

공간회귀분석을 활용한 통행발생모형 추정 연구 : 서울특별시를 중심으로

Developing Trip Generation Models Using Spatial Regression Analysis
: A Case for Seoul, Korea

진창종
Jin Changjong
홍익대학교 대학원 도시계획과 석사과정(제1저자)
M.S. Candidate, Dept. of Urban Planning,
Graduate School, Hongik Univ.(Primary Author)
(kor21cm@hanmail.net)

이향숙
Lee Hyangsook
홍익대학교 부설 환경개발연구원 연구위원
Research Fellow, Research Institute for Environmental
Development, Hongik Univ.
(hslee3060@gmail.com)

추상호
Choo Sangho
홍익대학교 도시공학과 교수(교신저자)
Prof., Dept. of Urban Design and Planning,
Hongik Univ.(Corresponding Author)
(shchoo@hongik.ac.kr)

목 차

- I. 서론
- II. 선행연구 검토
- III. 연구의 범위 및 방법
- IV. 통행발생모형 정립
 - 1. 자료 수집
 - 2. 공간자기상관의 탐색
 - 3. 모형의 정립
- V. 분석 결과
- VI. 결론

I. 서론

합리적인 교통계획을 위해서는 일련의 과정들을 거치게 된다. 문제 진단과 목표 설정, 자료의 수집 및 분석을 기초로 한 통행수요모형 작성과 대안 작성, 대안별 통행수요예측 및 비용·편익 산출, 대안평가와 최종 실행계획 및 집행이 그것이다. 각각의 단계들이 모두 중요한 특성을 지니지만 그중에서도 통행수요의 예측 단계는 시설의 규모를 정하는 데 있어 필수적이며, 잘못된 수요 예측으로 인해 발생하는 비용 산정의 문제 등과 직결되는 중요한 단계라 할 수 있다.

통행발생은 일반적으로 사용되는 4단계 교통수요 추정법(통행발생, 통행배분, 수단선택, 노선배정) 중 첫 번째 과정이다. 통행발생에 대한 예측이 잘 이루어질수록 통행수요의 예측은 더 정확해진다. 통행발생을 추정하기 위해서 우리나라에서는 자료의 수집이 용이한 행정구역 단위의 교통존을 기반으로 선형회귀식에 의한 통행발생모형 구축에 중점을 둔 반면, 외국에서 주로 사용되는 가구단위의 모형 적용은 자료수집 및 가구단위 변수의 예측에 한계가 있어 이를 정립하는 데는 어려움이 있었다.

한편, 교통존과 같이 공간을 준거로 하는 데이터들은 인접한 지역에서 서로의 값들에 대해 영향을 주고받는 상호의존적인 특성을 가지는데(김광구 2003), 이러한 특성을 공간자기상관이라 한다. 공간자기상관은 선형회귀모형의 기본전제인 무작위적 분포와 오차항의 독립성을 위배시켜 올바른 추정값을 얻지 못하게 한다. 즉, 통행발생 모형의 올바른 추정값을 얻기 위해서는 선형회귀모형에 대한 공간자기상관을 통제할 필요성이 있다.

본 연구에서는 선형회귀모형의 공간자기상관을 통제하기 위한 방법인 공간회귀분석을 사용하여 통행발생 유출·유입모형을 새롭게 정립하고자 한다. 또한 기존의 통행발생모형에 적용되었던 사회경제

지표와 더불어 지역적 특성을 나타내는 토지이용지표, 물리적 지표 등을 모형의 독립변수로 사용하여 분석의 정확도를 높이고자 한다.

II. 선행연구 검토

통행발생량을 추정하기 위한 방법으로 성장률법, 원단위법, 카테고리모형, 회귀모형 등이 있다. 국내의 경우 카테고리 분석법을 일부 이용하고 있으나, 자료구축의 한계로 인하여 자료수집이 비교적 용이한 회귀분석을 주로 이용한다. 유완 외(1999)는 1997년 서울시의 교통 센서스 구축자료를 이용하여 가구의 통행행태를 수요집단, 주중집단, 주말집단으로 분류하고, 선형회귀분석으로 각각의 집단에 대해 총통행수 및 통행/인에 대한 통행발생 예측모형을 정립하였다. 정지교(2001)는 동일 자료를 기반으로 서울시 동별 사회·경제적 특성을 고려하여 세 개의 지역(주거중심지역, 상업중심지역, 상업·업무중심지역)으로 구분하고 선형회귀분석을 통해 지역별, 통행목적별 통행발생예측모형을 도출하였다. 최근에 송재인 외(2011)는 서울시 5개 구와 경기도 6개 시·군에 대해 지역별 토지이용 특성을 변수로 활용하여 통행발생모형을 구축하였다.

Washington S. 외(1997)에서는 기존의 선형회귀분석을 이용한 통행발생모형의 한계점을 보완하기 위한 방법으로 계층적 구조 기반의 회귀분석(Hierarchical tree-based regression)을 사용하였고, 장수는 외(2000)는 '제1차 전국 물류현황 조사-지역간 화물 물동량 조사' 자료를 바탕으로 신경망 이론을 통행발생모형에 적용하였으며, 선형/비선형회귀모형과의 비교를 통해 역전파 신경망 모형의 적용 가능성을 입증하였다. 김현기 외(2004)는 베이지안 통계이론을 적용하여 변수 간 인과관계를 파악하고 종속적 카테고리 모형을 통해 통행발생량을 예측하였

