자유주제 – ⑧

기후 변수가 주요 품목의 생산성에 미치는 영향 : 준모수적 분석*

(Effects of Climate Variables on Rice and Vegetable Productivities : A Semi-parametric Analysis)

조현경** · 조은빛** · 권오상*** · 노재선***

ABSTRACT

The aim of this study is to analyze the effects of climate variables on the productivity of rice and garlic. In order to estimate these effects, we resorted to a semi-parametric model: a penalized spline model. The model allows us to identify flexible relations between climate variables and productivity. In addition, we used CART to detect the most significant variable among climate variables. According to the result, in yield of rice and garlic, the relation with temperature turns out to be nonlinear, while the relation with precipitation is linear.

Key words : climate change, semi-parametric analysis, penalty spline regression, CART(Classification And Regression Trees)

Ⅰ. 서 론

기후변화는 이미 전 세계적으로 대책을 마련하여 진행하고 있을 만큼 중요한 이슈

* 본 논문은 농촌진흥청 공동연구서(과제번호 : PJ008962)의 지원에 의해 이루어진 것임.
** 서울대학교 대학원 농경제사회학부 농업·자원경제학 전공
*** 서울대학교 농경제사회학부 농업·자원경제학 교수

이다. 더구나 기상청에 따르면 한반도는 세계의 평균보다도 빠른 속도로 기온과 강수량이 상승하는 동시에 그 집중호우와 가뭄이 심화될 것으로 전망되고 있다. 예를 들면 2012년 4월 기온이 급격히 상승하는 고온현상이 자주 나타나 관측 이래 일 최고 기온의 4월 극값이 경신된 지역이 많았을 정도로 날씨가 더워지고 있고, 작년(2011년)에는 기록적인 집중 호우로 피해가 상당했던 반면, 올해는 가뭄이 지속되고 있다. 권원태(2006)에 따르면 20세기에 들어 한반도의 평균 기온은 1.5도 상승하였고, 겨울철의 기온상승이 뚜렷하다고 한 바 있다. 이러한 기온 상승과 강수량 변동은 여러 가지 변화를 일으키겠지만 그 중 간과할 수 없는 것이 농작물 생산에 미치는 영향이다. 노재선 외(2012)에 따르면 주요 기후변수와 이상기후 발생여부가 쌀 농업생산 단수 변화에 영향을 준다고 분석한 바 있다.

본 연구에서는 미리 변수들 간의 선형이나 비선형 관계를 가정하고 분석하는 통상적인 회귀분석법은 정확한 분석법이라고 보기 어렵다는 판단 아래 기온과 강수가 농작물생산에 미치는 영향을 준모수적 방법 중에서도 벌칙 스플라인 회귀분석(penalty spline regression)을 사용하여 분석했다. 선행연구로 해외의 경우, Solomou and Wu (1999)와 Khatri et al.(1998)의 연구에서 유럽의 기후변화에 따른 농업생산성 변화를 준모수적으로 분석한 바 있으며 국내에서도 권오상, 김창길(2008) 연구에서 기후변화가 쌀 단수변화에 미치는 영향을 준모수적 및 비모수적으로 분석한 바 있다.

본 연구에서는 기온과 강수를 비모수적 변수로 사용하고, 그 밖에 생산기술을 대리하는 시간변수인 연도와 이산변수인 지역더미는 선형으로 가정하여 기후변화와 농작물 생산성과의 관계를 집중적으로 보고자 하였다. 또한 작물별로 생산성에 영향을 크게 미치는 특정 월별기온과 강수를 분류, 회귀나무 중 CART(Classification and Regression tree)알고리즘[1)]을 이용해 추출하고, 이를 반영하여 준모수적인 분석을 진행하였다. 여러 작물을 대상으로 준모수적 분석을 하였으나 본고에는 쌀과 마늘의 분석 내용만을 보이기로 한다. 그 이유는 쌀과 마늘이 본고의 목적인 준모수적인 분석법의 특징을 가장 잘 나타내는 분석 대상이었기 때문이다.

1) 분류, 회귀나무를 생성하는 대표적인 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection, Kass (1980)), CART(Classification and Regression tree, Breiman, Friedman Olshen, Stone(1984)), C4.5(Quinlan (1993))가 있다. 분류, 회귀나무를 제공하는 소프트웨어에는 R통계의 R part 패키지가 있고, Salford System의 CART패키지와 Rulequest's See5와 Cubist, S-Plus의 tree-based models가 사용된다. 본 연구에서는 R통계 프로그램의 rpart패키지를 사용하였다.

본 연구는 CART를 이용해 생산성에 영향을 미치는 특정 기간을 선별할 수 있고 이를 통해 준모수적인 분석법인 벌칙 스플라인 회귀분석을 했을 때 그 작물의 생산성과 어떠한 관계가 있는지 보여주는 과정에 주안점을 두었다. 이를 통해 다양한 작물에 대해서도 생육 특징을 반영한 연구가 가능하다는 점을 보이고자 하였다.

본고의 제Ⅱ장은 분석에 사용한 모형을 설명하고, 제Ⅲ장은 사용된 자료와 추정 결과를 보여준다. 제Ⅳ장에선 결과를 요약하고 결론을 도출한다.

