Issues in the Structural Equation Modeling of Complex Survey Data

Kenneth A. Bollen[1,4], Stephen J. Tueller[2], Daniel Oberski[3]
[1]University of North Carolina-Chapel Hill, Chapel Hill, NC, USA
[2]RTI International, Research Triangle Park, NC, USA
[3]Tilburg University, Tilburg, Netherlands
[4]Corresponding author: Kenneth A. Bollen, e-mail: bollen@unc.edu

Abstract

Structural equation models are routinely applied to data collected from complex samples. Methods to take account of the sample design and weights are discussed in a growing literature on the topic. This paper provides a literature review of the analysis of complex samples using structural equation models. The goal of the paper is to extract the points of agreement and disagreement from these publications; to assess and evaluate those areas of consensus; and to propose the research areas most in need of further research.

Key Words: Complex samples, latent variables, sample weights

1. Introduction

Structural Equation Models (SEMs) are widely used in the social and behavioral sciences as well as in marketing and information sciences, and they are growing in importance in health sciences and biostatistics. The ability of SEMs to handle latent variables, measurement error, and multiple indicators in systems of equations has contributed to their popularity. Tests of model fit have provided tools to assess the correspondence between models and data and to compare different models. Despite the vast number of SEM users, the discussions of taking account of weighting and complex samples have been limited. This is surprising in that many SEMs are applied to survey data with multistage sampling and unequal selection probabilities of individuals.

Part of this inattention is explainable by the history of the SEMs. Though we can trace these models back to Sewall Wright and path analysis in the first half of the 20th century, the contemporary forms of SEMs originating in the 1960s to 1980s grew out of factor analysis from psychology, simultaneous equation models from economics, and early syntheses of these models in the sociology literature. Although the sample survey tradition and attention to complex samples is strong among some social scientists, those working on SEMs typically were not survey methodologists. Most statistical estimation with SEMs originated with the implicit assumption of simple random sampling.

Fortunately, the last 20 or so years have seen this situation change. The literature on complex samples in latent variable SEMs is small but growing. Latent variable software (e.g., Mplus, LISREL, Stata, R package Lavaan) incorporate complex survey features such as sampling weights, clustering, and stratification, making it feasible to fit SEM models not covered in other software. The time is right to make an assessment of the major advice emerging from these publications.

The primary purposes of our paper are: (1) to review the major themes in the complex samples and SEMs literature, (2) to determine the points of consensus, and (3) to evaluate what questions remain to be answered. Our goal is not to provide a complete review of complex sample analysis in general nor is it to provide a comprehensive review of SEMs. Rather our focus is on the intersection of these topics. Our paper is organized as follows. The next section describes the scope of the literature reviewed. Then we discuss those areas of consensus on the properties and approaches to complex samples in SEMs. Following this is a section on areas of controversy. The Conclusions form the last section.

2. Literature Review Our target publications focus on complex sample analysis using SEMs. We searched for published articles or chapters in edited books. Our search was restricted to methodological articles that discussed this issue and we did not include substantive articles that applied complex samples corrections with SEMs.

In their most general sense, SEMs would include virtually all of the major statistical models in common use. To avoid casting such a large and less useful net, we focus on common forms of latent variable SEMs such as confirmatory factor analysis and latent variable models with continuous latent variables. The latter condition means that we did not consider latent class models. Articles were excluded if they were primarily concerned with other latent variable methods (e.g., item response theory), or with multilevel SEM in the absence of weighting or stratification issues (i.e., MLSEM with clustering only). We also excluded articles that were primarily applications. Articles were located using Google Scholar (terms “complex survey” and “structural equation model”). Additional articles were located searching the references and works citing the initial set of articles. The reference section provides the complete list of papers found in the literature review.

3. Points of Consensus

Muthén & Satorra (1995) discussed two modes of inference in SEM for complex surveys: aggregated and disaggregated analysis (see also Skinner, Holt & Smith, 1989, pp. 8-10). Disaggregated analyses define the parameters of interest as being within-cluster, i.e. the "conditional" parameters. Examples are multiple group and multilevel analyses. In contrast, aggregated analyses define the parameter of interest as the "marginal" parameter over all clusters: the target parameter is the "estimate" that would be obtained if the model were fitted to population data (Skinner, Holt & Smith, 1989, p. 81). Which of these two types of parameters is of interest will depend on the research question. In some cases, the two types of target parameters will correspond; in general, however, they do not (see Muthén,1989). Aggregated parameters are not necessarily estimates of disaggregated parameters (see Monte Carlo simulation by Wu & Kwok, 2012).

Psuedo-maximum likelihood (PML) for linearization estimation of asymptotic covariance matrices is frequently advocated for estimating SEM models with complex survey data (Muthén & Satorra 1995). Skinner, Holt, & Smith (1989, pp. 79-84) developed PML for complex samples for aggregated modeling, though the ideas may also be applied to obtain multilevel PML estimation. PML consists of two parts: (1) replacing sample covariances by weighted sample covariances, and (2) replacing inverse Fisher information with a sandwich estimator of variance. Studies show that not replacing the estimates by weighted estimates leads to bias, and not replacing Fisher information variance estimator with sandwich estimator leads to wrong standard errors (Stapleton, 2006; Asparouhov & Muthén, 2006; Asparouhov, 2005). Initial development of resampling methods including the jackknife repeated replication, balanced repeated replication, and bootstrapping are alternative approaches for standard error estimation (Stapleton, 2008) though their relative performance is largely unknown.