표 1 _ 통행발생모형 관련 선행연구 동향

연구자	주요내용	분석모형
Washington S. 외 (1997)	계층적 구조 기반의 회귀분석(HTBR)을 이용하여 기존의 선형회귀를 통한 통행발생모형을 보완	계층적회귀분석모형
김순관 (1998)	수도권을 대상으로 사회경제지표를 보완하여 통행발생예측모형 정립	선형회귀모형
유완 외 (1999)	가구특성 및 요일별 통행행태에 대한 군집분석을 통해 수요집단, 주중집단, 주말집단으로 나누고 각각에 대한 통행발생모형 정립	선형회귀모형
장수은 외 (2000)	신경망 이론을 통행발생모형에 적용하여 선형 · 비선형 회귀모형과 비교	신경망 모형
정지교 (2001)	사회경제적 특성을 고려하여 서울시 행정동을 주거중심지역, 상업중심지역, 상업 · 업무 중심지역으로 구분 후 목적별 통행발생예측모형 정립	선형회귀분석
김현기 외 (2004)	베이지안 통계 이론을 적용하여 변수 간 인과관계를 파악하여 통행발생 모형에 반영함	카테고리모형
김진자 외 (2004)	수도권의 통근·통학통행 발생률을 카테고리모형과 선형회귀모형을 이용, 예측하고 관측 값과 비교하여 카테고리 모형의 예측력이 우수한 결과를 도출함	카테고리모형 선형회귀모형
Yao L. 외 (2008)	통행자의 행동에 영향을 미치는 개인특성, 통행특성 및 통행서비스수준을 반영한 통행발생의 개별행태모형 도출	카테고리모형
Kim J. 외 (2009)	대구지역을 대상으로 인구, 고용자 수, 학생 수를 이용하고 공간회귀분석을 통해 통행발생모형 정립	공간회귀모형
송재인 외 (2011)	토지이용 특성을 반영하여 수도권 OD 및 PA기반 목적별 통행발생모형 구축	선형회귀모형

다. 김진자 외(2004)에서는 수도권의 통근·통학통행 발생률을 카테고리모형과 선형회귀모형을 이용하여 예측하고 이를 관측 값과 비교하여 카테고리 모형의 우수성을 검증하였다. 또한 Yao L. 외(2008)는 통행자의 개인특성, 통행특성 및 통행서비스수준을 반영, 카테고리모형을 도출하여 회귀모형보다 뛰어난 결과를 증명하였다.

한편, Kim J. 외(2009)에서는 공간종속성을 반영한 통행발생모형을 구축하였으나, 대구지역을 대상으로 하였고, 사용된 독립변수가 인구, 학생 수, 고용자 수 등으로 한정되어 있어 지역의 물리적 특성, 토지이용 지표 특성 반영에 대한 아쉬움이 있었다.

선행 연구의 검토 결과 회귀분석을 이용한 통행발생모형의 개선방안으로 새로운 변수를 적용하거나, 케이스의 집단 분류를 통해 각각의 통행발생모형을 정립하는 방법이 주를 이루었으나 인접지역의 특성을 반영하는 모형 연구는 미미한 것으로 나타났다.

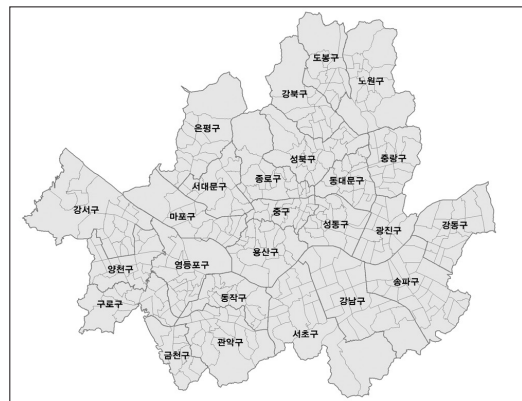
따라서 본 연구에서는 통행이 발생하는 지역들 간의 연관성을 고려하고, 다양한 지역특성 데이터를 활용하여 공간회귀분석을 수행하고자 한다. 또한, 그

결과를 보편적으로 이용되는 선형회귀모형과 비교 분석하였다.

III. 연구의 범위 및 방법

본 연구는 2006년 가구통행실태조사를 바탕으로 전수화된 수도권 목적별 통행량(OD) 중 서울지역의 통행 자료를 이용하여 각 목적별 통행발생모형을 개발하였다.

그림 1 _ 연구의 범위(서울특별시 행정동 전역)



분석은 다음의 네 가지의 단계로 진행되었다. 첫째, 통행 유출량 및 유입량과 사회경제지표, 토지이용지표, 물리적지표 간 상관분석을 통해 통행과 연관성이 있는 변수를 일차적으로 선별하였다. 둘째, 다중선형회귀분석의 단계선택방법(stepwise)을 통하여 변수를 재선별하고, 변수 간 다중공선성이 있는 경우(VIF검증 값이 10 이상)를 제외하여 최종변수를 추출하였다. 셋째, Moran's I 분석으로 종속변수의 공간자기상관을 탐색함으로써 공간회귀분석의 적용여부를 판별하였다. 넷째, 공간회귀분석을 통해 최종 모형을 도출하였다.

본 연구에서 이용된 공간데이터는 ArcGIS 9.0으로 구축하였다. 상관관계 분석과 다중선형회귀분석은 SPSS 18.0을 이용하였고, 공간자기상관의 탐색과 공간회귀분석은 OpenGeoDa1.0.1을 활용하였다.

IV. 통행발생모형 정립

1. 자료 수집

2006년 목적별 OD자료를 이용하여 서울지역의 동별 유출량 및 유입량을 추출하였다. 사회경제지표는 2006년 기준으로 수집되었으며, 토지이용지표와 물리적 지표는 구축이 가능한 2008년 자료를 이용하였다. 토지이용지표 중 건축물 용도별 연면적 데이터는 도로명주소 관리시스템(KLIS-rn)으로부터 수집된 건축물 데이터로부터 산출되었다.¹⁾ 기존의 연구에서 이용한 건축물 GIS데이터는 용도분류와 연면적의 결측치가 많아 분석에 한계점이 있었던 반면, KLIS-rn데이터는 건축물 주용도, 지상층수, 지하층수 등이 99% 이상 명시되어 있어 분석의 정확성을

높이는 데 유용하다. 수집된 자료의 세부정보는 <표 2>과 같다.

OD자료와 사회경제지표는 2006년 기준 소존단위인 522개의 행정동으로 구축되었으나, 토지이용지표와 물리적 지표는 2008년 기준인 424개의 행정동으로 구축되어 자료 매칭 시 문제점이 발생하였다. 이를 해결하기 위해 각 구청 및 주민센터의 자료를 기반으로 2006년 소존을 2008년 기준으로 변경하였다.