Ⅱ. 분석모형

1. 벌칙스플라인회귀분석(penalty spline regression)

본 연구에서는 기후변수가 단수에 미치는 영향에 구체적인 함수 형태를 가하지 않기 위해서 준모수적(semiparametric) 분석법을 사용하고자 한다. 비모수적(nonparametric) 분석법에 비해서 구체적인 함수 형태를 가하는 준모수적 분석법을 시도하는 이유는 많은 수의 관측치와 설명변수를 다루기 위함이다. 본고에서는 준모수적 분석법 중에서도 벌칙스플라인 회귀분석(penalty spline regression)을 사용하고자 한다.

(1) $$y_i = x_i\beta + m(z_i) + \epsilon_i \ , \ \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

이 분석 방법은 선형 관계라고 보아도 무방한 설명변수의 일부는 선형으로, 나머지는 함수 형태를 가정하지 않은 $m(z_i)$로 처리한다. 위 식의 경우, x_i는 선형 관계의 변수로, z_i는 비모수적 관계의 변수로 둔 것이다.

$m(z_i)$는 몇 개의 기저(basis) 함수를 이용해 생성된다. 기저함수는 연구자가 선택할 수 있는데, 본 연구에서는 3차 방사성(radial) 함수를 기저함수로 선택하였다. 이 경우 $m(z_i)$는 다음과 같이 근사할 수 있다.

(2) $$m(z_i) = \sum_{j=1}^{J} u_j |z_i - k_j|^3$$

식 (2)의 k_j는 매듭(knot)이라하며 연구자가 매듭의 수와 위치를 정할 수 있다. 본 연구에서는 매듭의 수를 20개로 정하였고(J=20), 그 위치는 비모수적 관계로 둔 변수의 표본을 중복되는 값없이 순서대로 나열한 후, 이 수열에서 $k_j = (\frac{j+1}{J+2})$번째 표본 분위수로 정하였다. 기저함수와 매듭의 개수 및 위치는 추정결과에 큰 영향을 미치지 않기 때문에(Ruppert et al. 2003, pp. 124-127) 기존 연구에서 보편적으로 사용되는 것들을 적용하였다.

상수항을 포함한 선형의 관계를 맺는 변수가 p개 있다고 하자. 식 (1)의 행렬식 표현을 $y = X\beta + \overline{Z}u + \epsilon$라 하면 이는 식 (3)을 최소화한다.

(3) $$\| y - X\beta - \overline{Z}u \|^2 + \lambda u^T D u$$

단, $\overline{Z_i} = \{|z_i - k_1|^3, ..., |z_i - k_J|^3\}$, $D = \begin{bmatrix} 0_{p \times p} & 0_{p \times J} \\ 0_{J \times p} & (\Omega^{1/2})^T \Omega^{1/2} \end{bmatrix}$, Ω의 (l, m)원소는 $|k_l - k_m|^3$

일반적인 선형최소자승모형과는 다르게 $\lambda u^T Du$가 더해졌기 때문에, 이렇게 추정하는 방식을 벌칙 스플라인 회귀분석이라 일컫는다. 여기서 λ는 제약식에 붙는 승수로, 이 값을 너무 크게 정하면 제약의 의미가 거의 없어지고 0으로 두면 회귀식 모형이 OLS 추정 방식과 다름이 없게 된다. 즉, λ가 너무 크거나 작으면 추정 결과가 선형으로 도출되게 된다. 따라서 앞서 제시한 매듭과 기저함수와는 다르게 λ의 값은 추정결과에 큰 영향을 미친다고 할 수 있다. 이 값을 정하는 방법으로는 잔차 제곱합(RSS)을 이용하는 Mallow의 Cp기준, Akaike의 정보기준(Akaike’s information criterion, AIC), 교차확인법(cross-validation, CV) 등과 잔차 제곱합을 이용하지 않고 추정과정에서 얻을 수 있는 분산의 추정값을 이용하는 우도함수법이 있다(Ruppert et al. pp. 112-123). 잔차 제곱합을 이용한 방법들은 표본이 많을 경우 시간이 오래 걸릴뿐더러, 때로는 구할 수 없기도 하는 경우가 있으므로, 본 연구에서는 가장 실용적이라 판단되는 우도함수법을 이용해 λ를 추정하고자 한다.

이 방식은 u를 추정할 파라미터가 아닌 $u \sim N(0, \sigma_u^2 \Omega^{-1/2} (\Omega^{-1/2})^T$를 만족하는 확률변수로 본다. 이 경우 추정모형은 다음과 같이 쓸 수 있다.

(4) $$Y = X\beta + \overline{Z}u + \epsilon, \quad cov\begin{pmatrix} u \\ \epsilon \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \Omega^{-1/2} (\Omega^{-1/2})^T & 0 \\ 0 & \sigma_\epsilon^2 I_n \end{pmatrix}$$

식 (4)에서 나오는 λ는 $\lambda=(\sigma_\epsilon^2/\sigma_u^2)$의 관계를 갖는다(Ruppert et al. p.113). 모형의 단순화를 위해 $b=\Omega^{1/2}u$, $Z=\overline{Z}\Omega^{-1/2}$로 두면 최종추정모형은 다음과 같게 된다.

$$Y=X\beta+Zb+\epsilon,\quad cov\begin{pmatrix} b \\ \epsilon \end{pmatrix}=\begin{pmatrix} \sigma_b^2 I_J & 0 \\ 0 & \sigma_\epsilon^2 I_n \end{pmatrix} \tag{5}$$

