

# Onion yield estimation using spatial panel regression model

Sungchun Choi<sup>a</sup> · Jangsun Baek<sup>a,1</sup>

<sup>a</sup>Department of Statistics, Chonnam National University

(Received May 9, 2016; Revised June 15, 2016; Accepted June 29, 2016)

---

## Abstract

Onions are grown in a few specific regions of Korea that depend on the climate and the regional characteristic of the production area. Therefore, when onion yields are to be estimated, it is reasonable to use a statistical model in which both the climate and the region are considered simultaneously. In this paper, using a spatial panel regression model, we predicted onion yields with the different weather conditions of the regions. We used the spatial auto regressive (SAR) model that reflects the spatial lag, and panel data of several climate variables for 13 main onion production areas from 2006 to 2015. The spatial weight matrix was considered for the model by the threshold value method and the nearest neighbor method, respectively. Autocorrelation was detected to be significant for the best fitted model using the nearest neighbor method. The random effects model was chosen by the Hausman test, and the significant climate variables of the model were the cumulative duration time of sunshine (January), the average relative humidity (April), the average minimum temperature (June), and the cumulative precipitation (November).

Keywords: panel analysis, spatial econometric model, meteorological data, onion yields

---

## 1. 서론

해마다 소비자들은 채소가격의 움직임에 예민하게 반응하며, 언론과 각종 기관에서는 채소 공급 및 소비에 대한 정확한 예측과 가격변동을 낮추는 방안을 마련하라고 요구한다. 특히 채소품목 중 양파는 전체 물량의 96% (RDA, 2013)가 가정용으로 소비될 정도로 식생활에 빼놓을 수 없는 양념채소로서, 공급량과 소비량의 불균형에 따른 가격변동은 생산자와 소비자 모두에게 피해를 주는 물가안정의 불안요소로 작용하고 있으며, 주된 원인은 기상악화로 인한 불안정한 생산량에서 비롯한다. 양파는 주로 텃밭에서 재배되기 때문에 기상변화에 예민하며, 이상기후 발생 시 예년에 비해 생산량 차이가 크며 지역별 편차도 커진다. 따라서 선제적인 수급량 조절을 위해 기후변화를 고려한 농작물 생산량 예측이 매우 중요하다. 기상과 양파의 생산량에 관련된 국내 연구로 Kim (2013)은 이상기후와 양파단수량의 연관성을 알아보고자 양파주산지인 해남, 영천, 밀양을 중심으로 패널분석을 한 결과 이상기후가 양파 단수에

---

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Advanced Production Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (315012-03-1).

<sup>1</sup>Corresponding author: Department of Statistics, Chonnam National University, 77, Yongbong-ro, Buk-Gu, Gwangju 61186, Korea. E-mail: [jbaek@jnu.ac.kr](mailto:jbaek@jnu.ac.kr)

부정적인 영향을 미치는 것으로 나타났다. Nam과 Choe (2015)는 자기회귀시차(autoregressive and distributed lags) 모형으로 재배면적과 단수를 예측하고, 양과 출하시기 도매가격을 전망하였다. 생산량을 배제한 채 가격의 패턴만 고려했던 타 연구와 달리, 기상자료와 생육특성을 고려한 생산량 변수가 포함되어 양과도매가격 예측력이 향상됨을 나타냈다. 여기서는 2006년부터 2015년까지 13곳의 양과 주산지의 단수(10a당 생산량(kg))자료를 활용해 패널회귀모형을 구축하여 기상변화에 따른 생산량 추정을 하고자 한다. 이와 더불어 주산지의 위치가 대부분 전남, 경남, 경북 등 남부지방에 집중되어 있어 지역간 상관성이 존재할 것이라 예상되어 공간계량모형을 고려한 분석을 시도하였다. 본 논문은 제 2장에서는 공간패널모형에 대해 설명하고, 제 3장에서는 분석에 사용된 자료를 설명한다. 제 4장에서는 분석결과를 제시하며, 제 5장에서는 결론 및 향후과제에 대해 제안했다.

## 2. 분석모형

### 2.1. 패널모형

어떠한 특성이나 성질이 시간의 순서대로 나열되어 있는 자료를 시계열 자료(time series data)라고 하며, 반대로 특정 시점에 여러 개체의 현상이나 특성을 모아 놓은 것을 횡단면 자료(cross sectional data)라고 한다. 패널자료(panel data)는 시계열자료와 횡단면자료를 하나로 합쳐놓은 자료이다.

다음과 같이 정의된 패널 선형회귀모형에서,

$$y_{it} = \alpha + \beta x_{it} + e_{it}, \quad (\text{단 } e_{it} = \mu_i + \varepsilon_{it}, \varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2)), \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T$$

$y_{it}$ 는  $i$ 번째 패널개체(지역)의  $t$ 시점에서의 반응변수이며,  $\alpha$ 는 상수항,  $\beta$ 는 회귀계수,  $x_{it}$ 는 설명변수,  $e_{it}$ 는 오차항이다. 이때 오차항  $e_{it}$ 는 두개의 항으로 구분할 수 있는데, 시간에 따라 변하지 않는 패널의 개체특성을 나타내는  $\mu_i$ 와 시간과 패널 개체에 따라 변하는 순수한 오차항인  $\varepsilon_{it}$ 로 구성되어 있다.