표 2_수집자료

구분		조사항목
종속 변수	OD 기반 통행량 자료	총 통행량, 목적별 통행량 (통근, 통학, 업무, 쇼핑, 기타)
	사회 경제 지표	인구 수(총, 연령별), 가구 수, 수용 학생 수 (초, 중, 고, 대학생 수), 승용차 수, 고용자 수, 종사자 수(1, 2, 3차 종사자 수), 학교 수(초, 중, 고, 대학, 전문대학 수)
독립 변수	토지 이용 지표	건축물 용도별 ¹⁾ 총연면적 자료, 용도지역 면적, 용도지역 혼합도 ²⁾
	물리적 지표	버스정류장 수, 지하철 역세권(반경 500m), 도시철도 연장, 실폭도로 면적, 공영주차장 면적

주: 1) 건축물 시행령 <별표 1>에 따른 용도 중간항목 분류는 다음과 같다.

1. 단독주택, 2. 공동주택, 3. 제1종근린생활시설, 4. 제2종근린생활시설, 5. 문화 및 집회시설, 6. 종교시설, 7. 판매시설, 8. 운수시설, 9. 의료시설, 10. 교육연구시설, 11. 노유자시설, 12. 수련시설, 13. 운동시설, 14. 업무시설, 15. 숙박시설, 16. 위탁시설, 17. 공장, 18. 창고시설, 19. 위험물 저장 및 처리시설, 20. 자동차 관련시설, 21. 동물 및 식물 관련시설, 22. 분뇨 및 쓰레기 처리시설, 23. 교정 및 군사 시설, 24. 방송통신시설, 25. 발전시설, 26. 묘지관련시설, 27. 관광휴게시설

$$2) Entropy Index = -\sum_{i=1}^n (P_i) \ln(P_i) / \ln(s), P_i = \frac{A_i}{A_1 + A_2 + \dots + A_n}, s: \text{가짓수}$$

(국토연구원, 2004 참조).

엔트로피지수(Entropy Index)는 혼합도 및 다양성을 판단할 수 있는 대표적인 기법이며 그 값이 1에 가까울수록 높은 혼합도를 나타냄. 본 연구에서는 주거지역, 상업지역, 공업지역, 녹지지역 총 4개의 용도지역에 대한 엔트로피지수를 구하였음.

1) 건축물 연면적 산정기준: 연면적 = 건폐면적 × (지상층수+지하층수)

도로명주소 관리시스템(KLIS-rn)의 건축물 데이터는 지상층수, 지하층수의 구축률이 99%로 매우 높으나 주요 항목에 건축물 연면적은 제외되어 있음. 따라서 연면적 산정을 위해 위의 식을 도입하여 계산하였음.

2. 공간자기상관의 탐색

공간상의 실체들은 공간에 무작위적으로 분포하지 않고 서로 영향을 주고받으며, 그 영향은 공간실체 간의 거리나 인접성이 높을수록 더 커진다고 할 수 있다(김광구, 2003). 이러한 특성을 공간자기상관이라 정의한다. 본 연구는 행정동을 분석단위로 하는 면(polygon)단위 자료를 사용하게 되는데, 이러한 데이터들은 공간단위와 공간현상의 불일치가 발생할 가능성이 있어 올바른 공간정보를 반영하지 못하기 때문에 공간자기상관을 일으키는 주요 요인으로 인식되고 있다(Anselin and Bera, 1998; 변필성, 2009). 기존의 통행발생모형에 일반적으로 이용되는 선형회귀분석은 오차항의 독립성을 가정하고 있으나,²⁾ 공간데이터의 공간자기상관으로 인해 이러한 가정이 위배될 가능성이 있어 이를 통제하기 위해 공간회귀분석을 사용한다.

공간자기상관의 탐색은 주로 Moran's I 공간자기상관계수를 통해 이루어진다. Moran's I 계수의 산출식과 공간자기상관의 판단 기준은 <표 3>에 나타나 있다.

여기서 w 는 공간가중치 행렬로 공간적 인접성(spatial contiguity)을 정의한다. 이는 서로 인접한 경우에 대해 가중치를 부여하는 것이다. 가중치는 인접기반(Contiguity-based)방식 또는 거리기반

(Distance-based)방식으로 부여될 수 있다(Anselin, 2005). 본 연구에서는 인접기반방식 중 하나인 Rock 방식(Rock-Based Contiguity)을 선택하였다. Rock 방식은 선을 공유하는 면에 대해 가중치 부여하는 방법인데, 본 연구의 공간단위인 행정동은 다양한 크기와 형태를 가지고 있기 때문에, 거리기반(예: Threshold distance 또는 K-nearest)방식의 가중치 설정보다 인접기반방식의 가중치 설정이 타당하다고 보아 Rock 방식을 선택하였다. 이러한 가중치는 Moran's I 분석뿐만 아니라 LISA(Local Indicator of Spatial Association)분석과 공간회귀분석에도 동일하게 사용되었다.

목적별(통근, 통학, 업무, 쇼핑, 기타) 통행 유출·유입량에 대한 Moran's I 값은 <표 4>와 같다. 모든 변수가 정적 공간자기상관을 보이는 것으로 나타났으며, 특히 통근통행 유입량, 업무통행 유출·유입량, 쇼핑통행 유출·유입량, 기타통행 유입량의 경우 Moran's I 계수 0.2 이상으로 공간자기상관이 비교적 높게 나타났다.

LISA 분석은 지역적으로 어떠한 공간적 군집 패턴이 발생하는지를 파악하기 위해 실시하는 분석이다(정경석 외, 2009). 본 연구에서는 Moran's I 계수가 상대적으로 높게 나온 목적통행에 대하여 추가적으로 LISA 분석을 실시하였다. 이를 통해 해당 지역

표 3_Moran's I 공간자기상관 계수

$$Moran\ I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2} \quad \begin{array}{l} x_i : \text{공간 } i \text{의 값} \\ \bar{x} : x \text{의 평균값} \\ n : \text{개체수} \end{array}$$

부적(negative) 공간자기상관	공간자기상관 없음	정적(positive) 공간자기상관
-1 <<<<<<<	0	>>>>>>> 1

표 4_Moran's I 공간자기상관 계수

변수(유출량)	Moran I	변수(유입량)	Moran I
총통행	0.1762	총통행	0.1776
통근통행	0.1668	통근통행	0.2738
통학통행	0.2131	통학통행	0.2489
업무통행	0.2616	업무통행	0.2547
쇼핑통행	0.2159	쇼핑통행	0.2489
기타통행	0.1570	기타통행	0.2248

2) 오차항(error term)은 종속변수의 관측치와 예측치 간의 차이를 말하며, 예측된 오차값들은 서로 독립적이라는 가정이 필요함. 즉 종속변수의 변화에 따라 오차항이 어떤 패턴을 가져서는 안 됨(이학식 외, 2011).