다음으로 설명변수가 목표변수에 통계적으로 영향을 미치는지, 더 나아가 선형의 영향을 미치는지 비선형적인 영향을 미치는지 확인하여야 한다. 두 가지 가설검정을 위한 귀무가설과 대립가설은 다음과 같이 쓸 수 있다.

$$H_0: E(Y|X,Z)=X\beta,\ H_1: E(Y|X,Z)=X\beta+Zb \tag{6}$$

$$H_0:\sigma_b^2=0,\ H_1:\sigma_b^2>0 \tag{7}$$

최우추정모형을 사용했으므로, 우도비 검정을 우선적으로 고려할 수 있으나 저 가설들을 위한 우도비 검정은 정확성에 문제가 있다고 알려져 있다.(Ruppert et al. pp. 105-107). 따라서 본고에서는 우도비 검정의 대안으로 F-검정법을 이용하고자 한다. R_1^2, df_1을 대립가설 모형의 R^2와 잔차 자유도라 하고, R_0^2, df_0을 귀무가설 모형의 R^2와 잔차 자유도라 하자. $df=df_0-df_1$라고 하면 다음의 통계량은 귀무가설 하에서 자유도 df와 df_1의 F-분포를 갖는다(Ruppert et al. pp. 145-151).

$$F=\frac{R_1^2-R_0^2}{(1-R_1^2)df/df_1} \tag{8}$$

본 연구에서는 위에서 설명한 작업을 수행하기 위해 R통계 프로그램의 SemiPar 패키지와 gam 패키지를 이용하였다.

2. 분류, 회귀나무 CART(Classification and Regression tree)

분류, 회귀나무는 분리기준(splitting criterion), 정지규칙(stopping rule), 가지치기(pruning)를 통해 형성된다. 분리기준은 각 마디를 두 그룹으로 분리하는 기준을 의미

하며 이는 그룹 내의 동질성과 그룹간의 이질적인 특성을 계량적으로 파악하는 기준이다. 정지규칙은 각 마디의 분리를 멈추고, 주어진 마디를 끝마디(terminal node)로 결정하는 규칙이며 이에 따라 최종나무가 결정된다. 나무를 생성할 때에는 어떤 설명변수를 이용하여 어떻게 분리하는 것이 목표변수에 가장 영향을 줄 수 있는지 계산하여 하위노드(child node)를 생성하는 분리를 한다. 이때 목표변수의 분포 정도를 순수도(purity) 또는 다른 기준을 사용해 측정하며 이를 기준으로 분리(split)을 한다. 여기서 순수도(purity)는 그 노드(node)에 속하는 자료들이 목표변수의 분포와 얼마나 유사한지를 측정하는 것이다. 고른 분포를 나타내는 노드는 순수도가 낮고 그렇지 않을수록 순수도가 높다. CART 알고리즘의 최적분리(splitting rules) 기준은 목표변수가 이산형일 때는 지니계수, 연속형일 때는 분산의 감소량(variation reduction)을 이용한다. 이 밖에도 이산형 목표변수인 경우 카이제곱통계량, 엔트로피계수, Twoing/Ordered Twoing 등의 분리기준이 사용될 수 있다. 목표변수가 연속형일 때 생성되는 나무(tree)는 회귀나무(Regression tree)라고 하며, 목표변수가 이산형일 때는 분류나무(Classification tree)라고 한다. 분류나무에서는, 그룹 내의 동질성을 각각의 범주가 나올 확률의 함수로 정의된 불순도함수(impurity function)로 측정한다. 각각의 범주의 확률이 같으면 불순도함수는 최댓값을 갖는다. 불순도의 차이는 식 (9)처럼 나타낸다(Breiman, L et al. 1984, pp. 25-27).

(9) $$\Delta i(s,t) = i(t) - p_R i(t_R) - p_L i(t_L)$$

(10) $$\Delta i(s^*,t) = \max_{s \in S} \Delta i(s,t)$$

중간마디 t에서 분리 s가 발생할 때, 중간마디 t에서 p_R 의 비율만큼 오른쪽 하위 마디(child node)인 t_R로 보내고, p_L의 비율만큼 왼쪽 하위마디인 t_L로 보내면, 식 (9)와 같은 불순도의 차이가 정의된다. 분리기준은 식 (10)과 같이 불순도의 차이인 $\Delta i(s,t)$를 최대화시키는 분리 s^*를 찾는 것이다(Breiman, L et al. p221-232).

(11) $$V = \frac{1}{n^{(t)}} \sum_{i=1}^{n^{(t)}} (y_i - \bar{y})^2$$

$n^{(t)}$는 노드 t의 관측값 수

회귀나무에서는, 식 (11)과 같은 분산을 고려하여 식 (9)의 방법처럼 예측오차를 최소화하는 것과 동일하게 생각할 수 있다. 즉, 목표변수의 분산감소량을 최대화하도록 분리한다. 이 방법은 식 (12)처럼 하위노드의 그룹 내 분산 $p_R V_R + p_L V_L$을 최소화하는 것과 동일하다.

(12) $$\Delta V = V - \frac{n_R^{(t)}}{n^{(t)}} V_R - \frac{n_L^{(t)}}{n^{(t)}} V_L$$

$n^{(t)}$는 노드 관측치 수, n_R, n_L은 하위노드의 관측치 수

이러한 분산의 감소량(ΔV)을 최대로 하는 설명변수와 그 변수의 최적 분리를 점으로 선택한다.

