

# 다중공선성하에서 Heuristic 추정량과 편의회귀 추정량에 관한 Simulation 연구

홍승표\* · 전치혁\*

\* 포항공과대학교 산업공학과

## Abstract

독립변수들 사이의 높은 상관관계가 있는 공정 데이터의 경우에는 다중공선성(multicollinearity)의 문제로 일반적인 회귀분석 방법인 선형다중회귀모형(multiple linear regression : MLR)을 적용하면 추정회귀계수가 불안정하여 추정회귀식의 예측력(prediction accuracy)이 떨어지는 문제점이 있다.

본 연구의 목적은 다중공선성하에서 MLR의 최소자승법(ordinary least square : OLS)에 의한 추정회귀계수의 문제점을 개선하기 위해 편의회귀모형인 능형회귀(ridge regression : RR), 주성분회귀(principal component regression : PCR)을 적용하고 Stein Estimation 등에 대한 구조적 차이 및 특성을 파악하여 새로운 Heuristic 추정회귀계수를 제안하는데 있다. 이러한 목적을 위해 시뮬레이션(simulation) 수치실험을 통해 OLS, RR, PCR 그리고 Stein에 의해 추정된 회귀계수와 새로운 Heuristic 추정회귀계수에 대한 비교 평가를 하였다. 여러 방법론에 의한 추정회귀계수의 평가 척도(measure)로는 추정회귀계수의 평균제곱오차(MSE)를 사용한다.

OLS를 제외하고는 모두 일종의 편의회귀(biased regression) 모형으로, 약간의 편의로 OLS 추정회귀계수에 비해 분산이 크게 줄어든다는 것은 편의추정회귀계수의 MSE가 OLS 추정회귀계수의 MSE보다 더 작아진다는 것을 의미한다. 즉 OLS 추정회귀계수에 약간의 shrinkage 효과를 줌으로써 다중공선성으로 인한 OLS 추정회귀계수의 문제점을 보완할 수 있다. 그러나 이러한 shrinkage 효과를 어느 정도 주어야 하는지를 정하기 위해서는 반복적인(iterative) 계산과정이 필요하다.

본 논문에서 제안되는 Heuristic 추정량은 다중공선성이 존재할 때 OLS 추정회귀에서 간단히 계산할 수 있는 일정상수  $c$ 를 OLS 추정회귀계수에 곱해서 새로운 회귀계수를 추정하는 것이다. 이때 추정된 회귀계수의 MSE를 최소화하는 상수  $c$ 를 추정하게 되는데, 이는 앞의 여러 편의회귀 모형과 반복적인 계산과정 없이 간단하게 구할 수 있는 장점이 있다.