과 주변지역의 공간자기상관이 모두 높은 값을 갖는 핫스팟(High-High)과 모두 낮은 값을 갖는 콜드스팟(Low-Low)지역을 표현해주는 클러스터맵(Cluster Map)을 구현하였다.

〈그림 2〉에 의하면 종로구와 을지로 등 서울 도심 지역과 강남구 일대가 핫스팟 지역으로 나타났다. 이는 높은 통행유입량을 보이는 군집을 뜻하는데, 실제로 위의 지역들은 출근시간 많은 인구가 한꺼번에 유입되는 서울의 대표적인 중심업무지역이다.

〈그림 3〉은 LISA 분석을 통해 도출된 LISA I 계수 값을 도면화한 것으로 지역적으로 공간자기상관이 높은 곳을 한눈에 알 수 있다.

그림 2_ 통근통행 유입량 LISA Cluster Map

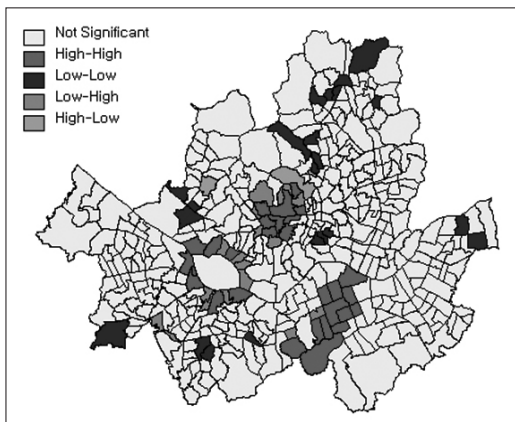
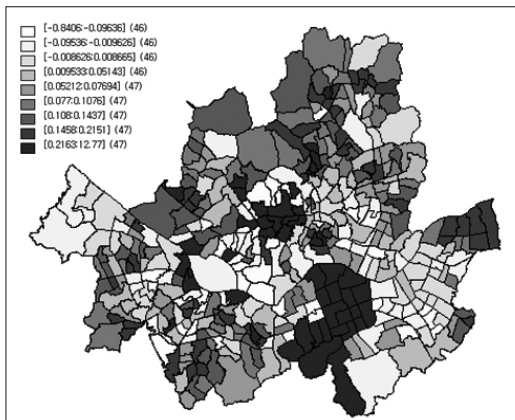


그림 3_ 통근통행 유입량 LISA I (Local Moran's I) Map



3. 모형의 정립

Moran's I 분석을 통해 각 목적별 통행 유출량 및 유입량에 공간자기상관이 탐색되었다면 이를 통제하기 위해 공간회귀분석을 수행한다. 공간회귀분석은 일반적으로 SEM모형(Spatial Error Model)과 SLM모형(Spatial Lagged Model)을 이용한다. 본 연구에서는 공간자기상관이 없는 경우 OLS모형(Ordinary Least Square Model, 선형회귀모형)을 구축하고, 공간자기상관이 있는 경우 SLM모형과 SEM모형을 구축하여 비교분석하였다.

OLS모형의 식은 다음과 같다.

$$y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \dots + \epsilon$$

y 는 종속변수, X 는 독립변수이고 δ 는 추정계수, ϵ 는 오차항 <식 1>

SEM모형은 공간오차모형이라고 하며, 오차항의 자기상관을 통제하기 위해 사용한다. SEM모형식은 다음과 같다.

$$y = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \dots + u$$

$$u = \lambda W u + e$$

y 는 종속변수, X 는 독립변수, u 는 공간자기상관을 갖는 오차항, W 는 공간가중치 행렬, δ, λ 는 추정계수, e 는 공간자기상관이 없는 오차항 <식 2>

SLM모형은 공간시차모형 혹은 공간지체모형이라고 하며, 종속변수와 오차항의 자기상관을 통제하기 위해 사용한다. SLM모형식은 다음과 같다.

$$y = \rho W y + \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \dots + \epsilon$$

y 는 종속변수, X 는 독립변수, W 는 공간가중치, δ, ρ 는 추정계수, ϵ 는 오차항 <식 3>

GeoDa 프로그램을 통한 공간회귀분석 모형의 선정은 다음과 같은 과정을 거친다. 먼저 변수의 다중공선성 및 이분산성을 진단하고, Moran's I, LM (Lagrange Multiplier) lag, LM error, Robust LM lag, Robust LM error 검정을 통해 적합한 모형을 선택한다. <그림 4>의 프로세스에 따라 공간자기상관 검정이 유의하지 않을 경우 OLS (Ordinary Least Square) 모형을 선택하고, 공간자기상관이 있는 경우 SLM 모형과 SEM 모형 중 더 유의한 모형을 선택한다.

V. 분석 결과

위의 과정을 통하여 총통행 및 5가지 목적통행(통근, 통학, 업무, 쇼핑, 기타)에 대한 유출·유입 모형이 각각 도출되었다. 동일 목적의 유입·유출 모형의 경우 대체로 같은 회귀분석 모형이 선택되었으며, 공간회귀분석은 SEM 모형이 주로 선택되었다. 변수 중 가구수는 모든 유출모형에 독립변수로서 선택되었으며,

통학목적은 제외한 대부분의 유입모형은 종사자 수 관련 지표와 제2종근린생활시설을 독립변수로 구성하고 있었다. 그 밖에 건축물 용도별 연면적, 도시철도연장이나 버스정류장 수, 지하철 역세권 등의 기존의 모형들에서 잘 다루지 않았던 지역의 특성을 나타내는 설명변수들이 통행유입·유출모형에 다수 포함되어 있었다.