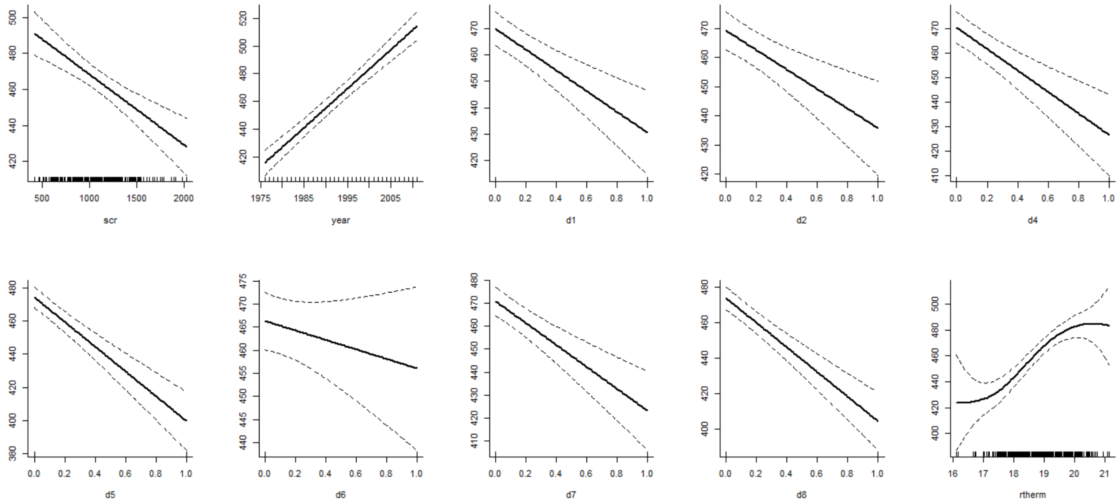
Ⅲ. 자료의 구축과 모형 추정

1. 쌀

작물별 단수 등의 자료는 농림부에서 펴낸 각 연도별『작물통계』로부터, 기후 관련 자료는 기상청으로부터 얻었다. 쌀의 경우, 1976년부터 2011년까지의 수도작 단수를 분석대상으로 활용하였다. 1976년부터 1992년까지는 일반벼와 통일벼를 구분하여 통계 조사를 실시하였는데, 본고에서는 해당기간의 단수자료로 일반벼를 채택하기로 하였다. 이는 두 품종 간 생산성이 크게 차이나는 것과 더불어 오늘날 통일벼는 거의 생산되고 있지 않다는 점을 고려했기 때문이다. 지역자료의 경우 제주도를 제외한 모든 도의 단수자료를, 기상자료의 경우 벼의 생육기간인 4월에서 10월까지의 평균 기온과 누적 강수량을 이용하였다.

생산성의 지역별 차이를 나타내기 위해 충남(d3)을 기준으로 하여 경기(d1), 충북(d2), 경북(d4), 경남(d5), 전북(d6), 전남(d7), 강원(d8)의 더미변수를 적용하였다. 우선적으로 기온과 강수량 모두 비모수적으로, 연도와 지역더미변수는 선형적으로 단수에 영향을 미친다고 가정하고 추정을 시도하였다. 그 결과 강수량의 평활 파라미터가

544200이라는 큰 값으로 추정되면서 선형의 관계를 맺는 것처럼 나왔다. 이 같은 이유로, 강수량을 선형의 관계로 놓고 추정을 시도했으며 그 결과는 <그림 1>, <표 1>과 같다. 더불어 선형 변수로 둔 강수량은 1% 수준에서 유의한 것으로 나왔다.



〈그림 1〉 쌀 벌칙스플라인 회귀분석 결과

〈표 1〉 쌀 벌칙스플라인 회귀분석 추정 결과

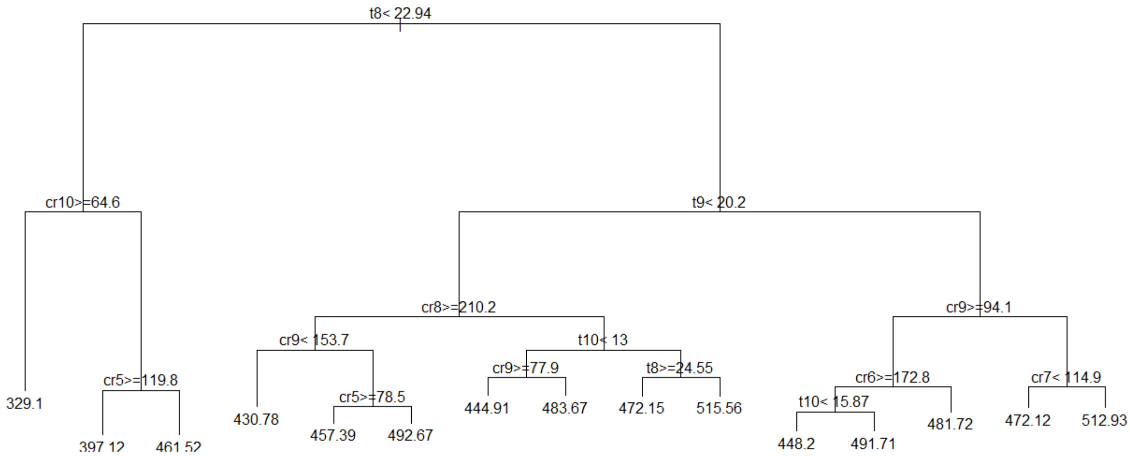
변　　수	경작기간 평균 기온과 강수량	
	추정치	t-값
상수항	-5535	-12.09***
강수량	-0.03899	-5.01***
연　도	2.82	13.01***
경기더미 d1	-39.5	-4.603***
충북더미 d2	-33.52	-3.92***
경북더미 d4	-43.74	-4.99***
경남더미 d5	-74.68	-8.213***
전북더미 d6	-10.27	-1.131
전남더미 d7	-47.98	-5.389***
강원더미 d8	-69.52	-8.084***
기온 평활 파라미터	4.819	

