

# 대구 지역 경제성장 요인 분석: 소재·부품·장비 산업구조와 회귀모형 및 랜덤포레스트 모형을 활용한 통계적 검증

진민준

경북대학교, 데이터사이언스학과

jinmin1569@nate.com

## Analysis of Factors Influencing Economic Growth in Daegu: Statistical Verification Using the Industrial Structure of Materials, Parts, and Equipment, Regression Models, and Random Forest

Min Joon Jin

KyungPook National Univ

### 요약

본 논문에서는 우리나라 16개 시도 소재, 부품, 장비 산업구조의 지역경제 다양성에 대해 분석하고 대구 지역경제 성장에 미치는 요인에 대해 기술한다. 또한 본 논문에서는 회귀모형과 랜덤효과 모형과 머신러닝 모델인 랜덤포레스트(Random Forest)를 활용하여 대구 지역경제 성장에 영향을 미치는 요인에 대해 통계적으로 검증하고 요인들을 추출하여 대구지역 경제 성장에 도움이 될 수 있는 정보들을 제공한다.

### I. 서론

소재, 부품, 장비 산업은 원자재에서 중간재, 최종재에 이르는 과정에서 중간재에 해당하며 모든 산업의 근간이 되는 산업이다. 구체적으로 소부장 산업은 고부가가치 산업이면서 기술집약적이며 중소기업 비중이 높은 산업으로(오지환, 정기호 2012)[1] 소부장 산업이 활성화가 되면 경제 전반의 양극화 및 일자리 창출 효과를 기대할 수 있다(이양호, 배준영 2016)[2]. 또한, 소재와 부품은 최종 완제품의 성능 및 품질 등을 결정하는 핵심요소로 완제품 생산 능력이 평균화 되는 시기에 제품의 차별화 및 경쟁력을 높일 수 있는 요인으로 그 중요성이 더욱 커지고 있다(이양호, 배준영 2016)[2].

지방도시에 소부장 산업이 활성화되면 지역경제활동의 파급효과를 창출한다. 특히 지역내 총 생산량이 증가함에 따라 지역경제가 활성화 되고 고용창출 및 지역내 인구유입 효과를 기대할 수 있다.

이에 본 논문에서는 지역별로 엔트로피극대화지수를 이용하여 소부장 산업의 다양성을 확인하고 엔트로피극대화지수를 보완하는 다른 경제지표로 임지계수를 이용하여 특정 산업이 지역 경제에서 얼마나 중요한지를 파악하는데 사용하였다. 또한 랜덤포레스트를 활용하여 grdp에 영향을 미치는 중요 변수들을 추출하였고 또 다른 방법으로 회귀모형과 랜덤효과 모형을 활용하여 grdp에 영향을 미치는 요인들을 통계적으로 검증하였다.

### II. 본론

본 논문에서는 먼저 소재, 부품, 장비 산업 데이터를 활용하여 지역별로 소부장 산업의 다양성을 확인하였다.

#### 2-1. 엔트로피극대화지수

$$E_{rt} = - \sum_{M=1}^M \left( \frac{e_{mrt}}{e_{rt}} \right) \ln \left( \frac{e_{mrt}}{e_{rt}} \right)$$

$E_{rt}$ 는 엔트로피극대화지수를 이용하여 t시점에서 측정된 r지역의 산

업구조 다양성지수이고,  $e_{mrt}$ 는 t시점에서 r지역의 m산업 고용자들

수,  $e_{rt}$ 는 t시점 r지역의 총고용자수, M는 해당 지역의 산업의 수이다.

결국에는 엔트로피극대화지수의 값이 클수록 지역의 소부장 산업구조의 다양성이 높다는 것을 의미하고, 엔트로피극대화지수의 값이 작을수록 소부장 산업중 특정 산업에 집중되어있다는 것을 의미한다.

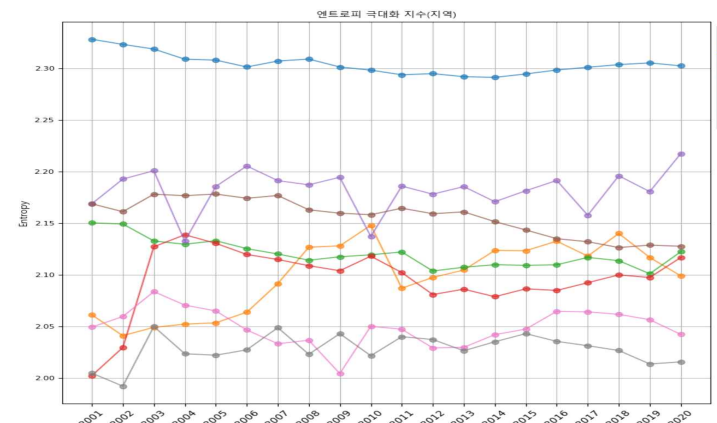


그림1 소부장 엔트로피극대화지수 특별시 및 6대광역시 지역

그림1의 경우 2000년대 초반에는 대구가 빠르게 소부장산업을 다양화하는 방향으로 나타났으며 서울 또한 소부장 산업을 다양화하는 방향으로 나아가고 있다. 반면에 인천과 부산은 해가 지날수록 소부장산업 구조가 빠르게 특화되어가고 있는 것을 볼 수 있다.

