

Expectile Regression を用いた Heterogeneous モデルの分位点推定

東京大・情報理工 高須 佑哉
東京大・情報理工 清 智也

1 はじめに

分位点は確率変数を特徴付ける代表的な値の 1 つであり、回帰モデルにおける分位点の推定は重要な課題の 1 つである。代表的な手法としては分位点回帰が挙げられるが、Heterogeneous モデル、つまり残差の分布が説明変数に依存するような回帰モデルの下では、通常分位点回帰による推定量は漸近有効性を示さないことが知られている。また、漸近有効な推定量を構成するためには、残差の密度関数の推定が必要となる。

分位点回帰以外の手法としては、Waltrup et al. [1] によって提案された、expectile を用いた推定方法が挙げられる。expectile は平均値と分位点をあわせたような概念で、1 次元確率変数 Y の τ -expectile m_τ ($0 < \tau < 1$) は

$$m_\tau := \underset{m}{\operatorname{argmin}} \mathbf{E} [\omega_\tau(Y - m)^2]$$

として定義される値である。ただし ω_τ は

$$\omega_\tau = \begin{cases} 1 - \tau & (Y < m) \\ \tau & (Y \geq m) \end{cases}$$

であり、 Y と m に依存する。[1] では、expectile の回帰である expectile regression を繰り返し用いて複数の τ における expectile を推定し、これらの値を用いて確率分布関数を推定することで分位点の推定値を求める、という手法を提案している。

2 提案手法

本研究では、[1] の手法を改良することにより、Heterogeneous モデルにおける新しい分位点の推定手法を提案する。分位点回帰の場合と同様に、Heterogeneous モデルにおいて通常 expectile regression は漸近有効性を示さないことが知られている。漸近有効な推定量を構成するためには、説明変数 X と従属変数 Y の実現値の組 (X_i, Y_i) ($i = 1, \dots, n$) に対して

$$\mathbf{E} [\omega_\tau(Y_i - m_\tau(X_i))^2 | X = X_i], \quad i = 1, \dots, n \quad (1)$$

の推定を行う必要があるが、密度関数を推定する必要はない。

提案手法では、まず通常 expectile regression を用いて [1] の手法を行い、その過程で得られた分布関数の推定量を用いて (1) を推定した上で、再度 [1] の手法を行うことを考える。発表では提案手法の詳細に加えて、数値実験を通じた提案手法の推定誤差の検証結果を紹介する。

参考文献

- [1] Waltrup, L. N., Sobotka, F., Thomas, K. and Kauermann, G. (2015): Expectile and quantile regression – David and Goliath? *Statistical Modeling*, **15**(5), 433–456.