단, *** 1%에서 유의, ** 5%에서 유의, * 10%에서 유의

생산성을 나타내는 단수는 시간이 지날수록 증가하는 것으로 나타났다. 지역더미의 경우, 모든 도가 충남보다 단수가 낮은 것으로 나타났으나, 전북의 경우 통계적 유의성이 높지 않다. 평균기온이 낮을 때는 평균기온 증가에 따라 생산성도 증가하는 것으로 보이나, 약 19℃를 넘어서는 구간부터는 그 증가분이 줄어들다가 20℃를 넘어서게 되면 생산성은 증가하지 않고 오히려 감소하는 것으로 나타난다. 다만, <그림 1>에서 확인할 수 있듯이 약 21℃ 이상 구간에서는 통계적 유의성이 높지 않으므로 기온상승이 쌀의 단수를 감소시킬 것이라는 해석은 신뢰도가 높지는 않다.

다음으로 기온변수가 단수에 통계적으로 영향을 미치는지, 더 나아가 선형의 영향을 미치는지 비선형적인 영향을 미치는지 확인하여야 한다. 기온이 단수에 영향을 미치는지에 대한 가설 검정의 경우, F-통계량이 25.759646이며 이는 1% 수준에서 기각되었다. 기온이 비선형적인 영향을 미치는지에 대한 가설 검정의 F-통계량은 4.3728104로 5% 수준에서 기각되었다.

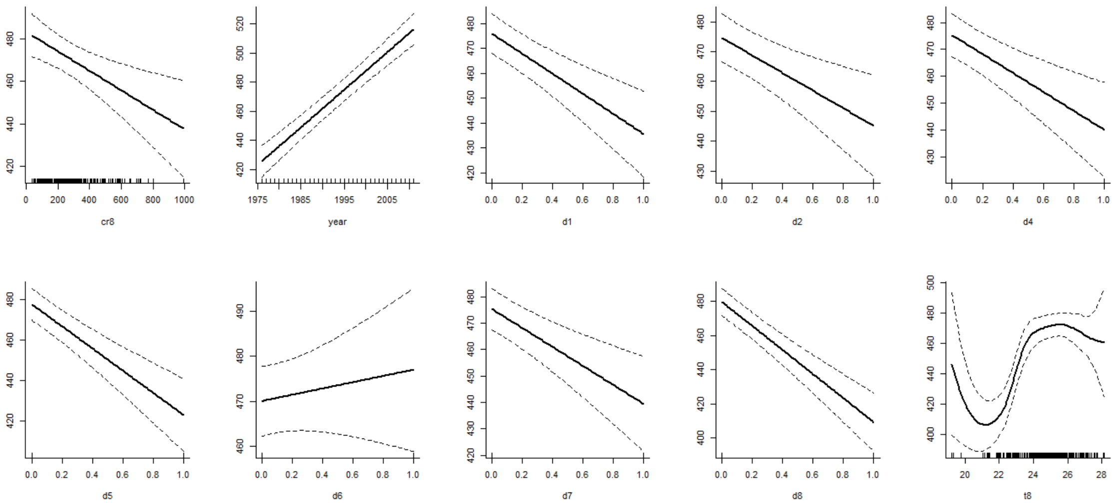
<그림 2>는 rpart 패키지를 이용해 생육기간의 월별 기온, 강수량 변수 가운데 가장 설명력 있는 변수를 추출한 결과이다.



〈그림 2〉 쌀 분류, 회귀나무 분석 결과

분류, 회귀나무 분석 결과, 8월 기온이 가장 영향력을 미치는 변수인 것으로 나타났다. 수확 전 벼가 익는 시점인 8월이 중요하게 나온 것은 작물학적으로도 어느 정도 타당하다고 볼 수 있다. <그림 3>, <표 2>는 8월 기온을 비모수적 변수로, 8월 강수량, 연도, 지역 더미를 선형 변수로 넣은 결과이다.

이전과는 다르게 전북이 충남보다 생산성이 높은 것으로 나왔으나 통계적으로 유의하지 않다. 기온의 경우, 22℃에서 26℃에 이르는 구간에서는 8월 평균 기온이 상승할수록 단수가 증가하나 26℃를 넘어서는 구간에서는 단수가 하락하는 것으로 나타났다. 강수량의 경우, 8월 누적 강수량이 증가할수록 단수가 하락하는 것으로 나타났다. 이는 수확 직전 시기에 벼가 잘 익기 위해서 고온과 적은 강수가 요구된다는 경험적 사실과 일치한다. 그러나 일정 정도를 넘어서는 고온은 오히려 생산성을 감소시킨다는 점을 시사한다.