패널 회귀모형에서 오차항  $\mu_i$ 를 고정효과로 볼 것인지, 확률효과로 볼 것인지에 따라 추정방법이 달라진다. 고정효과 모형에서는  $\mu_i$ 를 패널 개체별로 고정되어 있는 모수로 해석하는 반면, 확률효과 모형에서는 확률분포를 따르는 확률변수로 해석한다. 고정효과 모형인지 확률효과 모형인지 판단할 때 일차적으로 중요한 기준은 데이터에서 패널 개체의 특성을 의미하는  $\mu_i$ 에 대한 추론이다. 패널 개체들이 모집단에서 무작위로 추출된 표본의 개념이라면 오차항  $\mu_i$ 는 확률분포를 따른다고 가정할 수 있다. 그러나 주어진 패널 개체들이 모집단에서 무작위로 추출된 표본이 아니라 특정 모집단 그 자체라면 오차항  $\mu_i$  확률분포를 따른다고 말할 수 없다. 다른 측면으로는 설명변수와 오차항  $\mu_i$ 의 상관성 유무를 기준으로 고정효과모형과 확률효과모형을 구분할 수 있다. 만약 설명변수와 오차항  $\mu_i$  간의 상관성이 없는 귀무가설( $H_0 : \text{cov}(x_{it}, \mu_i) = 0$ )인 경우 고정효과 모형의 추정량( $\hat{\beta}_{FE}$ )과 확률효과 모형의 추정량( $\hat{\beta}_{RE}$ ) 둘다 일치추정량으로서 동일한 값을 가지게 된다. 하지만  $\mu_i$  값을 모수로서 추정하는 고정효과 모형에 비해 자유도를 적게 사용하는 확률효과 추정량( $\hat{\beta}_{RE}$ )이 보다 효율적이다. 설명변수와 오차항  $\mu_i$  사이에 상관성이 존재하는 경우( $H_1 : \text{cov}(x_{it}, \mu_i) \neq 0$ ), 고정효과 추정량( $\hat{\beta}_{FE}$ )은 일치성을 만족하지만 확률효과 추정량( $\hat{\beta}_{RE}$ )은 일치성을 만족하지 못하게 된다. 다음과 같이 카이제곱 분포를 기반으로 하는 하우스만 검정 통계량을 통하여 고정효과 모형과 확률효과 모형 중 어느 모형이 더 적합한가를 판정한다 (Baltagi, 2005; Greene, 2012).

$$W = \frac{(\hat{\beta}_{FE} - \hat{\beta}_{RE})^2}{\text{Var}(\hat{\beta}_{FE}) - \text{Var}(\hat{\beta}_{RE})} \sim \chi^2(k-1).$$

하우스만 검정을 통해 오차항  $\mu_i$ 를 확률변수로 간주하는 확률효과모형인지, 추정해야할 모수로 간주하는 고정효과모형인지를 선택한 후, 각각의 모형에서 확률효과 또는 그룹효과의 유의성을 검정한다. 만약 확률효과모형에서  $\sigma_\mu^2 = 0$ 의 결과가 나온다면 오차항  $\mu_i$ 는 확률변수가 아닌 상수항이 되며 패널모형이 아닌 일반회귀모형의 성격을 가지게 된다 (Min과 Choi, 2012).

## 2.2. 공간가중패널회귀모형(spatial autoregressive model; SAR)

패널 모형에는 다음과 같은 기본가정이 성립해야 한다. (가정 1) 모든  $i$  및  $t$ 에 대하여,  $E(\varepsilon_{it}) = 0$ , (가정 2) 모든  $i$  및  $t$ 에 대하여,  $\text{var}(\varepsilon_{it}) = \sigma^2$ , (가정 3) 모든  $i \neq j$  및  $t \neq s$ 에 대하여,  $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$ , (가정 4) 모든  $i$  및  $t$ 에 대하여,  $\text{cov}(x_{it}, \varepsilon_{it}) = 0$ . 만약 이러한 가정 중에서 개체 간의 상관성이 존재하여  $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$ 을 만족하지 못하게 되면, 공간적인 자기상관이 존재한다고 할 수 있으며, 자기상관을 고려한 모형을 사용해야 한다. 즉, 패널선형회귀모형의 반응변수  $y_{it}$ 가 각 패널들의 지리적 위치나 거리에 따라 서로 상관관계가 존재하면 공간적 종속성이 나타난다. 예를 들면  $t$ 시점에서  $\text{cov}(\varepsilon_{it}, \varepsilon_{js}) \neq 0$ ,  $i \neq j$ 인 경우 오차항은 공간적으로 상관관계가 존재하며 따라서 공분산은 특정한 이웃관계(neighbor relation)를 따르게 된다. 이러한 이웃관계는 소위 공간가중치행렬(spatial weight matrix)로서 표현된다.  $W_N$ 를  $N \times N$  행렬로서 지역간의 상관성을 나타내는 공간가중치행렬이라고 하자. 행렬의 원소인  $w_{ij}$ 는 지역  $i$ 와 지역  $j$  간의 상호의존성의 크기를 나타내며, 모든 지역이 인접성(근접성) 정도에 따라 동일한 가중치를 가진다고 가정한다. 또한 시간에 걸쳐 변하지 않는다고 가정하며 모든 행원소의 합은 1이 되도록 표준화 시킨다.

가중치 행렬을 구하는 방법은 크게 인접성(contiguity)을 기준으로 구축하는 방법과, 거리(distance)를 기준으로 구축하는 방법이 있다. 인접성을 고려하는 경우 인접하는 기준을 어떻게 세우느냐에 따라 방법이 달라지고, Rook, Queen, Bishop 방법 등이 있다. Rook 방법은 두 지역의 경계선이 공유되면  $w_{ij} = 1$ 을, 그렇지 않으면  $w_{ij} = 0$ 을 부여하는 방법이다. 이와 비슷하게 Bishop 방법은 두 지역이 모서리를 공유하는 경우, Queen 방법은 두 지역이 면이나 모서리를 공유하는 경우  $w_{ij} = 1$ 을 설정하게 된다. 거리를 고려하는 경우 중심선택방법, 임계거리(threshold distance) 설정 등에 따라 여러 가중치 행렬을 구할 수 있다. 예를 들어 지역의 행정구역 중심점 간의 거리를 계산하여 임계거리가 300m 안에 있는 경우  $w_{ij} = 1$ 을 부여 하거나, 중심점 간의 거리의 역수를 계산하여 가중치 행렬의 원소로 사용하는 경우이다 (Lee와 Noh, 2013). 공간가중치행렬은 한 시점에서의 지역간의 상관성을 나타내는 행렬이지만, 패널모형은 총  $T$ 시점의 시차를 고려해야 한다. 따라서 다음과 같이  $NT \times NT$  크기의  $W_{NT}$  행렬을 생성하게 된다.

