

Robust Estimation with Discrete Explanatory Variables

Pavel Čížek*

Abstract

The least squares estimator is probably the most frequently used estimation method in regression analysis. Unfortunately, it is also quite sensitive to data contamination and model misspecification. Although there are several robust estimators designed for parametric regression models that can be used in place of least squares, these robust estimators cannot be easily applied to models containing binary and categorical explanatory variables. Therefore, I design a robust estimator that can be used for any linear regression model no matter what kind of explanatory variables the model contains. Additionally, I propose an adaptive procedure that maximizes the efficiency of the proposed estimator for a given data set while preserving its robustness.

JEL: C20, C24, C61

Keywords: discrete explanatory variables, linear regression, robust statistics, least trimmed squares

* Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin and Charles University, Center for Economic Research and Graduate Education, Prague

Abstrakt

Metoda nejmenších čtverců je jednou z nejčastěji používaných metod v regresní analýze. Její hlavní nevýhodou je velká citlivost vůči chybné specifikaci modelu a kontaminaci dat. Toto řeší tak zvané robustní odhady, kterých existuje pro lineární regresní model celá řada. Většinu robustních odhadů však není možno použít v případě, že model obsahuje diskretní proměnné. V tomto článku proto navrhuji robustní odhad, který je možno aplikovat v libovolném lineárním regresním modelu bez ohledu na charakter vysvětlujících proměnných. Nezanedbatelnou částí je i návrh a studium adaptivní metody, která maximalizuje účinnost odhadu při zachování jeho robustnosti.