〈그림 3〉 쌀 벌칙스플라인 회귀분석 결과 : 8월 기온

〈표 2〉 쌀 벌칙스플라인 회귀분석 추정 결과 : 8월 기온

변 수	8월 평균 기온과 누적 강수량	
	추정치	t-값
상수항	-5125	-8.705***
8월 누적 강수량	-0.04582	-3.171***
연 도	2.594	12.16***
경기더미 d1	-40.32	-4.634***
충북더미 d2	-29.44	-3.381***
경북더미 d4	-35.16	-3.994***

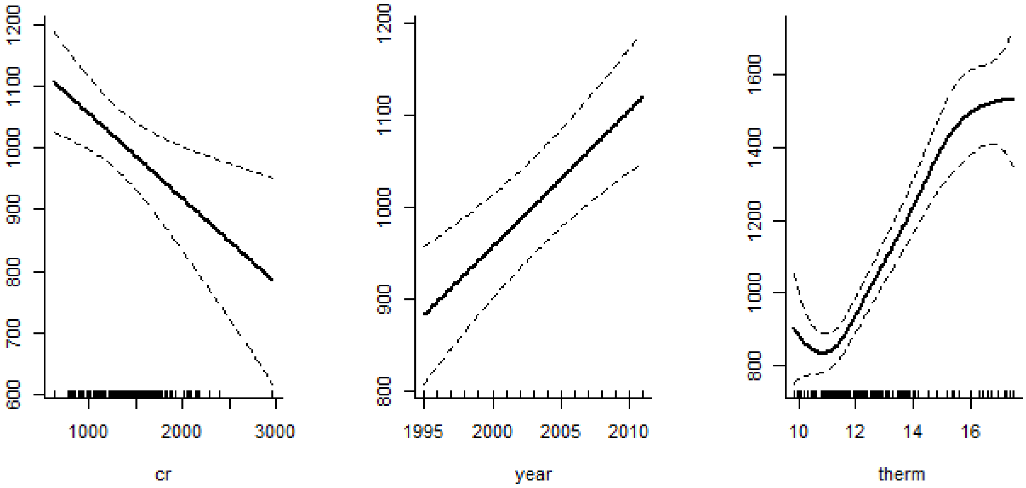
변　　수	8월 평균 기온과 누적 강수량	
	추정치	t-값
경남더미 d5	-54.78	-6.148***
전북더미 d6	6.969	0.7848
전남더미 d7	-35.91	-4.046***
강원더미 d8	-70.44	-8.096***
8월 기온 평활 파라미터	3.106	

단, *** 1%에서 유의, ** 5%에서 유의, * 10%에서 유의

2. 마늘

마늘의 경우 1995년부터 2011년까지의 통계청에서 얻은 단수자료를 분석대상으로 하였고, 경기, 충북, 충남, 경북, 경남, 전북, 전남, 강원, 제주의 단수 자료를 이용하였다. 기후자료는 동일기간에 이천, 충주, 서산, 의성, 밀양, 정읍, 해남, 원주, 서귀포 관측소에서 관측된 기창청의 자료를 이용했다. 마늘의 경우 우리나라는 10월에 파종을 하여 여름에 수확하는 작물이다. 그래서 생육기간만을 한정하지 않고 연평균 기온과 누적강수를 사용하였다. 또한 마늘은 지역별로 단수차이가 크지 않아 지역더미를 설명변수로 넣지 않았다.

<그림 4>와 <표 3>은 연평균기온을 비모수적 변수로 설정하고, 누적강수와 연도를 선형 변수 가정하여 분석한 결과다. 누적강수를 선형 변수로 가정한 이유는, 기온과 함께 누적강수를 비모수적 변수로 분석했을 때 누적강수량의 평활 파라미터가 5180으로 상당히 크게 나왔기 때문이다. 또한 강수량을 선형변수로 가정했을 때 1% 수준에서 유의한 결과가 나왔다.



〈그림 4〉 마늘 벌칙스플라인 회귀분석 결과

〈표 3〉 마늘 벌칙스플라인 회귀분석 추정 결과

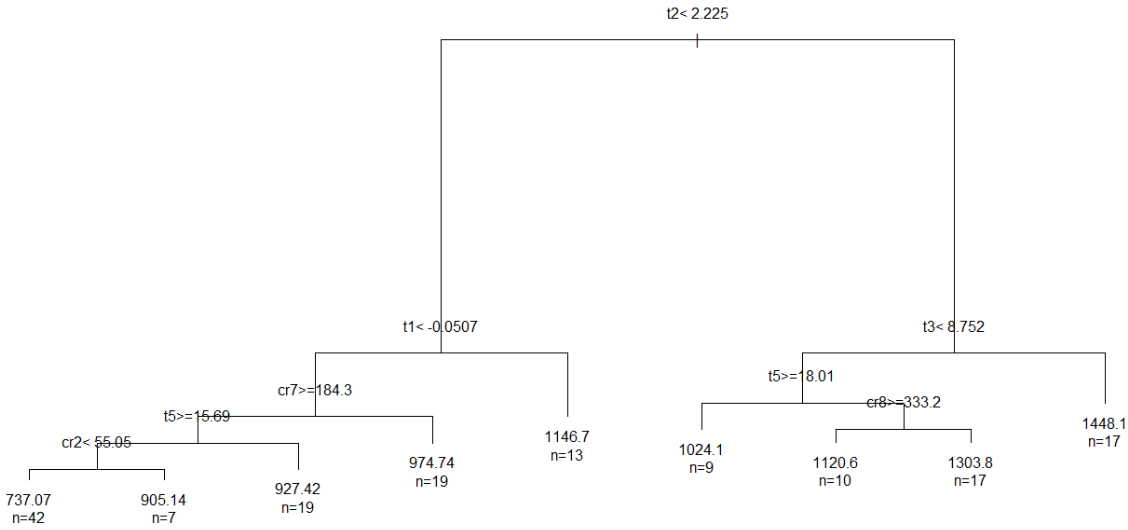
변　　수	연평균 기온과 누적강수량	
	추정치	t-값
상수항	-3.079e+04	-4.781***
연　도	1.480e+01	4.654***
누적강수	-1.375e-01	-2.889***
기온 평활 파라메터	3.817	