$$W_{NT} = I_T \otimes W_N$$

지역간의 상관성이 존재하여 추정치의 편의가 발생하는 경우 다음과 같이 공간시차(spatial lag)를 반영한 공간 패널 모형을 구축하여 공간적인 자기상관성을 고려할 수 있다.

$$y = \rho(I_T \otimes W_N)y + X\beta + (\iota_T \otimes \mu) + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_{NT}).$$

이때  $y$ 는  $NT \times 1$  벡터로서 반응변수이며,  $\rho$ 는 공간 자기상관 회귀 계수,  $X$ 는  $NT \times k$  행렬로서 설명변수,  $\beta$ 는  $k \times 1$  벡터로서 회귀계수,  $\mu$ 는  $N \times 1$  벡터로서 개별특성효과,  $\varepsilon$ 는  $NT \times 1$  벡터로서 오차항,  $W_N$ 는  $N \times N$  공간가중치행렬이다.  $I_T$ 는  $T \times T$  크기의 단위행렬이며  $\iota_T$ 는 원소가 모두 1의 값을 가지는  $T \times 1$  크기의 벡터이다. 위의 모형에서 오른쪽에 있는  $y$ 와 관련된 항을 왼쪽으로 옮긴 후  $(I_N - \rho W_N)$ 의 역행렬이 존재한다고 가정한다면 다음과 같이 정리 할 수 있다.

$$y = [I_T \otimes (I_N - \rho W_N)]^{-1} X\beta + [I_T \otimes (I_N - \rho W_N)]^{-1}[(\iota_T \otimes \mu) + \varepsilon].$$

종속변수 혹은 독립변수에 공간적 종속성이 있는 경우 특정 독립변수의 한계적인 변화는 해당지역의 종속변수에게 영향을 미칠뿐만 아니라 다른 지역의 종속변수에게도 영향을 미칠 수 있다. 따라서 이러한 상황에서 추정회귀계수는 더이상 한계적인 총효과를 의미하지 않는다. LeSage와 Pace (2009)는 총효과를 특정독립변수의 변화가 동일 지역의 종속변수에 미치는 효과인 직접효과와, 다른 인접지역의 종속변수에 영향을 미치는 간접효과로 구분하였다. 이를 유도하기 위해 위의 모형을  $t$ 시점에서 다시 정리하면 다음과 같이 표현할 수 있다.

$$y_t = \sum_{r=1}^k S_r(W)x_r + (I_N - \rho W_N)^{-1}\mu + (I_N - \rho W_N)^{-1}\varepsilon_t,$$

이때

$$S_r(W) = (I_N - \rho W_N)^{-1}I_N\beta_r, \quad (I_N - \rho W_N)^{-1} = I_N + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots$$

이며,  $y_t$ 는  $N \times 1$  벡터로서  $t$ 시점에서  $N$ 개 지역의 반응변수값이며,  $x_r$ 은  $N \times 1$  벡터로서  $t$ 시점에서  $r$ 번째 설명변수에 대한 총  $N$ 개 지역의 관측값 벡터이다.  $\varepsilon_t$ 는  $N \times 1$  벡터로서  $t$ 시점에서 오차항이다.

특정  $t$ 시점에서 지역  $i$ 의  $r$ 번째 설명변수  $x_{ir}$ 의 한계효과를 구하기 위해 위의 모형의 양변을  $x_r$ 에 대해 편미분을 하면 다음과 같다.

$$\frac{\partial y_t}{\partial x_r} = S_r(W) = (I_N - \rho W)^{-1} \begin{pmatrix} \beta_r & 0 & \dots & 0 \\ 0 & \beta_r & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_r \end{pmatrix}.$$

이때 직접효과는  $S_r(W)$ 의 대각행렬 요소의 평균값( $n^{-1}\text{tr}(S_r(W))$ )이며 총효과는  $S_r(W)$ 의 행 또는 열의 평균( $n^{-1}\iota_n' S_r(W) \iota_n$ )이다. 간접효과는 총효과에서 직접효과와의 차이가 된다. 본 논문에서는 R package(splm) (Millo와 Piras, 2012)를 활용하여 직접효과와 간접효과의 추정치를 계산하였다. 모수에 대한 추정은 최우추정방법을 이용하며 로그우도는 아래와 같다.

$$L = \ln |I_t \otimes (I_N - \rho W_N)| - \frac{NT}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \varepsilon' \varepsilon, \quad \varepsilon = y - \rho(I_T \otimes W_N)y - X\beta.$$

### 3. 분석 자료

공간패널모형에서 반응변수로 사용한 자료는 2006년부터 2015년 까지 통계청에서 제공하는 13개 시군의 10a당 양파 생산량이다 (Table 3.1). 해당 시군은 전국 양파 주요 산지이며, 전국대비 양파 수확면적 비율은 (2010년 농림어업총조사 기준, 통계청) 무안군 19.97%, 신안군 9.35%, 창녕군 6.76%, 해남군 6.53%, 제주시 6.17%, 함천군 5.48%, 함평군 5.45%, 고흥군 2.82%, 함양군 2.60%, 청도군 1.24%, 군위군 1.11%, 의성군 1.09%, 영천시 1.01%를 점유하고 있다. 설명변수로는 농촌진흥청 (RDA, 2013)의 자료를 참조하여 양파 생육에 영향을 미치는 주요 기상변수로서 온도, 강수량, 습도, 햇빛 등을 고려하였다. 기상자료는 기상청에서 제공하는 2006년부터 2015년까지 월별 기상자료를 사용하였다 (Table 3.2). 여기서 1월부터 12월까지의 평균기온, 평균최저기온, 평균최고기온, 누적강수량, 평균상대습도, 누적일조시간, 평균지면온도에 대하여 기상자료를 구축하였고 총 84개의 월별 설명변수를 사용하였다. 양파의 수확시기인 7월을 기준으로 양파의 연간 단위당 생산량이 변화하기 때문에, 해당연도의 양파 생산량에 미치는 기상자료를 전년도 7월부터 해당년도 6월까지로 설정하여 설명변수 값으로 사용하였다.

