 가까운 지역에서 유의미한 부(-)의 영향력을 보였다. 셋째, 평균 초혼 연령은 모든 지역에서 유의미한 정(+)의 영향력을 보였으며, 특히 이는 GWR 모형의 분석 결과와 반대되는 방향성이라는 점에서 GWR 모형 적합 결과에 있어서 심슨의 역설과 같은 현상이 나타날 수 있음을 의미하였다. 마지막으로, 잔차 분석 결과 SGWR 모형을 적합했을 때도 잔차의 공간적 자기상관은 존재하지 않았다.

본 연구의 한계점으로는 다음의 세 가지 측면을 제시하고자 한다. 첫째, 종속변수인 합계출산율에 영향을 미치는 시차를 단순 1년 전으로 가정하였다는 점이다. 실제로 출산력에 대한 영향력은 시계열적으로 더 길거나 짧을 수 있으며, 변수별로 작동 스케일 역시 상이할 것으로 예상된다. 따라서 이를 더욱 정밀하게 고려하기 위해서는 시공간 데이터를 다루기 위한 GTWR 등의 방법론을 이용하고, 그 역시 속성 유사성 가중치를 반영하는 방법을 개발할 필요가 있다. 둘째, 출산력에 영향을 미치는 설명변수가 누락되었을 수 있다는 점이다. 특히 정책적 변수의 경우 일관된 정량화에 어려움이 있어 출산장려금 등의 효과를

고려하지 못했는데, 정책적 시사점을 더욱 강화하기 위해서는 이러한 변수들을 추가로 고려할 필요가 있다. 마지막으로 SGWR 모형의 한계점으로, 아직 일반적인 상황에서 어떤 이유로 설명력 및 정확성이 제고되는지에 대한 정당화가 부족하다는 점과, 복잡한 파라미터 공간에서 국소 최적에 갇히지 않고 전역적 최적해를 찾을 수 있음과 동시에 연산 효율성이 높은 방법이 아직 특정되지 않았다는 점이다. 개인적으로는 AICc가 과연 SGWR에 적절한 적합도 평가 지표인지 의문이 남았으며, 특히 오버피팅을 방지하기 위해서는 대역폭의 탐색 반경에 대한 적절한 제약을 부여하거나 ENP에 대한 페널티를 더 크게 부여해야 할 것 같다는 생각이 들었다. 특히 마지막 측면은 SGWR 방법론의 확장성 및 신뢰성을 높이기 위하여 관련된 후속 연구가 조속히 진행되어야 할 것이며, 이를 통해 국내 출산력에 영향을 미치는 요인들의 공간적 이질성을 새로운 시각에서 조망할 수 있게 되기를 기대한다.

## VI. 참고 문헌

### 1. 국내 문헌

- 고원식. (2019). 출신국가별 합계출산율이 결혼이민여성의 출산력에 미치는 영향. *응용경제*, 21(3), 41-68.
- 이상일, & 이소영. (2023). 인구이동이 인구재분포에 미치는 영향력의 시공간적 역동성 탐색: 우리나라 국내 인구이동에의 적용. *한국지도학회지*, 23(1), 1-19. <https://doi.org/10.16879/jkca.2023.23.1.001>
- 장인수, & 정찬우. (2022). 출산지원금이 지역 출산력에 미치는 영향에 대한 공간적 변이 탐색: 다중범위지리가중회귀모형의 적용과 함의. *보건사회연구*, 42(4), 305-325. <https://doi.org/10.15709/hswr.2022.42.4.305>
- 정명구. (2017). 『지리가중회귀모형을 이용한 지역별 차별 출산력 분석』 (석사학위논문). 서울대학교 보건대학원.
- 이영서. (2024). 『지리가중 회귀분석을 활용한 출산장려금과 기초자치단체 출산율의 관계 탐색 연구』 (석사학위논문). 서울대학교 간호학과.

