Statistica Sinica **32** (2022), 1911-1938
doi:https://doi.org/10.5705/ss.202021.0068

A GENERALIZED HECKMAN MODEL WITH VARYING SAMPLE SELECTION BIAS AND DISPERSION PARAMETERS

Fernando de Souza Bastos[1,2], Wagner Barreto-Souza[2,3] and Marc G. Genton[3]

[1] *Universidade Federal de Viçosa,* [2] *Universidade Federal de Minas Gerais and* [3] *King Abdullah University of Science and Technology*

Abstract: Many proposals have emerged as alternatives to the Heckman selection model, mainly to address the non-robustness of its normal assumption, which is often illustrated using the 2001 Medical Expenditure Panel Survey data. In this paper, we propose a generalization of the Heckman sample selection model by allowing the sample selection bias and dispersion parameters to depend on covariates. We show that the non-robustness of the Heckman model may be due to the assumption of the constant sample selection bias parameter, rather than the normality assumption. Our proposed methodology allows us to understand which covariates explain the sample selection bias phenomenon, rather than to simply form conclusions about its presence. Furthermore, our approach may attenuate the non-identifiability and multicollinearity problems faced by existing sample selection models. We explore the inferential aspects of the maximum likelihood estimators (MLEs) for our proposed generalized Heckman model. More specifically, we show that this model satisfies some regularity conditions such that it ensures consistency and asymptotic normality of the MLEs. Proper score residuals for sample selection models are provided, and model adequacy is addressed. Simulated results are presented to check the finite-sample behavior of the estimators, and to verify the consequences of not considering a varying sample selection bias and dispersion parameters. We show that the normal assumption for analyzing medical expenditure data is suitable, and that the conclusions drawn using our approach are coherent with the findings from prior studies.

Key words and phrases: Asymptotics, heteroscedasticity, regularity conditions, score residuals, varying sample selection bias.

1. Introduction

Heckman (1974, 1976) introduced a model for dealing with the sample selection bias problem, using a bivariate normal distribution to relate the outcome of interest and a selection rule. A semiparametric alternative to this model, known

Corresponding author: Wagner Barreto-Souza, Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. E-mail: wagner.barretosouza@kaust.edu.sa.