

A comparison of parametric and semi-parametric estimators of farm profit-accounting for sample design in household surveys

Dean Jolliffe

Economic Research Service, USDA, Washington, DC, USA
and the Institute for the Study of Labor (IZA), Bonn, Germany

Abstract

Much empirical research assumes that data are selected from a sample frame using a simple random draw, when in practice farm-household survey data most typically results from complex, multi-stage sample designs. This paper shows that understanding sample design is important to determining an appropriate estimation strategy, and argues that the least absolute deviations (LAD) and censored least absolute deviations (CLAD) regression are useful tools when using survey data which comes from a multi-stage draw. LAD and CLAD estimators are used to predict the effect of education on farm and non-farm income and labor supply. CLAD standard errors, which are robust to violations of homoscedasticity and independence, are generated by a boot-strap method which replicates the two-stage sample design.

Keywords: farm households, off-farm labor, human capital, least absolute deviations, censored least absolute deviations, sample design, Ghana

JEL: C14, C42, I2

1. Introduction

Much empirical research assumes that the data being analyzed are taken from a sample frame using a simple random draw, when in practice household survey data most typically results from complex, multi-stage sample designs. Consider a nationally representative household survey which uses a national census as the sample frame. Rarely is such a survey designed to randomly draw sample households in one stage from the census data. One example of a typical design is to first stratify on geographic regions or socio-economic characteristic, then draw primary sampling units (which sometimes will be villages or city blocks), and finally to randomly draw the ultimate sampling units such as households. This is only one example of many types of designs,