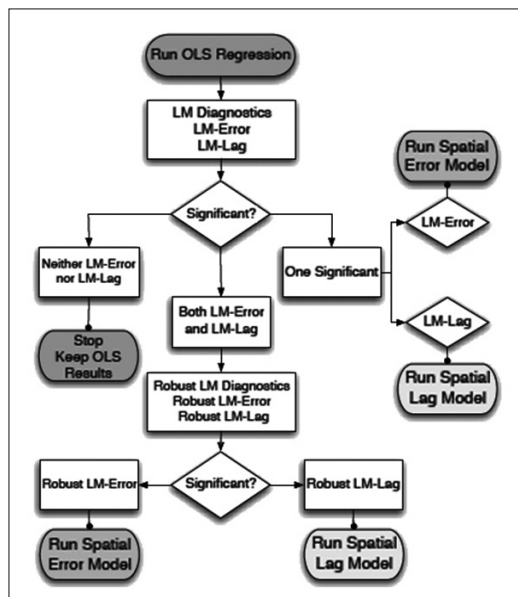
총통행 유출모형<표 5>의 경우 공간종속성 검정

표 5_ 공간회귀분석 결과(총통행 유출량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	-1.24E+03	0.58	-1.17E+03	0.57
3차 종사자 수	2.49E+00	0.00*	2.55E+00	0.00*
학생 수	9.40E-01	0.00*	9.92E-01	0.00*
가구 수	2.76E+00	0.00*	2.77E+00	0.00*
차량 수 /1,000명	6.04E+00	0.01*	6.22E+00	0.00*
공동주택	7.29E-03	0.00*	6.22E-03	0.00*
위락시설	1.61E+00	0.00*	1.63E+00	0.00*
중심상업지역	9.85E-02	0.05*	1.13E-01	0.02*
도시철도연장	1.69E+00	0.03*	1.14E+00	0.10**
R-squared	0.852399		0.858901	
Lagcoeff.			(Lambda): -0.31, p=0.00*	

주: *p < 0.05, ** p < 0.1.

그림 4_ 공간회귀모형 선택과정



자료 : Anselin, 2005.

표 6_ 공간회귀분석 결과(총통행 유출량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	1.22E+03	0.54	1.55E+03	0.38
종사자 수	2.08E+00	0.00*	2.20E+00	0.00*
학생 수	1.03E+00	0.00*	1.14E+00	0.00*
가구 수	2.76E+00	0.00*	2.72E+00	0.00*
제2종 근생시설	2.12E-02	0.00*	1.83E-02	0.00*
1종일반지역	6.07E-03	0.08**	4.54E-03	0.13
중심상업지역	1.14E-01	0.02*	1.26E-01	0.01*
지하철역세권	5.63E-03	0.00*	4.19E-03	0.00*
R-squared	0.869007		0.878112	
Lagcoeff.			(Lambda): -0.39, p=0.00*	

주: *p < 0.05, ** p < 0.1.

결과 SEM모형이 적합한 것으로 나타났다. 사회·경제적 지표로 3차산업 종사자 수와 총 학생 수, 가구 수, 차량대수/1,000명 등이 독립변수로 포함되었으며, 공동주택, 위락시설 총 연면적 등의 건축물 용도 관련 지표와, 중심상업지역, 도시철도연장 지표 등 지역의 물리적 특성을 지닌 지표가 다수 포함되었다.

총통행 유입모형(표 6) 역시 검정결과 SEM모형이 최종 선택되었으며, 종사자수와 학생 수, 가구 수, 제2종근린생활시설, 중심상업지역(용도지역), 지하철역세권 등의 지표가 총통행 유입량에 영향을 주는 독립변수로 나타났다.

통근통행 유출 및 유입 모형(표 7, 8)은 SEM모형이 선택되었다. 한편 20~64세 인구가 통근통행 유출 모형의 변수로 선택되었는데, 이는 잠재적 취업가능 인구의 연령대를 구성한 지표로, 통근유출량에 영향을 미치는 것으로 사료된다. 또한 2·3차 종사자 수와 제2종근린생활시설, 도시철도연장 등의 지표가 통근통행 유입량에 영향을 미치는 독립변수로 선택되었다.

통학목적의 통행 유출모형(표 9)은 공간중속성 검정 결과 OLS모형이 최종 선택되었다. 가구 수와 단독주택 등 주거관련 지표가 독립변수로 선정되었다. 통학통행 유입모형(표 10)의 경우 등록 학생 수, 대학·전문대학 수 버스정류장 수 등 통학통행을 유발하는 변수가 지표로 설정되었다. 한편 5~29세 인구는 유치원부터 대학·대학원까지의 잠재적 통학가능인구를 구성한 지표인데, 통학통행 유출량과 유입량 둘 다 영향을 미치는 것으로 나타났다.

쇼핑목적의 통행 유출·유입모형(표 11, 12)는 모두 SLM모형이 선택되었다. 3차산업 종사자수가 두 모형의 공통 독립변수로 선정되었으며, 쇼핑목적의 통행을 유발하는 판매시설지표와 상업시설이 집중해 있는 지하철역세권이 통행유입모형의 지표로 선정되었다. 쇼핑통행유출모형의 경우 단독주택, 가구 수, 제3종일반주거지역이 독립변수로 포함되어

표 7_공간회귀분석 결과(통근통행 유출량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	-6.65E+01	0.73	-1.11E+02	0.59
가구 수	6.39E-01	0.00*	6.61E-01	0.00*
20~64세 인구	2.27E-01	0.00*	2.18E-01	0.00*
R-squared	0.871556		0.876474	
Lagcoeff.			(Lambda): 0.25, p=0.00*	

주: *p<0.05.

표 8_공간회귀분석 결과(통근통행 유입량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	-1.95E+01	0.96	-5.95E+01	0.82
2,3차 종사자 수	9.90E-01	0.00*	1.05E+00	0.00*
제2종 근린생활시설	7.84E-03	0.00*	6.18E-03	0.00*
도시철도 연장	6.92E-01	0.01*	4.02E-01	0.07**
R-squared	0.893362		0.908385	
Lagcoeff.			(Lambda): -0.58, p=0.00*	

주: *p<0.05, ** p<0.1.