**Table 3.1.** Onion yield(kg/10a) in major planting regions

시도	시군	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006
전남	무안	6,017	6,051	5,713	5,500	6,977	6,787	7,949	7,090	7,646	5,759
	신안	5,978	6,152	5,785	5,570	6,669	7,047	8,284	6,633	6,965	5,727
	합평	5,562	6,975	6,283	5,567	5,926	5,788	6,817	5,936	6,521	5,978
	해남	4,909	5,690	5,148	4,182	5,796	6,074	7,299	6,815	6,224	5,551
	고흥	6,808	6,638	5,862	5,419	6,964	6,509	6,542	6,168	6,247	5,016
제주	제주	6,412	7,213	6,580	6,338	7,032	5,078	5,790	6,147	6,231	5,918
경남	창녕	6,711	7,086	7,716	5,988	7,827	7,264	8,548	6,882	7,255	6,042
	합양	6,645	7,577	8,317	7,588	8,611	7,911	7,982	6,915	6,860	6,577
	합천	5,685	7,519	8,321	6,012	7,060	6,941	8,581	6,701	7,014	5,787
경북	영천	5,329	6,894	7,585	7,410	7,350	6,094	6,744	6,657	6,837	6,298
	의성	5,270	6,075	7,363	5,709	7,381	7,529	6,487	7,416	6,824	6,306
	군위	5,921	7,647	7,448	6,250	7,484	6,038	6,395	6,539	6,060	6,059
	청도	6,552	7,523	7,975	6,992	7,608	7,492	8,727	7,560	8,447	7,979

**Table 3.2.** Variables of the panel regression model

		변수	세부자료
<i>y</i>		단수: 생산량(kg/10a)	2006-2015, 지역별 10a당 양파 생산량자료
		월 평균기온(°C)	
<i>x</i>		월 평균최저기온(°C)	
		월 평균최고기온(°C)	2006-2015, 지역별 월별 기상데이터
		월 누적강수량(mm)	(7월-12월: 전년도 기상자료,
		월 평균상대습도(%)	1월-6월: 해당연도 기상자료)
		월 누적일조시간(hr)	
		월 평균지면온도(°C)	

예를 들어 2015년도 양파 단위당 생산량에 미치는 기상요인들은 2014년 7월부터 2015년 6월까지의 자료가 사용된다. 기상관측장비가 없는 지역인 경우에는 가장 인접해 있는 지역의 기상자료로 대체하였다.

설명변수 선택은 생산량과 84개의 기상데이터를 활용하여 다중회귀분석 단계선택 방법을 통해 유의한 변수만을 선택하였다 (유의수준: 진입 0.05, 제거 0.1). 다중회귀분석-단계선택 방법을 통하여 1월 누적일조시간, 4월 평균상대습도, 6월 평균최저기온, 6월 평균기온, 8월 평균최저기온, 11월 누적강수량이 선택되었다. 사용된 월별 기상자료의 기초통계량 값이 Table 3.3에 정리되어 있다.

모형의 적합성은 아카이케 정보 기준(Akaike information criterion; AIC), 베이저안 정보 기준(Bayesian information criterion; BIC)값을 활용하여 모형간의 적합도를 비교하고 전체 설명력을 확인하였다. 이때  $N^*$ 은 관측치의 수,  $k^*$ 는 추정된 모수의 수, likelihood는 최대우도를 의미한다.

$$AIC = -2 \times \ln(\text{likelihood}) + 2 \times k^*,$$

$$BIC = -2 \times \ln(\text{likelihood}) + \ln(N^*) \times k^*.$$

결정계수( $R^2$ )값은 모형선택에 의하여 설명변수를 교체함에 따라 공간자기상관회귀계수인  $\rho$ 값의 변화를 초래하여 일관적으로 결정계수( $R^2$ )의 증가(감소)를 보장할 수 없다. 대안으로서 반응변수와 예측값의

