University of Central Florida
College of Business

QMB 6912
Capstone Project in Business Analytics

Solutions: Problem Set #2

The author "James J. Heckman" in this paper introduced the issues of bias caused by the using nonrandomly selected samples to estimate behavioral relationships as an ordinary specification error or omitted variables bias.

In his paper he discussed sample selection bias as a specification error and presented a simple consistent estimation method that eliminates the specification error for the case of censor samples.

He underlined the cause of sample selection bias in two reasons. The first is self-selection by individuals or data units being investigated. The second, sample selection decisions by analysts or data processors operate in much the same way as self-selection.

Then he introduced examples of self-selection bias. The first one was the observations in market wages for working women whose market wage exceeds their home wage at zero hours of work, second one was the wages for union members who found their nonunion alternative less desirable, the third one was wages of migrants that do not afford a reliable estimate of what nonemigrants would have earned had they migrated.

According to the author, in each of the examples, wage or earning functions estimated on selected samples do not estimate the population wage functions. Which caused by biased estimate of the effect of a random "treatment" the sample.

He suggested that data can also be nonrandomly selected because of

decisions taken by data analysts. And that has the same effect on structural estimates as self-selection.

In conclusion, this paper introduced a computationally tractable technique that enables analysts to use simple regression techniques to estimate behavioral functions free of selection bias with censored sample.