표 9_공간회귀분석 결과(통학통행 유출량)

구분 변수	OLS	
	계수	p
상수	-2.50E+02	0.07**
5~29세 인구	3.88E-01	0.00*
초중고 학생 수	5.53E-02	0.03*
가구 수	1.58E-01	0.00*
단독주택	-1.81E-03	0.00*
R-squared	0.771968	

* p< 0.05, ** p<0.1

있는데 모두 거주인구와 관련된 지표이며, 쇼핑을 목적으로 한 통행이 보통 가정에서부터 이루어진다는 것을 고려하면 합당한 지표로 사료된다. 또한 가구 수의 경우 공간회귀분석을 통해 모형의 설명력을 향상시켰음을 알 수 있다.

업무목적의 통행 유출·유입모형(표 13, 14)는 검정 결과 SEM모형이 선택되었다. 업무통행 유출모

표 10_공간회귀분석 결과(통학통행 유입량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	-4.92E+02	0.03*	-4.66E+02	0.01*
5~29세 인구	1.15E-01	0.00*	9.68E-02	0.00*
학생 수	7.32E-01	0.00*	7.52E-01	0.00*
버스정류장 수	6.48E+00	0.00*	7.16E+00	0.00*
대학· 전문대학 수	1.80E+03	0.00*	1.81E+03	0.00*
R-squared	0.843723		0.859506	
Lagcoeff.			(Lambda): -0.47, p=0.00*	

주: *p < 0.05.

표 11_공간회귀분석 결과(쇼핑통행 유출량)

구분 변수	OLS		SLM	
	계수	p	계수	p
상수	1.52E+02	0.32	-2.93E+02	0.10**
자동차 수	1.06E-01	0.00*	8.46E-02	0.00*
3차종사자 수	6.92E-02	0.00*	6.37E-02	0.00*
가구 수	4.75E-02	0.05*	7.24E-02	0.00*
단독주택	1.40E-03	0.00*	1.21E-03	0.00*
위락시설	6.47E-02	0.00*	6.69E-02	0.00*
3종일반주거 지역	1.04E-03	0.01*	9.01E-04	0.01*
준주거지역	-3.12E-03	0.01*	-2.92E-03	0.02*
R-squared	0.528969		0.547303	
Lagcoeff.			(Rho): 0.20, p=0.00*	

주: *p < 0.05, ** p < 0.1.

표 12_공간회귀분석 결과(쇼핑통행 유입량)

구분 변수	OLS		SLM	
	계수	p	계수	p
상수	6.75E+02	0.00*	3.74E+02	0.05*
2차종사자 수	5.28E-01	0.00*	5.00E-01	0.00*
3차종사자 수	4.81E-02	0.06**	4.37E-02	0.08**
판매시설	5.50E-03	0.00*	5.30E-03	0.00*
공장	-4.27E-03	0.00*	-3.95E-03	0.00*
운수시설	1.13E-02	0.01*	1.22E-02	0.01*
숙박시설	7.33E-03	0.01*	7.37E-03	0.00*
지하철 역세권	6.00E-04	0.02*	5.51E-04	0.03*
R-squared	0.574489		0.584551	
Lagcoeff.			(Rho): 0.15, p=0.00*	

주: *p < 0.05, ** p < 0.1.

형의 독립변수 중 업무시설지표가 OLS모형에서는 유의하였으나 SEM모형에서는 유의하지 않은 것으로 나타났는데, 이는 OLS모형이 공간자기상관을 통제하지 못하여 변수의 유의성 검증에서 부정확한 추론에 도달한 것으로 판단된다(김광구, 2003).

기타목적의 통행유입·유출모형(표 15, 16)은 모두 OLS모형이 선택되었는데, 이는 특정한 목적을 가진 통행의 종류 하나가 정해진 것이 아니라 여러 기

표 13_공간회귀분석 결과(업무통행 유출량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	-1.10E+02	0.74	-1.08E+02	0.72
종사자 수	3.72E-01	0.00*	3.82E-01	0.00*
가구 수	1.69E-01	0.00*	1.79E-01	0.00*
건축물 총연 면적	8.38E-04	0.00*	7.02E-04	0.01*
문화 및 집회 시설	8.08E-03	0.02*	8.69E-03	0.01*
공장	-2.18E-03	0.00*	-2.14E-03	0.00*
방송통신시설	4.06E-01	0.00*	3.43E-01	0.00*
업무시설	-2.79E-03	0.07*	-2.03E-03	0.16
도시철도연장	0.29	0.02*	0.19	0.09**
R-squared	0.847648		0.851806	
Lagcoeff.			(Lambda): -0.25, p=0.00*	

주: *p < 0.05, ** p < 0.1.

표 14_공간회귀분석 결과(업무통행 유입량)

구분 변수	OLS		SEM	
	계수	p	계수	p
상수	2.83E+02	0.44	4.15E+02	0.21
2, 3차 종사자 수	3.99E-01	0.00*	4.24E-01	0.00*
고용자 수	7.23E-02	0.01*	6.13E-02	0.01*
제2종 근생시설	3.98E-03	0.00*	3.30E-03	0.00*
관광휴게시설	6.65E-01	0.00*	6.12E-01	0.00*
지하철 역세권	1.07E-03	0.00*	8.48E-04	0.00*
R-squared	0.864148		0.873279	
Lagcoeff.			(Lambda): -0.38, p=0.00*	

주: *p < 0.05.

표 15_공간회귀분석 결과(기타통행 유출량)

구분 변수	OLS	
	계수	p
상수	-3.93E+02	0.30
3차종사자수	1.89E-01	0.00*
대학생수	2.21E-01	0.00*
가구수	4.75E-01	0.00*
차량수/1,000명	1.39E+00	0.00*
공동주택	2.15E-03	0.00*
제2종근린생활시설	4.00E-03	0.00*
위락시설	2.42E-01	0.00*
도시철도연장	5.03E-01	0.00*
R-squared	0.78702	

주: *p < 0.05.