**Table 3.3.** Simple statistics

월	통계량	평균기온	평균최저기온	평균최고기온	누적강수량	평균상대습도	누적일조시간	평균지면온도
1월	Mean	0.29	-4.57	5.92	21.14	62.81	159.74	0.92
	Std. Dev	2.50	3.50	2.16	21.94	9.03	38.43	2.01
	Min	-7.02	-14.94	-1.81	0.00	42.25	33.90	-5.19
	Max	7.41	5.05	10.16	156.80	83.46	239.70	6.23
2월	Mean	2.60	-2.56	8.46	43.94	62.53	156.94	3.59
	Std. Dev	2.21	2.88	2.22	30.91	8.57	32.35	1.95
	Min	-2.61	-10.90	2.36	0.30	43.76	59.60	-0.68
	Max	9.09	5.91	13.34	137.50	81.27	225.10	9.00
3월	Mean	6.91	1.09	13.17	67.05	60.43	200.45	8.83
	Std. Dev	1.46	2.15	1.89	39.57	8.25	39.53	1.43
	Min	2.96	-4.45	8.53	7.00	42.35	109.30	5.25
	Max	10.78	7.52	17.30	171.90	80.17	263.20	12.36
4월	Mean	12.09	5.99	18.53	96.99	61.83	200.82	14.86
	Std. Dev	1.40	1.95	1.94	55.34	6.87	25.38	1.65
	Min	7.98	1.47	13.38	20.00	50.81	150.30	10.66
	Max	15.08	11.95	22.75	311.10	83.30	247.10	18.39
5월	Mean	17.76	11.87	24.25	113.20	65.50	225.28	21.55
	Std. Dev	0.97	1.42	1.94	55.45	6.74	38.37	1.66
	Min	14.82	8.41	19.83	22.50	53.20	66.40	17.39
	Max	19.83	15.85	27.98	297.00	82.43	298.70	25.74
6월	Mean	21.60	17.34	26.81	125.23	74.01	156.67	25.56
	Std. Dev	1.13	1.21	1.84	83.09	6.98	32.83	1.85
	Min	17.20	14.73	21.72	23.60	58.09	71.00	20.57
	Max	23.66	20.28	30.52	431.10	91.13	244.80	30.42
7월	Mean	25.12	21.87	29.38	280.29	80.68	133.22	27.97
	Std. Dev	1.37	1.33	1.77	152.54	6.31	33.54	2.00
	Min	21.37	18.89	24.66	14.70	70.69	39.30	23.59
	Max	28.66	25.73	32.99	738.50	95.02	247.10	34.02
8월	Mean	25.84	22.40	30.49	262.85	79.23	162.23	28.95
	Std. Dev	1.52	1.59	1.95	150.52	6.37	50.20	2.14
	Min	21.47	18.92	24.81	48.10	65.38	66.10	24.40
	Max	29.14	26.35	34.75	656.10	96.29	280.00	33.85
9월	Mean	21.36	17.25	26.73	146.21	76.92	157.81	24.36
	Std. Dev	1.34	1.92	1.27	149.97	5.26	30.88	1.61
	Min	18.32	13.03	23.33	9.00	61.50	61.40	20.86
	Max	25.07	22.81	29.46	880.00	93.33	219.00	29.99
10월	Mean	15.40	9.98	22.21	51.03	71.45	198.92	17.68
	Std. Dev	1.59	2.57	1.34	34.18	5.11	26.66	1.50
	Min	11.94	4.57	19.09	4.00	55.17	131.90	14.53
	Max	19.86	16.67	25.45	167.50	87.79	244.20	22.01
11월	Mean	8.64	3.40	14.93	48.02	67.58	157.71	9.24
	Std. Dev	2.21	3.34	1.41	43.56	6.48	29.82	2.01
	Min	4.16	-3.38	11.23	0.00	51.33	67.20	5.20
	Max	15.69	13.42	18.38	231.30	85.67	220.30	14.91
12월	Mean	1.80	-2.94	7.33	30.33	65.13	147.98	2.20
	Std. Dev	2.55	3.40	2.05	25.68	8.19	35.20	2.09
	Min	-4.85	-12.02	2.25	0.10	48.56	25.70	-2.57
	Max	9.57	7.46	11.77	131.00	84.36	207.30	8.48

**Table 3.4.** Minimum threshold method

	고흥	군위	무안	신안	영천	의성	제주	창녕	청도	함양	함평	합천	해남
고흥	0	0	0.1844	0.1441	0	0	0	0	0	0.1485	0.1800	0.1253	0.2177
군위	0	0	0	0	0.2419	0.3551	0	0.1021	0.1350	0.0706	0	0.0952	0
무안	0.1070	0	0	0.3215	0	0	0	0	0	0	0.3033	0	0.2682
신안	0.0958	0	0.3680	0	0	0	0	0	0	0	0.3666	0	0.1696
영천	0	0.2853	0	0	0	0.1860	0	0.1296	0.2220	0.0739	0	0.1032	0
의성	0	0.4264	0	0	0.1894	0	0	0.0962	0.1176	0.0760	0	0.0945	0
제주	0	0	0	0	0	0	0	0	0	0	0	0	1
창녕	0	0.1173	0	0	0.1263	0.0919	0	0	0.2717	0.1251	0	0.2677	0
청도	0	0.1561	0	0	0.2177	0.1132	0	0.2736	0	0.0907	0	0.1487	0
함양	0.0972	0.0982	0	0	0.0873	0.0880	0	0.1516	0.1091	0	0.0898	0.2787	0
함평	0.1089	0	0.3163	0.3340	0	0	0	0	0	0.0830	0	0	0.1578
합천	0.0668	0.1079	0	0	0.0992	0.00892	0	0.2642	0.1457	0.2270	0	0	0
해남	0.1662	0	0.3528	0.1948	0	0	0.0872	0	0	0	0.1990	0	0

**Table 3.5.** Maximum threshold method

	고흥	군위	무안	신안	영천	의성	제주	창녕	청도	함양	함평	합천	해남
고흥	0	0.0510	0.1215	0.0950	0.0501	0.0473	0.0589	0.0737	0.0603	0.0978	0.1189	0.0825	0.1434
군위	0.0309	0	0.0258	0.0257	0.2051	0.3010	0.0165	0.0866	0.1145	0.0598	0.0288	0.0807	0.0245
무안	0.0782	0.0275	0	0.2348	0.0260	0.0264	0.0397	0.0338	0.0293	0.0482	0.2216	0.0388	0.1959
신안	0.0681	0.0304	0.2616	0	0.0285	0.0295	0.0392	0.0360	0.0316	0.0522	0.2606	0.0417	0.1206
영천	0.0351	0.2371	0.0283	0.0279	0	0.1546	0.0190	0.1078	0.1845	0.0615	0.0312	0.0858	0.0273
의성	0.0337	0.3536	0.0291	0.0292	0.1571	0	0.0187	0.0797	0.0975	0.0630	0.0325	0.0784	0.0275
제주	0.1136	0.0527	0.1192	0.1055	0.0523	0.0507	0	0.0634	0.0578	0.0717	0.1015	0.0663	0.1453
창녕	0.0479	0.0930	0.0341	0.0327	0.1001	0.0729	0.0214	0	0.2154	0.0992	0.0379	0.2122	0.0332
청도	0.0407	0.1273	0.0306	0.0297	0.1775	0.0923	0.0202	0.2231	0	0.0739	0.0337	0.1212	0.0298
함양	0.0777	0.0785	0.0595	0.0578	0.0697	0.0703	0.0295	0.1211	0.0872	0	0.0717	0.2227	0.0543
함평	0.0814	0.0327	0.2364	0.2496	0.0306	0.0314	0.0361	0.0400	0.0343	0.0620	0	0.0474	0.1179
합천	0.0545	0.0880	0.0398	0.0384	0.0809	0.0727	0.0227	0.2155	0.1189	0.1852	0.0456	0	0.0378
해남	0.1193	0.0337	0.2533	0.1399	0.0324	0.0321	0.0626	0.0425	0.0368	0.0569	0.1429	0.0475	0