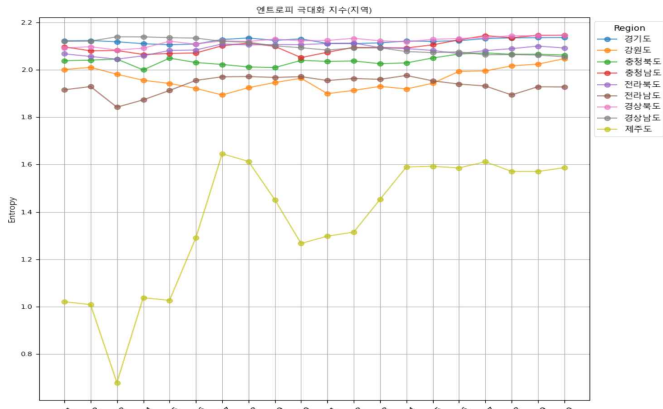


그림2 소부장 엔트로피극대화지수 도지역

한국의 9개 도 지역의 엔트로피극대화지수를 살펴보면 경상남도가 2000년 대 초반부터 중반까지 소부장 산업의 다양성 정도가 가장 크게 나타났으며 제주도는 엔트로피극대화지수가 가장 낮으며 애초에 소부장산업의 수가 가장 적었기 때문에 엔트로피극대화지수가 가장 낮은 수치라고 예상한다.

## 2-2. 입지계수

입지계수는 특정 지역의 산업구조를 다른 지역이나 전체 국가와 비교하여 해당 산업의 상대적 집중도를 측정하는 지표이다. 이는 특정 산업이 지역경제에서 얼마나 중요한지를 파악하는데 사용된다. 따라서 대구의 GRDP에 영향을 미치는 요인으로 소재, 부품, 장비 입지계수를 독립변수로 활용하였다.

$$LQ_{mrt} = \frac{(e_{mrt} / \sum_{M=1}^M e_{mrt})}{(e_{rt} / \sum_{M=1}^M e_{rt})}$$

여기서  $LQ_{mrt}$ 는 t시점 r지역의 m산업에 대한 입지계수,  $e_{mrt}$ 는 t시점 r지역 m산업의 고용자수,  $e_{rt}$ 는 t시점 국가 전체의 m산업 고용자수, M는 r지역 산업의 수를 의미한다.

## 2-3. 랜덤포레스트

대구 GRDP에 영향을 미치는 요인을 분석하기 위해 GRDP를 종속변수로 설정하고, 독립변수로 GDP, 인구수, 엔트로피 지수, 외환보유액, 소재·부품·장비 입지계수를 활용하였다.

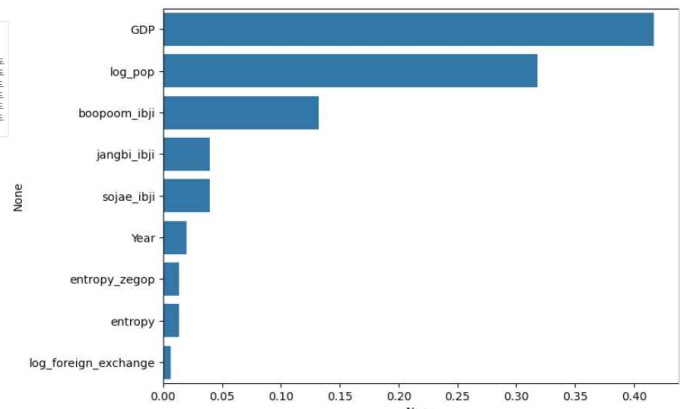


그림3 대구grdp에 영향을 미친 중요 변수

변수들의 중요도를 평가한 결과, GDP, 인구수, 부품 및 장비 입지계수가 대구 GRDP에 가장 큰 영향을 미치는 요인으로 나타났다.

## 2-3. 회귀모형(OLS)

변수	OLS모형
엔트로피지수	-3801.44
엔트로피지수(제곱)	898.12
GDP	1.10 ***
외환보유액	0.0002
소재산업입지계수	189.78 **
부품산업입지계수	421.24
장비산업입지계수	37.22
인구수	128.83
R-squared	0.861
Adj. R-squared	0.735

## III. 결론

본 논문에서는 소부장 데이터를 이용하여 대구 grdp에 대한 중요요인에 대해 분석하기 위해 통계분석과 랜덤포레스트 2가지 방법론을 소개하였다. 종속변수로 대구 GRDP 지수를 활용하였고 독립변수로 GDP, 인구수, 엔트로피 지수, 외환보유액, 소재·부품·장비 입지계수를 활용하였다. 통계모형과 랜덤포레스트에서 공통적으로 중요한 변수는 GDP였고, 랜덤포레스트는 GDP, 인구수, 부품입지계수가 중요한 변수였고, 통계모형은 GDP와 소재입지계수가 통계적으로 유의하다는 결과를 도출할 수 있었다.

## 참 고 문 헌

- [1] 오지환, 정기호 “DEA 모형을 이용한 부품소재산업의효율성 분석”, 대한경영학회
- [2] 이양호, 배준영 “한국 소재·부품산업의 경쟁력 분석 모델 제안”, 한국 무역연구원
- [3] 최기흥, 류수열, 윤성민 “지역산업구조 변화가 지역경제성장에 미치는 영향”