표 16_공간회귀분석 결과(기타통행 유입량)

구분 변수	OLS	
	계수	p
상수	6.18E+02	0.10
학생수	1.99E-01	0.00*
3차산업 종사자수	6.61E-01	0.00*
제2종 근린생활시설	8.91E-03	0.00*
중심상업지역	9.60E-02	0.00*
버스정류장수	1.12E+01	0.07**
도시철도연장	7.86E-01	0.00*
R-squared	0.854275	

주: *p < 0.05, ** p < 0.1.

타 목적이 혼합되어 있기 때문에 공간자기상관을 탐색하기 어려운 것으로 보인다. 또한 기타목적 통행발생모형에 도시철도연장, 차량 수/1,000명, 버스정류장수 등 특정목적의 통행을 유발하는 것이 아닌 교통과 관련된 지표가 포함되어 있는 것이 특징이며, 제2종근린생활시설과 3차산업 종사자수가 유출, 유입모형에 모두 영향을 주는 것으로 나타났다.

이 밖에 OLS모형과 공간회귀분석모형에 대한 변

수들의 유의수준(p-value)을 비교한 결과 대체적으로 공간회귀모형의 변수들의 유의수준이 높은 것으로 분석되었다.

한편 본 연구 모형의 검증을 위해 수도권교통본부에서 2002년 수도권 가구통행실태조사를 바탕으로 예측한 2006통행발생 유출·유입량 값과 RMSE%(Root Mean Square Error)³⁾를 비교해보았다(<그림 5> 참조). 쇼핑통행 유입모형과 기타통행 유입모형을 제외한 10개의 모형이 우수한 것으로 나타났다. 목적별로 적게는 1%에서 많게는 18% 이상의 감소를 나타내었다.

VI. 결론

본 연구는 통행발생모형 구축에 주로 이용되는 선형 회귀분석의 공간자기상관에 대한 문제점을 지적하고 이를 개선하기 위해 공간자기상관을 탐색하였으며, 공간회귀분석을 활용하여 목적별 통행발생모형을 구축하였다. 기타목적 유입·유출 모형과 통학통행 유출모형을 제외한 9개의 모형이 공간종속성 검정결과에 따라 공간회귀모형으로 선택되었으며, 회귀모형과의 비교를 통해 설명력이 상승하였음을 확인할 수 있었다.

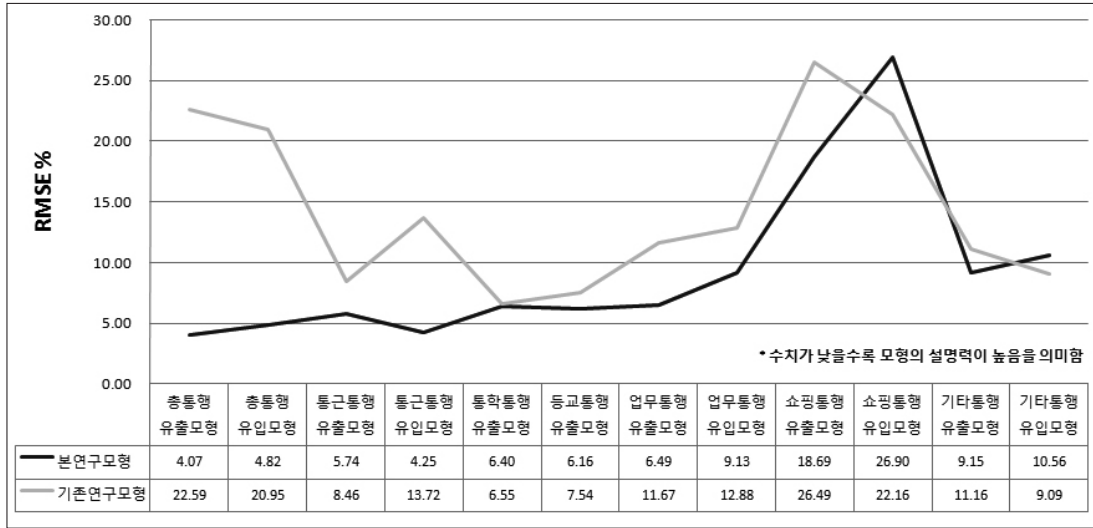
또한 기존 연구들에서 시도되었던 지역적·물리적 특성 지표를 보완하고 새롭게 추가하여 모형의 설명력을 높였으며, 모형 정립을 통해 각 목적별 통행발생량에 상식적으로 영향을 미치는 지역적 특성 지표(예: 판매시설 연면적과 지하철 역세권 지표가 쇼핑통행 유출모형의 독립변수로 포함되었음)에 대하여 그 영향력을 입증할 수 있었다.

따라서 본 연구는 사회경제지표를 이용하여 선형

3) $RMSE = \sqrt{\frac{\sum(T_i^o - T_i^p)^2}{N}}$, $RMSE\% = \frac{RMSE}{\text{실측지평균}} \times 100$, T_i^o : i지역 실측치, T_i^p : i지역 예측치, N: 대상개수

※ 이진 예측자료와의 소준(행정동) 개수 차이로 인해 중준(구)단위로 예측 값을 종합 후 RMSE%를 도출하여 비교함.

그림 5_ 기존 연구 예측 값과 본 연구 예측 값의 RMSE% 비교



회귀모형을 구축한 기존 연구들과는 달리, 토지이용 및 지역적 특성자료를 통한 공간회귀모형을 구축함으로써 통행발생량추정모형의 설명력을 향상시켰다는 점에서 의의가 있다고 할 수 있다. 하지만 분석이 서울로 한정되어 지역적 범위가 다소 제한적이므로 향후 인천 및 수도권 지역을 포함하는 광범위한 분석이 필요할 것이다. 또한 한 개 연도를 기준으로 분석되었다는 한계점을 극복하기 위하여 향후 시계열 자료를 이용한 추이분석으로 서울의 다양한 통행특성 변화를 연구할 필요가 있을 것이다.