상관계수 제곱값인  $\text{corr}^2(Y, \hat{Y})$ 을 적합도 검정의 측도로 사용하였다.

$$\text{corr}^2(Y, \hat{Y}) = \frac{[(Y - \bar{Y})'(\hat{Y} - \bar{\hat{Y}})]^2}{[(Y - \bar{Y})'(Y - \bar{Y})][(\hat{Y} - \bar{\hat{Y}})'(\hat{Y} - \bar{\hat{Y}})]},$$

이때  $\bar{Y}$ 는 반응변수의 평균이며  $\bar{\hat{Y}}$ 는 예측값의 평균이다. 특정 변수의 변동으로 인해 일관적인 값을 갖지 못하는 결정계수( $R^2$ )와는 달리 예측값과 실제값간의 상관성만을 확인하기 때문에 잔차에 영향을 받지 않아 얼마나 잘 적합이 되었는지 확인하는데 이점을 가지고 있다 (Verbeek, 2000, p.21).

공간가중치행렬( $W_N$ )의 원소( $w_{ij}$ )를 결정하는 방법은 한곳에 밀집되어 있지 않고 전남, 경남, 경북, 제주로 떨어져 있어 인접여부의 방법( $w_{ij} = 1$  or  $w_{ij} = 0$ )은 사용하지 않고, 지역간 거리를 활용하여 생성하였다. 지역간의 공간적인 자기상관성을 나타내는 기준은 임계치 설정방법과 최근거리 설정방법으로 구분하였다.

**Table 3.6.** Closest distance method  $m = 2$ 

	고흥	군위	무안	신안	영천	의성	제주	창녕	청도	함양	함평	합천	해남
고흥	0	0	0.4586	0	0	0	0	0	0	0	0	0	0.5414
군위	0	0	0	0	0.4052	0.5948	0	0	0	0	0	0	0
무안	0	0	0	0.5145	0	0	0	0	0	0	0.4855	0	0
신안	0	0	0.5010	0	0	0	0	0	0	0	0.4990	0	0
영천	0	0.5624	0	0	0	0	0	0	0.4376	0	0	0	0
의성	0	0.6925	0	0	0.3075	0	0	0	0	0	0	0	0
제주	0	0	0.4506	0	0	0	0	0	0	0	0	0	0.5494
창녕	0	0	0	0	0	0	0	0	0.5038	0	0	0.4962	0
청도	0	0	0	0	0.4431	0	0	0.5569	0	0	0	0	0
함양	0	0	0	0	0	0	0	0.3523	0	0	0	0.6477	0
함평	0	0	0.4864	0.5136	0	0	0	0	0	0	0	0	0
합천	0	0	0	0	0	0	0	0.5378	0	0.4622	0	0	0
해남	0	0	0.6393	0	0	0	0	0	0	0	0.3607	0	0

**Table 3.7.** Closest distance method  $m = 3$ 

	고흥	군위	무안	신안	영천	의성	제주	창녕	청도	함양	함평	합천	해남
고흥	0	0	0.3168	0	0	0	0	0	0	0	0.3092	0	0.3740
군위	0	0	0	0	0.3305	0.4851	0	0	0.1844	0	0	0	0
무안	0	0	0	0.3600	0	0	0	0	0	0	0.3397	0	0.3003
신안	0	0	0.4070	0	0	0	0	0	0	0	0.4054	0	0.1876
영천	0	0.4115	0	0	0	0.2683	0	0	0.3202	0	0	0	0
의성	0	0.5815	0	0	0.2582	0	0	0	0.1603	0	0	0	0
제주	0.3005	0	0.3152	0	0	0	0	0	0	0	0	0	0.3843
창녕	0	0	0	0	0.1897	0	0	0	0.4082	0	0	0.4021	0
청도	0	0.2412	0	0	0.3362	0	0	0.4226	0	0	0	0	0
함양	0	0	0	0	0	0	0	0.2811	0.2023	0	0	0.5166	0
함평	0	0	0.3915	0.4133	0	0	0	0	0	0	0	0	0.1953
합천	0	0	0	0	0	0	0	0.4148	0.2288	0.3564	0	0	0
해남	0	0	0.4725	0.2610	0	0	0	0	0	0	0.2655	0	0

임계치 설정방법은 지역간 거리가 특정 임계치를 기준으로 가까울 경우는 서로 자기상관성이 존재한다고 판단하는 방법이다. 만약 지역간 거리가 임계치보다 가까운 경우 행렬의 원소값은 지역간 거리의 역수( $w_{ij} = \text{dist}(\text{지역}i, \text{지역}j)^{-1}$ )를 사용하며, 지역간 거리가 임계치보다 멀리 떨어진 경우는 원소의 값을 0으로( $w_{ij} = 0$ )사용한다. 지역간 자기상관성이 최소한 1곳 이상 존재하도록 설정하기 위하여 제주와 해남간의 거리인 141,900m를 최소임계치로 설정하는 방법과, 모든 지역이 서로 자기상관성이 존재하기 위해 424,037m를 최대임계치로 설정하는 방법을 사용하였다 (Table 3.4, Table 3.5). 최근거리 설정방법은 특정지역을 기준으로 가장 가까운  $m$ 번째 까지 지역은 서로 자기 상관성이 존재한다고 판단하는 방법으로 예를 들면 제주를 기준으로  $m = 3$ 인 경우 제주와 가장 가까운 고흥, 무안, 해남의 지역만 제주와 공간적인 자기 상관성을 가지게 된다. 행렬의 원소값은 위와 동일하게 자기상관성이 존재하는 경우 지역간 거리의 역수( $w_{ij} = \text{dist}(\text{지역}i, \text{지역}j)^{-1}$ )를 사용하며 그 외 원소의 값은 0으로( $w_{ij} = 0$ )사용한다. 최근거리는  $m = 2$ ,  $m = 3$ ,  $m = 4$  총 3가지 방법을 사용하였다 (Table 3.6, Table 3.7, Table 3.8). 공간데이터분석을 수행하는 GeoDa프로그램(<http://geodacenter.asu.edu>)을 활용하여 가중치행렬을 생성하였으며 행을 기준으로 표준화 시켜 행원소의 합이 1이 되도록 변경하였다.



