

Macroeconomic Forecasting and Variable Selection with a Large Number of Mixed Frequency Data

日本学術振興会特別研究員 PD (統計数理研究所) 植松 良公
小樽商科大学 田中 晋矢

1 概要

近年のデータ環境の発展により、大規模経済データセットを用いたマクロ経済予測が注目を集めている。しかし、全ての変数をモデルに含めた場合、推定誤差の増大により予測精度が低下する可能性がある。こうした問題を解決する手段として、Lasso に代表される罰則付き推定法が知られている。この推定値はスパースになるため、節約的なパラメータ推定と同時に変数選択を行うことができる。しかし、高次元時系列モデルの場合の理論は必ずしも明らかではない。本研究では、その理論を提供するとともに、大規模月次データセットを用いた米国 GDP の MIDAS (MIxed DAta Sampling) 回帰による予測を通じてこの手法の有用性を示す。

2 理論分析

高次元線形時系列モデル $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}$ を考える。ただし、 \mathbf{X} は定常な p 個の説明変数から成る $T \times p$ 行列、 \mathbf{u} はマルチンゲール差分列から成る誤差項とする。ここで $p > T$ を許容するが、 \mathbf{X} は不要な (回帰係数が 0 である) 説明変数を多数含む、すなわち係数 $\boldsymbol{\beta}_0$ はスパースであるとする。このとき、推定量 $\hat{\boldsymbol{\beta}}$ を次の目的関数を \mathbb{R}^p 上で最小化するものとして定義すると、スパースな推定値が得られる：

$$Q_T(\boldsymbol{\beta}) := (2T)^{-1} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \|p_\lambda(\boldsymbol{\beta})\|_1$$

ただし $p_\lambda(\boldsymbol{\beta}) := (p_\lambda(|\beta_1|), \dots, p_\lambda(|\beta_p|))^T$ で p_λ は非負の罰則関数である。正則化項 $\lambda > 0$ はゼロに収束する様を選ぶ。罰則関数 p_λ は例えば ℓ_1 -penalty (Lasso) のほか、SCAD や MCP と呼ばれるものを含む (folded-concave penalties)。本研究では、当該モデルに関して適当な条件を課したうえで、(i) 予測誤差の非漸近的上限 (Oracle inequality)、並びに (ii) 変数選択の一致性及び $\hat{\boldsymbol{\beta}}$ の非ゼロ要素が「真の」モデルから得られる MLE と漸近同等であること (Oracle property) を示した。

3 実証分析

McCracken and Ng (2014) による米国大規模月次データセット FRED-MD を用いて、罰則付きスパース MIDAS 回帰 (S-MIDAS) による米国 GDP 予測を行った。その結果、超短期予測 (Nowcasting) において、S-MIDAS が大規模データセットを用いた MIDAS 予測で現在最も有力な手法といえる Factor MIDAS と同等以上のパフォーマンスを持つことが確認され、S-MIDAS による経済予測の有用性が示された。また、モデルの変数選択の結果として、予測に用いた月次 124 系列にラグを含めた計 1116 個の説明変数のうち、鉱工業生産指数 (IP)、個人消費支出 (PCE)、FF レートのラグなど計 31 個 (全体の約 3%) が GDP 予測に有用であることが明らかになった。