

編集委員会依頼論文

健康保険データに基づく医療費予測モデリング
—正則化 two-part モデルによるアプローチ—

小暮 厚之* 小林凌雅†

2018年8月31日投稿

概要

本論文では、健康保険データに基づく医療費の予測モデリングについて考察する。特に、医療費の代表的な統計モデルである two-part モデルを取り上げ、ラッソのような正則化手法の適用により予測精度が高まるか否かを検討する、わが国の健康保険データへ応用した結果、説明変数の個数が大きい高次元の場合には、正則化により予測精度が向上する可能性が示された。

キーワード：医療費の予測，健康保険データ，正則化，two-part モデル，スパース性，ラッソ

1 はじめに

本論文では、健康保険データに基づく医療費の予測モデリングについて考察する。特に、医療費の代表的な統計モデルである two-part モデルを取り上げ、ラッソのような正則化手法の適用により予測精度が高まるか否かを検討する、ここで用いるデータは、わが国の複数の健康保険の組合員の集団から無作為に抽出された 10,000 人の 2010 年から 2012 年までの 3 年間のレセプト及び健康診断のデータである。まず、2010 年の変数を説明変数、2011 年の総医療費を目的変数として予測モデルを構築した。次に、この予測モデルの評価を行うために、構築した予測モデルの説明変数に 2011 年の値を代入して、2012 年の医療費を予測した。この結果、説明変数の個数が大きい高次元の場合には、two-part モデルを正則化することにより、予測精度が向上する可能性が示された。

正則化による医療費の予測は、Loginov et. al (2013) によって既に報告されている。ただし、Loginov et. al (2013) では、通常の回帰モデルに正則化を適用している。本稿では、two-part モデルに対する正則化の適用について議論し、ラッソ法 (Tibshirani, R., 1996) と弾性ネット法 (Zhou and Hastie, 2005) の 2 種類のスパース正則化の適用を考察した。また、スパース性をもたない伝統的な正則化手法であるリッジ法も試みた。

本論文の構成は以下の通りである。第 2 節では、医療費の代表的な統計モデルであるトービット法、標本選択法、two-part モデルについて概説する。第 3 節では two-part モデルの推定と予測について詳述する。第 4 節では、予測モデリングにおけるバイアス-分散トレードオフについて注意を与える。第 5 節ではラッソ法を中心として two-part モデルへの正則化の適用を議論し、第 6 節でわが国健康保険データへの応用を考察する。第 7 節で結語を述べる。

* 東京経済大学 経営学部 e-mail: kogure@tku.ac.jp

† 慶應義塾大学大学院 政策・メディア研究科博士課程

‡ 本論文は、株式会社 JMDC との共同研究を通じて作成されたものであり、データの提供を含め同社から多大な支援を受けた。ここに深く感謝申し上げます。