**Table 3.8.** Closest distance method  $m = 4$ 

	고흥	군위	무안	신안	영천	의성	제주	창녕	청도	함양	함평	합천	해남
고흥	0	0	0.2524	0	0	0	0	0	0	0.2032	0.2464	0	0.2980
군위	0	0	0	0	0.2900	0.4257	0	0.1224	0.1619	0	0	0	0
무안	0.1070	0	0	0.3215	0	0	0	0	0	0	0.3033	0	0.2682
신안	0.0958	0	0.3680	0	0	0	0	0	0	0	0.3666	0	0.1696
영천	0	0.3467	0	0	0	0.2260	0	0.1575	0.2697	0	0	0	0
의성	0	0.5141	0	0	0.2283	0	0	0.1159	0.1417	0	0	0	0
제주	0.2349	0	0.2464	0.2182	0	0	0	0	0	0	0	0	0.3005
창녕	0	0	0	0	0.1597	0	0	0	0.3436	0.1582	0	0.3385	0
청도	0	0.1961	0	0	0.2734	0	0	0.3437	0	0	0	0.1867	0
함양	0	0.1540	0	0	0	0	0	0.2378	0.1711	0	0	0.4371	0
함평	0.1188	0	0.3450	0.3642	0	0	0	0	0	0	0	0	0.1721
합천	0	0.1449	0	0	0	0	0	0.3547	0.1956	0.3048	0	0	0
해남	0.1820	0	0.3865	0.2135	0	0	0	0	0	0	0.2180	0	0

**Table 4.1.** Estimated regression coefficients by threshold methods

변수	최소임계치		최대임계치	
	Estimate	$p$ 값	Estimate	$p$ 값
1월 누적일조시간	3.3824	0.0599·	3.4981	0.0512·
4월 평균상대습도	-43.6638	0.0000***	-40.8556	0.0001***
6월 평균최저기온	173.0204	0.0050**	143.3902	0.0202*
11월 누적강수량	-6.0565	0.0000***	-5.5035	0.0000***
$\phi$	0.2643	0.0827·	0.2949	0.0787·
$\rho$	0.1990	0.0552·	0.3056	0.0134*
Hausman	0.9976		0.9372	
AIC	2052.3790		2050.3530	
BIC	2066.7170		2064.6900	
corr <sup>2</sup>	0.4520		0.4500	

(\*\*\*):  $p$ -values between 0.000 and 0.001, (\*\*): between 0.001 and 0.01, (\*): between 0.01 and 0.05, (·): between 0.05 and 0.1; AIC = Akaike information criterion, BIC = Bayesian information criterion.

#### 4. 결과

공간가중치행렬 구성방법에 따른 공간패널효과 모형의 추정결과는 각각 Table 4.1과 Table 4.2와 같다. 모형 추정시 회귀계수가 유의하지 않은 변수가 발생할 경우 해당되는 변수는 제외시킨 후 다시 추정하는 과정을 반복하여 6월 평균기온, 8월 평균최저기온이 제외되었으며, 최종 유의한 설명변수는 전년도 11월 누적강수량, 당해년도 1월 누적일조시간, 당해년도 4월 평균상대습도, 당해년도 6월 평균최저기온이 선택되었다. 하우스만 검정결과 모든 모형에서 귀무가설을 채택하게 되어 확률효과모형으로 추정하였다. 공간적인 자기상관성을 나타내는  $\rho$ 의 경우  $m = 3$ ,  $m = 4$ 의 방법에서 유의수준 0.001하에 유의하였으며, 최대임계치,  $m = 2$  방법에서 유의수준 0.05하에 유의하게 나타났다. 확률효과 모수인  $\phi = \sigma_{\mu}^2 / \sigma_{\varepsilon}^2$ 는 모든 모형에서 유의수준 0.1하에 유의한 결과가 나왔다.  $\text{corr}^2 = 0.464$ 로 가장 높았고 AIC, BIC 값을 비교할 시  $m = 3$ 인 경우가 가장 좋은 모형이었다.

$m = 3$ 인 경우 계수값을 직접효과와 간접효과로 구분하여 추정한 결과는 Table 4.3과 같다. 예를들어 4월 평균상대습도의 경우 직접효과는 -39.63439, 간접효과는 -13.001309, 총 효과는 -52.635699이며

**Table 4.2.** Estimated regression coefficients by closest distance methods

변수	$m = 2$		$m = 3$		$m = 4$	
	Estimate	$p$ 값	Estimate	$p$ 값	Estimate	$p$ 값
1월 누적일조시간	3.8077	0.0303*	3.7264	0.0334*	3.4728	0.0491*
4월 평균상대습도	-42.2541	0.0000***	-38.7214	0.0001***	-38.8506	0.0001***
6월 평균최저기온	162.3443	0.0069**	159.1240	0.0080**	155.9772	0.0100**
11월 누적강수량	-5.8942	0.0000***	-5.4603	0.0000***	-5.5797	0.0000***
$\phi$	0.2341	0.0936·	0.2561	0.0844·	0.2643	0.0818·
$\rho$	0.2071	0.0124*	0.2644	0.0031**	0.2672	0.0057**
Hausman	0.996		0.9997		0.9991	
AIC	2049.975		2047.9180		2048.9280	
BIC	2064.313		2062.2560		2063.2650	
corr <sup>2</sup>	0.457		0.4640		0.4610	

(\*\*\*):  $p$ -values between 0.000 and 0.001, (\*\*): between 0.001 and 0.01, (\*): between 0.01 and 0.05, (·): between 0.05 and 0.1; AIC = Akaike information criterion, BIC = Bayesian information criterion.