단, *** 1%에서 유의, ** 5%에서 유의, * 10%에서 유의

쌀과 마찬가지로 시간이 지날수록 생산성은 증가하는 것으로 나타났다. 한편, 연평균기온이 낮을 때는 기온이 상승함에 따라 마늘의 단수가 감소하는 구간이 나타났으나 약 12℃ 이상부터는 기온상승에 따라 마늘 단수도 증가하고 있다. 하지만 16℃ 이상으로 높아질 경우 신뢰구간의 폭이 커지고 있으므로 통계적 유의성이 높지 않다. 따라서 기온이 16℃ 이상으로 높을 경우 계속적인 기온 상승에 따라 마늘 단수도 증가한다는 해석은 신뢰도가 높지 않다. 기온 평활 파라미터는 3.817로 마늘 단수와 기온간의 관계가 비선형 모양으로 나왔다. 누적강수의 경우 마늘의 단수와 음의 관계를 갖는 것으로 나타나, 누적강수가 많아질수록 마늘 생산이 감소됨을 알 수 있다.

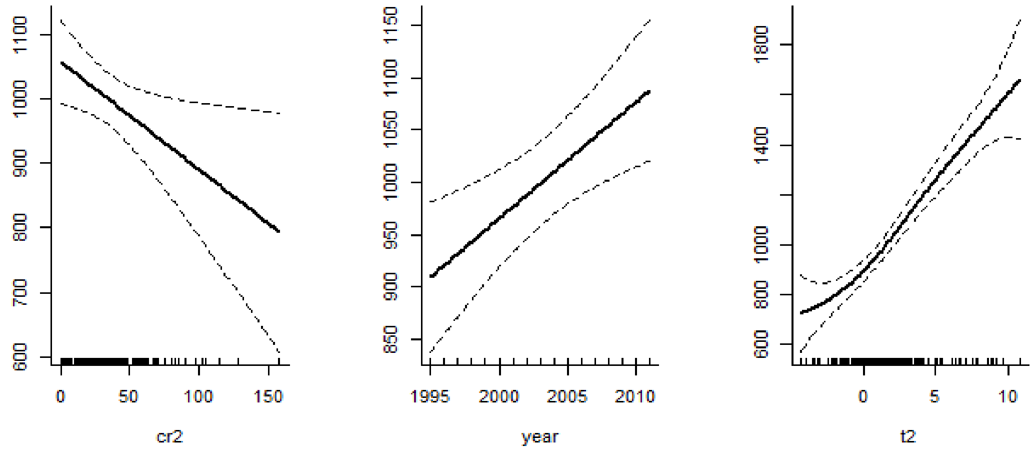
다음으로, 기온변수가 단수에 통계적으로 영향을 미치는지, 또한 선형으로 영향을

미치는지, 비선형으로 영향을 미치는지 확인하기 위해 가설검정을 시행하였다. 기온이 단수에 영향을 미치는지에 대한 가설 검정의 경우, F-통계량이 69.582374이며, 이는 1% 수준에서 기각되었다. 기온이 선형적인 영향을 미치는지에 대한 가설 검정의 F-통계량은 9.753529로 1% 수준에서 기각되었다. 이에 따라 마늘 단수에 기온이 비선형으로 영향을 미친다는 해석이 유의하다고 볼 수 있다.



〈그림 5〉 마늘 분류, 회귀나무 분석 결과

<그림 5>는 rpart 패키지를 이용해 마늘 생산에 영향을 미치는 특정 기간의 기온변수와 누적강수 변수를 추출한 결과이다. 연평균기온과 누적강수를 설명변수(explanatory variable)로 늘어놓고, 목표변수(target variable)를 마늘 단수로 하여 가장 설명력 있는 변수를 추출하였다. 그 결과 마늘생산에 대한 뿌리마디(root node)로 2월 평균기온이 나왔다. 이는 마늘이 월동작물이라는 점을 고려해볼 때, 작물학적으로도 타당해 보인다. 이를 토대로 2월 평균기온을 비모수적으로, 2월 누적강수 변수를 선형 관계를 갖는다고 가정하고 추정하였다. 그 결과는 <그림 6>과 <표 4>와 같다.



