

経済学会第 568 回例会

2017 年 12 月 20 日

Efficient Regularization Parameter Selection for OLS Post-Selection Estimator

末石 直也

近年、高次元データと呼ばれるサンプルサイズよりも説明変数の数が大きいデータを用いた回帰分析が盛んに行われている。説明変数の数がサンプルサイズよりも大きいとき、そのままでは最小 2 乗法でパラメータを推定できないため、何らかの方法により変数選択を行い、説明変数の数を減らす必要がある。そのためのポピュラーな方法が、LASSO に代表される罰則付き推定法である。最小 2 乗推定量の目的関数にある種の罰則項をつけて最小化問題を解くことにより、変数選択とパラメータの推定を同時に行うことができる。

罰則付き推定量を用いる際には、罰則の大きさを決める正則化パラメータ (regularization parameter) の選択が重要であり、これは分析者が自分で決定する必要がある。正則化パラメータの値が大きすぎると必要な変数がモデルから除外される可能性があり、反対に、正則化パラメータの値が小さすぎると不要な変数がモデルに入れられる可能性がある。

本研究では、罰則付き推定量を用いた 2 段階推定法 (OLS post-selection 推定量) を採用し、そのためのデータドリブンな正則化パラメータの選択方法を提案する。OLS post-selection 推定量では、はじめに LASSO などの罰則付き推定量を用いて変数選択を行い、さらに選択された変数を用いて最小 2 乗法でパラメータを推定しなおす。このような推定量は Belloni and Chernozhukov (2013, Bernoulli) などでも考察されており、よい性質を持つことが知られている。本研究では、提案手法は漸近的に効率的であることを示す。すなわち、提案された方法で正則化パラメータを選択すれば、漸近的には L2 の意味での損失を最も小さくするような正則化パラメータを選ぶことが可能であることを示す。