**Table 4.3.** Marginal effects of the estimated model

변수	Direct	Indirect	Total
1월 누적일조시간	3.814243	1.251190	5.065433*
4월 평균상대습도	-39.634390	-13.001309	-52.635699***
6월 평균최저기온	162.876038	53.428394	216.304432**
11월 누적강수량	-5.589036	-1.833377	-7.422413***

(\*\*\*):  $p$ -values between 0.000 and 0.001, (\*\*): between 0.001 and 0.01, (\*): between 0.01 and 0.05, (·): between 0.05 and 0.1; AIC = Akaike information criterion, BIC = Bayesian information criterion.

유의한 결과가 나왔다. 이는 4월 평균상대습도가 1% 증가할 경우 양파 10a당 52.635kg 생산량이 감소하는 효과가 있다는 것을 의미하며, 해당 지역에는 -39.63kg, 인접지역에는 -13.00kg의 영향을 미치게 된다. 즉 4월 평균상대습도가 양파 생산량에 미치는 영향력은 공간적인 상관성 때문에 총 효과의 75%는 해당지역에, 25%는 인접지역에 영향력이 미치게 된다.  $m = 3$ 인 추정된 모형을 이용하여 연도별 양파생산량을 각 지역별로 추정하여 실제 생산량과 비교한 결과가 Figure 4.1에 나타나 있다. 모든 지역에서 연도별 추정값의 추이가 실제 생산량과 일치함을 알 수 있으며 따라서 제안된 공간패널회귀모형이 잘 적합됨을 확인할 수 있다.

## 5. 결론 및 향후과제

노지에서 재배되는 양파의 생산량은 기후환경에 영향을 많이 받으며, 특정 지방에서 많이 생산되는 지역적인 특성을 가지고 있다. 본 논문에서는 공간적인 자기상관을 고려한 공간패널모형을 이용하여 기상변화에 따른 생산량 예측을 하였다. 하우스만 검정을 통해 확률효과모형을 선택하였으며, 공간가중치 행렬의 선택방법에 따라 공간 자기상관 회귀계수가 유의한 것으로 나타났다. 누적일조시간(1월), 평균상대습도(4월), 평균최저기온(6월), 누적강수량(11월)이 양파 생산량에 유의하였으며, 각각의 직접효과와 간접효과도 추정하였다.

본 논문에서는 양파 주산지만을 고려하였으며, 기상자료의 부족으로 기상대체시군을 활용하는 한계가 있었다. 향후 확장된 공간모형을 활용하거나 다른 방법의 공간가중치 행렬을 사용할 경우 예측력이 높아질 가능성이 있어 추후 더 많은 지역 자료를 사용한 연구가 필요하다.



Figure 4.1. Comparison between the yield and the estimate in 13 major onion planting regions.

## References

- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3th ed.), John Wiley & Sons, New York.
- Greene, W. H. (2012). *Econometric Analysis* (7th ed.), Pearson.
- Kim, Y. J. (2013). *Abnormal climate's effect on crop yield and its volatility: a case study of onions* (Thesis paper), Seoul National University.
- Lee, H. Y. and Noh, S. C. (2013). *Advanced Statistical Analytics*, MOONWOOSA.
- LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*, CRC Press Taylor & Francis Group.
- Millo, G. and Piras, G. (2012). splm: Spatial panel data models in R, *Journal of Statistical Software*, **47**, 1–38.
- Min, I. S. and Choi, P. S. (2012). *STATA Panel Data Analysis*, The Korean Association of STATA.
- Nam, K. H. and Choe, Y. C. (2015). A study on onion wholesale price forecasting model, *Journal of Agricultural Extension & Community Development*, **22**, 423–434.
- Rural Development Administration (RDA) (2013). *Agricultural Technology Guide: Onion* (e-book), <http://www.rda.go.kr>
- Verbeek, M. (2000). *A Guide to Modern Econometrics*, Wiley, Chichester.

# 공간 패널 회귀모형을 이용한 양파 생산량 추정

최성천<sup>a</sup> · 백장선<sup>a,1</sup>

<sup>a</sup>전남대학교 통계학과

(2016년 5월 9일 접수, 2016년 6월 15일 수정, 2016년 6월 29일 채택)

---

## 요약

노지에서 재배되는 양파 생산량은 기후환경에 의하여 영향을 받으며, 특정 지역에서 많이 생산되는 지역적인 특성을 가지고 있다. 따라서 생산량 예측시 기상과 지역을 동시에 고려하는 접근이 필요하다. 본 논문에서는 공간 패널 회귀모형을 이용하여 기상변화에 따른 생산량을 추정하였다. 양파 주산지 13곳에 대한 2006년부터 2015년까지의 기상 패널자료를 사용하여, 공간시차를 반영한 공간자기회귀(spatial autoregressive)모형을 사용하였다. 공간가중치 행렬은 임계치 설정방법과 최근거리 설정방법으로 나누어 분석하여, 최근 3곳까지 거리 설정방법을 사용한 모형이 최종 모형으로 선택되었으며, 자기상관성이 유의함을 보였다. 하우스만 검정을 통해 채택된 확률효과모형으로 분석한 결과 누적일조시간(1월), 평균상대습도(4월), 평균최저기온(6월), 누적강수량(11월) 등이 양파 생산량 예측에 유의한 변수로 나타났다.

주요용어: 패널분석, 공간계량모형, 기상데이터, 양파생산량

---

본 결과물은 농림축산식품부의 재원으로 농림수산식품기술기획평가원의 첨단생산기술개발사업의 지원을 받아 연구되었음 (315012-03-1).

<sup>1</sup>교신저자: (61186) 광주광역시 북구 용봉로 77, 전남대학교 통계학과. E-mail: jbaek@jnu.ac.kr,