〈그림 6〉 마늘 벌칙스플라인 회귀분석 결과 : 2월 기온

〈표 4〉 마늘 벌칙스플라인 회귀분석 추정 결과 : 2월 기온

변 수	2월평균기온과 누적강수	
	추정치	t-값
상수항	-21310.000	-3.028^{***}
연 도	11.110	3.158^{***}
2월 누적강수	-1.675	-2.284^{**}
2월 평균기온 평활 파라미터	2.582	

단, *** 1%에서 유의, ** 5%에서 유의, * 10%에서 유의

2월 평균기온의 경우 기온이 상승할수록 마늘 생산이 증가하는 것으로 보인다. 2월 평균기온의 평활 파라미터는 2.582로 큰 값이 나오지 않았고 그래프도 약간의 굴곡을 갖는 비선형 모양을 보인다. 가설검정 결과 F-통계량이 3.3773163로 5% 수준에서 기각되어, 2월 평균기온은 마늘 단수와 비선형의 관계가 있다고 볼 수 있다.

Ⅳ. 요약 및 결론

본 연구는 기후변수인 기온과 강수가 작물의 생산성과 어떠한 관계를 가지는지 준

모수적 분석법인 벌칙 스플라인 회귀분석(penalty spline regression)으로 분석하였다. 이는 통상적인 회귀분석에서 설명변수간의 관계를 선형이나 비선형관계로 미리 가정하고 분석하는 데 따른 한계점을 극복할 수 있는 방법이다. 그러나 모든 변수를 비모수적 변수로 설정하여 분석하는 비모수적 분석법(nonparametric analysis)은 설명변수가 여러 개인 경우 차원의 저주(curse of dimension)문제가 발생할 수 있다. 이 점을 고려하여 본 연구에서는 기본적으로 기온과 강수변수만 비모수적인 변수로 설정하고 그 외의 변수인 연도와 지역더미는 선형으로 설정하여 작물의 생산성과 기후변수간의 관계를 파악하고자 하였다. 또한, CART를 바탕으로 작물의 생산성에 영향을 미치는 특정한 월 평균기온과 강수변수를 선별하여 작물의 생육기간 중 어느 기간의 기온과 강수가 작물 생산에 큰 영향을 미치는지 파악하였고, 그 기간의 기후변수는 작물 생산과 어떤 관계를 갖는지 파악하였다.

쌀의 경우 기온은 단수와 비선형 관계라는 해석이 유효하다고 할 수 있었고, 생육기간 평균기온이 약 20℃를 넘어서게 되면 생산성은 증가하지 않고 오히려 감소하는 것으로 나타났다. 강수의 경우 선형으로 생산성에 음의 영향을 미치는 것으로 나왔다. CART의 적용결과, 8월의 기온이 쌀 생장에 중요한 요인을 미치는 것으로 나왔으며 8월 평균기온이 26℃를 넘어서면 생산성은 감소하는 것으로 나타났다. 마늘의 경우도 기온은 마늘의 단수와 비선형관계로 영향을 미친다는 해석이 유효하였다. 또한 마늘 단수는 기온 상승에 따라 증가하는 것으로 나왔다. 누적강수의 경우 선형으로 단수에 영향을 미쳤으며, 누적강수가 증가함에 따라 마늘 단수는 감소하는 것으로 나타났다. CART를 적용하면 2월 기온이 마늘 생산에 중요한 영향을 미치는 것으로 나왔다. 이를 바탕으로 준모수적 분석을 한 결과, 2월 평균기온상승에 따라 마늘 단수는 증가했고 비선형 관계를 보였다. 2월 누적강수는 증가할수록 마늘 단수를 감소시키는 것으로 나왔으며 선형으로 영향을 미쳤다.

본고의 분석에서 변수 간의 비선형관계가 유의하지 않을 때 그 대안으로 선형 관계를 가정하고 다시 추정하였는데 그 결과는 모두 통계적으로 유의하였다. 하지만 이 점이 작물학적인 측면에서 의구심이 들게 해 보완해야 할 한계로 남아있다. 또한 분석에 사용된 자료는 지역을 기반으로 하여 수집한 자료이다. 그래서 전국구를 대상으로 합계를 낸 단수자료와, 전국의 관측소에서 관측한 기후자료를 혼합(pooling)하여 이용하는 것보다는 일반화된 결론을 내릴 때 한계가 있을 수 있다. 다만 본고의 분석은 CART를 통해 작물별로 생육주기와 특성을 고려한 준모수적 분석을 시도했다는 점에

서 상대적 장점이 있다고 할 수 있다. 따라서 향후에는 자료보완과 더불어 작물학적으로도 타당한 분석이 진행되어야 할 필요가 있다.

참 고 문 헌

권원태(2006),『한반도 기후변화 현황, 전망과 기업의 대응』, 기상연구소.

권오상 · 김창길(2008), “기후변화가 쌀 단수변화에 미치는 영향 : 비모수적 및 준모수적 분석”,『농업경제연구』49(4): 45-64.

노재선 · 권오상 · 조승현(2012), “기후변수와 쌀 단수간의 인과성 및 이상기후가 쌀 단수에 미치는 영향 분석”,『농업경제연구』53(1): 21-39.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.(1984), Classification and regression trees, Chapman and Hall, Belmont, CA, Wadsworth.

Ruppert, D., M. P. Wand and R. J. Carroll.(2003), Semiparametric Regression, Cambridge University Press.

John Ross Quinlan(1993), C4.5:programs for machine learning, Morgan Kaufmann Publishers.

Kass, Gordon V.(1980), “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, *Applied Statistics*, 29(2): 119–127.

Solomou, S. and W. Wu(1999), “Weather Effects on European Agricultural Output, 1850-1913”, *European Review of Economic History* 3: 351-737.