참고문헌 ●●●●

국토연구원. 2004. 공간분석기법. 서울 : 한울아카데미.
 김광구. 2003. “공간자기상관(spatial autocorrelation)의 탐색과 공간회귀분석(spatial regression)의 활용”. 정책분석평가학회보 제13권 제1호. pp273-294.
 김순관. 1998. 서울시 교통수요 예측모형 정립(I)-통행발생 및 통행분포. 서울 : 서울시정개발연구원.
 김진자 · 이종호. 2004. “수도권지역의 통행발생모형의 검증: 회귀

모형과 카테고리모형을 중심으로”. 대한교통학회지 제 22권 제3호. pp49-58.
 김현기 · 이상민 · 김강수. 2004. “베이지안 망을 이용한 통행발생 모형의 설계 및 구축”. 대한교통학회지 제22권 제7호. pp79-90.
 변필성. 2009. “공간계량경제모델링-공간적 자기상관과 공간회귀모형”. 알기 쉬운 국토연구방법론. 경기 : 국토연구원. pp315-326.
 송재인 · 나승원 · 추상호. 2011. “토지이용 특성을 반영한 통행발생모형 추정 연구”. 한국ITS학회논문지 제10권 제6호. pp126-139.
 수도권교통본부. 2009. 수도권 장래교통 수요예측 및 대응방안 연구. 서울 : 수도권교통본부 기획조정부 교통정보팀.
 유완 · 김형진 · 권성진. 1999. “개별통행행태를 고려한 통행발생 예측”. 대한교통학회 제36회 학술발표회논문집. pp77-82.
 이학식 · 임지훈. 2011. SPSS 18.0 매뉴얼. 서울 : 집현재.
 장수은 · 김대현 · 임강원. 2000. “신경망 이론을 이용한 통행발생모형연구 : 선형/비선형 회귀모형과의 비교”. 대한교통학회지 제18권 제4호. pp95-105.
 정경석 · 문태현 · 정재희 · 허선영. 2009. “GIS와 공간통계기법을 이용한 시 · 공간적 도시범죄 패턴 및 범죄발생 영향요인 분석”. 한국지리정보학회지 제12권 제1호. pp12-25.
 정지교. 2001. “지역특성을 고려한 통행목적별 통행발생 예측모형”. 연세대학교 석사학위논문.

- Anselin, Luc. & Bera, Anil K. 1998. *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics: Handbook of Applied Economic Statistics*. New York : Marcel Dekker.
- Anselin, Luc. 2005. *Exploring Spatial Data with GeoDa: A Workbook*. Urbana : University of Illinois Urbana-Champaign.
- Kim, J. H., Chung, J. H., Kim, I. and Lim, U. 2009. "Spatial Dependence in Trip Generation Models: A Spatial Econometric Perspective". TRB 88th(2009) Annual Meeting Compendium of Papers DVD.
- Washington, S. and Wolf, J. 1997. "Hierarchical Tree-Based Versus Ordinary Least Squares Linear Regression Models: Theory and Example Applied to Trip Generation". *Journal of the Transportation Research Board* vol.1581. pp82-88.
- Yao, L., Guan, H. and Yan, H. 2008. "Trip Generation Model Based on Destination Attractiveness". *Tsinghua Science and Technology* vol.13, no.5. pp632-635.

-
- 논문 접수일: 2012. 4. 9
 - 심사 시작일: 2012. 5. 2
 - 심사 완료일: 2012. 5. 27

Developing Trip Generation Models Using Spatial Regression Analysis : A Case for Seoul, Korea

Keywords: Trip Generation Model, Spatial Regression, Spatial Autocorrelation, Household Travel Survey, Land Use, Seoul City

The paper develops trip generation models using the 2006 Seoul Household Travel Survey data for trip production and attraction models by the total and purpose(commute, school, shopping, business, others). The existing models frequently use linear regression models assuming independence among the error terms. However, traffic analysis zones(TAZs) are not simply distributed at random while having autocorrelation each other. To control such autocorrelation, spatial regression models need to be considered rather than linear regression models. Therefore, the paper examines the autocorrelation for trip production and attraction, and verify if the spatial regression models are suitable. Then, SEM(Spatial Error Model) or SLM(Spatial Lagged Model) is estimated according to the spatial regression decision process. The results indicate that spatial regression model is more reliable for most trip generation models except the school attraction model and other production/attraction models. The paper provides the improved trip generation models by applying the spatial interactions among TAZs and land use data.

공간회귀분석을 활용한 통행발생모형 추정 연구 : 서울특별시를 중심으로

주제어: 통행발생모형, 공간회귀분석, 공간자기상관, 가구통행실태조사, 토지이용특성, 서울특별시

본 논문은 4단계 교통수요 추정법 중 첫 번째 단계인 통행발생모형을 개선하기 위한 연구이다. 2006년 가구통계조사 전수화자료를 바탕으로 총통행 및 목적별 통행 유입·유출량에 대한 모형 정립을 목적으로 한다. 대상지는 서울특별시 전역으로 하며 분석단위는 행정동(소존)이다. 기존의 통행발생모형들을 살펴보면 대다수가 선형회귀모형을 이용하여 통행발생모형을 구축하였다. 선형 회귀모형은 기본적으로 오차항의 독립성을 가정하고 있는데, 행정동(소존)별 통행량과 같이 공간을 근거로 하는 데이터들은 공간상에 무작위적으로 분포하고 있지 않으며 공간자기상관의 특징을 가지므로 오차항의 독립성이라는 기본가정을 위배하게 된다. 이러한 공간자기상관을 통제하기 위해 일반회귀분석이 아닌 공간회귀분석을 이용하게 되는 것이다. 따라서 본 연구에서는 총통행량 및 목적별 통행량(통근, 통학, 쇼핑, 업무, 기타)에 대한 공간자기상관 존재를 판단하고, 그 여부에 따라 일반회귀분석 또는 공간회귀분석을 선택하며, 공간회귀모형 선택 시 그 프로세스에 따라 SEM모형(Spatial Error Model, 공간오차모형), SLM모형(Spatial Lagged Model, 공간시차모형)을 선택하여 최종 모형을 도출하였다. 분석 결과 기타목적 유입·유출 모형과 통학통행 유출모형을 제외한 9개의 모형이 공간회귀모형으로 선택되었으며, 회귀모형과의 비교를 통해 설명력이 상승하였음을 확인할 수 있었다. 또한 본 연구에서는 기존의 모형 연구에서 다루었던 변수 이외의 토지이용지표, 사회경제지표등을 추가 및 보완하여 분석 결과의 설명력을 상승시켰다.