### 2. 외국 문헌

- Anselin, L., & Li, X. (2020). Tobler's law in a multivariate world. *Geographical Analysis*, 52(4), 494-510. <https://doi.org/10.1111/gean.12237>
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Longman Scientific & Technical.
- Brazil, N. (2022). Environmental inequality in the neighborhood networks of urban mobility in U.S. cities. *Proceedings of the National Academy of Sciences*, 119(17), Article e2117776119. <https://doi.org/10.1073/pnas.2117776119>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281-298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Cho, S.-Y. (2018). A bride deficit and marriage migration in South Korea. *International Migration*, 56(6), 100-119. <https://doi.org/10.1111/imig.12432>
- da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48(3), 233-247. <https://doi.org/10.1111/gean.12084>
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247-1265. <https://doi.org/10.1080/24694452.2017.1352480>
- Fotheringham, A. S., Oshan, T. M., & Li, Z. (2023). *Multiscale geographically weighted regression: Theory and practice*. CRC Press. <https://doi.org/10.1201/9781003435464>
- Fukuda, S. (2020). Marriage will (continue to) be the key to the future of fertility in

- Japan and East Asia. *Vienna Yearbook of Population Research*, 18, 71–80. <https://doi.org/10.1553/populationyearbook2020.deb07>
- Griffith, D. A. (2019). Negative spatial autocorrelation: One of the most neglected concepts in spatial statistics. *Stats*, 2(3), 388–415. <https://doi.org/10.3390/stats2030027>
- Jing, F., Li, Z., Qiao, S., Zhang, J., Olatosi, B., & Li, X. (2022). Investigating the relationships between concentrated disadvantage, place connectivity, and COVID–19 fatality in the United States over time. *BMC Public Health*, 22, 2346. <https://doi.org/10.1186/s12889-022-14779-1>
- Jung, M., Ko, W., Choi, Y., & Cho, Y. (2019). Spatial variations in fertility of South Korea: A geographically weighted regression approach. *ISPRS International Journal of Geo–Information*, 8(6), Article 262. <https://doi.org/10.3390/ijgi8060262>
- Kiefer, J. (1953). Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3), 502–506. <https://doi.org/10.2307/2032161>
- Lessani, M. N., & Li, Z. (2024). SGWR: Similarity and geographically weighted regression. *International Journal of Geographical Information Science*, 38(7), 1232–1255. <https://doi.org/10.1080/13658816.2024.2342319>
- Myrskylä, M., Kohler, H.–P., & Billari, F. C. (2009). Advances in development reverse fertility declines. *Nature*, 460(7256), 741–743. <https://doi.org/10.1038/nature08230>
- Ni, S., Wang, Z., Wang, Y., Wang, M., Li, S., & Wang, N. (2022). Spatial and attribute neural network weighted regression for the accurate estimation of spatial non–stationarity. *ISPRS International Journal of Geo–Information*, 11(12), 620. <https://doi.org/10.3390/ijgi11120620>
- OECD. (2024). *Society at a glance 2024: OECD social indicators*. OECD Publishing. <https://doi.org/10.1787/918d8db3-en>
- Sachdeva, M., & Fotheringham, A. (2023). A Geographical Perspective on Simpson's Paradox. 1–25. 10.5311/JOSIS.2023.26.212.
- Storn, R., & Price, K. (1997). Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. <https://doi.org/10.1023/A:1008202821328>
- Sui, D. Z. (2004). Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, 94(2), 269–277. <https://doi.org/10.1111/j.1467-8306.2004.09402003.x>
- Yu, S., Hu, X., Sheng, Y., & Zhao, C. (2025). Similarity and geographically weighted regression considering spatial scales of features space. *Spatial Statistics*, 67, Article 100897. <https://doi.org/10.1016/j.spasta.2025.100897>
- Zhou, X., Assunção, R., Shao, H., Huang, C., Janikas, M., & Asefaw, H. (2023). Gradient–based optimization for multi–scale geographically weighted regression. *International Journal of Geographical Information Science*, 37(10), 2101–2128. <https://doi.org/10.1080/13658816.2023.2246154>

Zhu, A.-X., & Turner, M. S. (2022). How is the third law of geography different? *Annals of GIS*, 28(1), 57-67. <https://doi.org/10.1080/19475683.2022.2026467>