Parameter Estimation Method of Latent Variable Models: SIMEX Approach

Sahika Gokmen[1,2] **and Johan Lyhagen**[2]

[1] *Ankara Haci Bayram Veli University, Ankara, Turkey*
[2] *Uppsala University, Uppsala, Sweden*

Abstract

It is widely accepted that economic variables cannot be measured, directly observed or only measured with an indirect variable. In any of case, errors in the measurement process are inevitable too. These situations cause problems from the very beginning (compilation of data) for a valid analyzes of economic data. If only indirect measurement is possible for any non-measurable variables such as the productivity of a worker, permanent income, etc, these variables are described as "latent"[1]. Thereby, according to Wansbeek and Meijer (2000), latent variable consists of the observed variable and its difference from its latent counterpart. On the other side, the measurement error definition is more general and refers to the difference between the observed variable and its real value. With regard to measurement error, the observed variable in place of the true variable is named errors-in-variable or error-prone variable.

In the literature, latent variables are incorporated in the analysis via Structural Equation Models (SEM). However, since the error-prone variables cause biased parameter estimations in regression models, ignoring measurement error can lead to serious problems. In this context the statistical models used to analyze mismeasured data are called measurement error models or error-in-variables model. Among the measurement error models, Simulation-Extrapolation (SIMEX) method is considered as one of the most successful one in the lately related field. SIMEX method is a unbiased parameter estimation technique based on simulation improved by Cook and Stefanski (1994)[2].

Among the current studies in the literature, the differences between these concepts are frequently mentioned and the importance of this emphasized. Especially Hu (2017) noted in his study that the relationship between latent variables and their measures are important in defining the observed and unobserved relationships in economic models[3]. Although not through SEM, Küchenhoff et al. (2006) used the SIMEX method for binary variables with a focus on classification success[4]. In addition, even though Simonetto (2008) considers the processing of estimation methods in measurement error models on SEM, there is no methodological improvement was made such as updating SIMEX[5]. Rutkowski and Zhou (2015), on the other hand, deal with latent variables and measurement error separately, but their research only contains an application on the education system through the study of

Küchenhoff et al. (2006) [6]. Overall, Hu (2017)'s study draws attention as the first social science study that takes into account business economics in this field. In his study, latent variables considered as mismeasured and then some measurement error estimation methods are applied[3]. However, a concrete adaptation of SIMEX for the estimation of SEM has not been proposed yet. Thereby, in this study, the SIMEX method, which is very effective and popular estimation method in measurement error models, is adapted to a new field. It is aimed to obtain unbiased estimations for regression models with a latent explanatory variable in which it is taken into consideration that latent variable also is an error-prone variable. Combining measurement errors and latent variables together plays an important role in terms of providing more reliable results for both survey studies and economic analyzes.

The main purpose of this study is that the theoretical adaptation of SIMEX method in terms of SEM will be discussed. Also, the adaptation will be supported by simulation studies under different conditions. In this way, it is expected that the study will serve as a bridge between SEM and measurement error models and it will contribute to the estimation process of SEM.

Keywords

Structural equation models, error-prone variables, SIMEX.

Acknowledgements

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the TUBITAK BIDEB 2219-International Postdoctoral Research Scholarship Programme.

References

[1] Wansbeek, T. and Meijer, E. (2000). *Measurement Error and Latent Variables in Econometrics*. North Holland, Elsevier.

[2] Cook, J., R. and Stefanski, L., A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association 89*, 1314–1328.

[3] Hu, Y. (2017). The Econometrics of Unobservables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics. *Journal of Econometrics 200*, 154-168.

[4] Küchenhoff, H., Mwalili, S. M. and Lesaffre, E. (2006). A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX. *Biometrics 62(1)*, 85-96.

[5] Simonetto, A. (2008). *Estimation Procedures for Latent Variable Models with Psychological Traits*. PhD Thesis, Universita Degli Studi Di Milano – Bicocca, Milano.

[6] Rutkowski, L. and Zhou, Y. (2015). Correcting Measurement Error in Latent Regression Covariates via the MC-SIMEX Method. *Journal of Educational Measurement 52(4)*, 359-375.