International Statistical Review (2006), **74,** 3, 403–418, Printed in Wales by Cambrian Printers

Estimation in Two-Stage Models with Heteroscedasticity

John Buonaccorsi

University of Massachusetts, Amherst, MA, USA. E-mail: johnpb@math.umass.edu

Summary

A surprising number of important problems can be cast in the framework of estimating a mean and variance using data arising from a two-stage structure. The first stage is a random sampling of "units" with some quantity of interest associated with the unit. The second stage produces an estimate of that quantity and usually, but not always, an estimated standard error, which may change considerably across units. Heteroscedasticity in the estimates over different units can arise for a number of reasons, including variation associated with the unit and changing sampling effort over units. This paper presents a broad discussion of the problem of making inferences for the population mean and variance associated with the unobserved true values at the first stage of sampling. A careful discussion of the causes of heteroscedasticity is given, followed by an examination of ways in which inferences can be carried out in a manner that is robust to the nature of the within unit heteroscedasticity. Among the conclusions are that under any type of heteroscedasticity, an unbiased estimate of the mean and the variance of the estimated mean can be obtained by using the estimates as if they were true unobserved values from the first stage. The issue of using the mean versus a weighted average which tries to account for the heteroscedasticity is also discussed. An unbiased estimate of the population variance is given and the variance of this estimate and its covariance with the estimated mean is provided under various types of heteroscedasticity. The two-stage setting arises in many contexts including the one-way random effects models with replication, meta-analysis, multi-stage sampling from finite populations and random coefficients models. We will motivate and illustrate the problem with data arising from these various contexts with the goal of providing a unified framework for addressing such problems.

Key words: Mean, meta-analysis; Measurement error, Random coefficients; Random effects; Replication; Sampling; Variance.