Statistica Sinica **35** (2025), 1-21
doi:https://doi.org/10.5705/ss.202022.0331

A LINEAR ERRORS-IN-VARIABLES MODEL WITH UNKNOWN HETEROSCEDASTIC MEASUREMENT ERRORS

Linh H. Nghiem*[1] and Cornelis J. Potgieter[2,3]

[1] *University of Sydney,* [2] *Texas Christian University and* [3] *University of Johannesburg*

Abstract: In the classic measurement error framework, covariates are contaminated by independent additive noise. This paper considers parameter estimation in such a linear errors-in-variables model where the unknown measurement error distribution is heteroscedastic across observations. We propose a new generalized method of moment (GMM) estimator that combines a moment correction approach and a phase function-based approach. The former requires distributions to have four finite moments, while the latter relies on covariates having asymmetric distributions. The new estimator is shown to be consistent and asymptotically normal under appropriate regularity conditions. The asymptotic covariance of the estimator is derived, and the estimated standard error is computed using a fast bootstrap procedure. The GMM estimator is demonstrated to have strong finite sample performance in numerical studies, especially when the measurement errors follow non-Gaussian distributions.

Key words and phrases: Asymmetric distributions, bootstrap, generalized method of moments, nutrition, phase function, variance heterogeneity.

1. Introduction

The errors-in-variables linear model arises when certain covariates suffer from measurement error contamination. This can stem from sources like instrumentation and self-reporting errors, as well as the inadequate use of short-term measurements as proxies for long-term variables. Ignoring measurement error can result in biased estimators, see Carroll et al. (2006) regarding the importance of measurement error correction in understanding the effects of the covariates on the outcome. This paper considers a heteroscedastic measurement error setting, allowing the measurement error covariance to vary across observations. This observation-specific measurement error variance structure, treated as unknown, requires estimation from replicate data. We adopt the classic additive measurement error model wherein the contaminated covariates, i.e. the surrogates, are treated as the sum of the true covariates and independent measurement errors, so surrogate variances exceed true covariate variances.

*Corresponding author. E-mail: linh.nghiem@sydney.edu.au