Although initial research under simple sample schemes and models suggested that sampling weights had negligible effect on parameter estimates (Kaplan & Ferguson, 1999; Hahs-Vaughn & Lomax, 2006), more recent research has demonstrated that ignoring sampling weights can lead to parameter bias when the probability of selection is informative on the SEM parameters (Asparouhov 2005; Stapleton, 2006). On the other hand, when unequal selection probabilities do not produce bias, analyzing with weights can inflate standard errors. Whether to include weights therefore depends on whether sampling is informative about the parameters. This can be tested by classical tests for informativeness (Pfefferman 1993; Fuller 2009, chapter 6); Asparouhov & Muthén (2007) provided an adjustment of Pfefferman's test that performs better in small samples. They also suggest that another application of tests for informativeness is to compare estimates obtained with original and trimmed weights If the test indicates no difference, the trimmed weights may provide equally consistent estimates but with lower variance.

In multilevel SEM, it becomes relevant how the first-level weights are scaled. A range of possible scaling methods has been suggested, such as scaling to the within-cluster sample size or effective sample size (Pfefferman *et al.*, 1998; Stapleton 2002; Grilli & Pratesi 2004). Asparouhov (2006) suggested an extension of Grilli & Pratesi's (2004) estimation method and performed a Monte Carlo evaluation of six candidate methods of scaling within-cluster weights; he suggested to scale weights to the within-cluster sample size.

Model-based estimation performs better than design based estimation (Wu & Kwok, 2012) when the model is correct (Skinner, Holt & Smith 1989). An advantage of design-based estimation may be that it does not rely on model correctness, while a disadvantage can be loss of efficiency and small-sample bias. Skinner & de Toledo Vieira (2007) show that misspecified model-based estimation – i.e. using random intercepts to deal with clustering – can lead to serious underestimation of standard errors. Wu & Kwok (2012) argued that model-based inference is more appropriate for disaggregated parameters. Stapleton (2002) initially suggested replacing sample size with “effective sample size” to correct standard errors, but later found evidence against this (Stapleton, 2006).

4. Points of Controversy

Correcting test and fit statistics for complex sampling has been studied less than point and variance estimation. The Satorra-Bentler (1994) first-order correction to the overall chi-square test is usually applied (du Toit, 2012; Oberski frth; Muthén & Satorra, 1995). A second-order correction is another possibility (Muthén & Satorra, 1995) but no study has systematically investigated which is better. Modification indices and expected parameter changes should be adjusted as well. One route, for instance taken by the lavaan.survey software, applies the overall Satorra-Bentler correction to each modification index, but this approximation is not evaluated anywhere. Wu and Kwok (2012) conclude that fit indices (RMSEA, CFI, SRMR, chi-square) could not reliably detect misspecified higher order models.

There is also little evaluation of estimators that are used as alternatives to PML outside the SEM literature. In particular, weighted least squares or the generalized estimating equation (GEE) framework are alternatives that may perform better than PML. Asparouhov & Muthén (2005) compared PML with GEE estimation, and de Toledo Vieira & Skinner (2006) compared PML with Asymptotically Distribution-Free (ADF) estimation. Oberski (frth) suggested to apply Yuan & Bentler (1998) "Gamma" matrix as estimation weights. But it remains to be seen how well this works in general. Thus there are still many possible choices of estimation weight matrix and smoothed variance estimators that could still be applied to complex SEM.

Casewise Maximum Likelihood ("FIML") is subsumed in Asparouhov's (2005) framework, allowing for missing data. Multiple imputation is another popular way of dealing with missing data, but when sampling weights are involved this method may be more problematic (Kott 1995; Kim, Brick, Fuller, and Kalton 2006).

Design: surveys are often designed with univariate statistics in mind. For example the European Social Survey requires the effective sample size to be at least 1500, which is evaluated by calculating design effects for means of some key variables in the survey (see ESS website). But design effects for SEM parameters may be very different from those for means (Skinner 1986; Skinner, Holt & Smith 1989). Therefore design choices currently made for surveys may be inadequate for the purpose of SEM.

Models in the reported literature are typically simple factor models with three or fewer factors and typically only covariances (rather than regression relationships) are specified between the factors (see Hahs-Vaughn & Lomax, 2006, for an exception). Mediation, multiple group, longitudinal, and other generalizations of SEM have not been examined.

5. Conclusions

Much of the data subject to analysis by SEMs comes from surveys collected using complex samples. Yet the majority of SEMs analyses ignore the sample design and report results that implicitly assume simple random sampling. The consequences of this practice depend on the degree of departure from simple random sampling, whether sample design variables are part of the model, the degree of model misspecification, and whether the distributional assumptions are correct. The impact can range from nearly correct estimates and significance tests to severely biased parameter estimates or significance tests. One unanswered question is when the results assuming simple random sampling will be robust to complex sampling. The literature also suggests design-based and model-based corrections to analyzing complex samples. Which approach works best depends on a number of factors such as whether the cluster level model has a similar structure to the individual level one, but certainly more research is needed to address the relative performances of these approaches. In addition, design-based corrections can increase the variance of the estimator and failing to correct can affect the accuracy of estimates and their significance tests. It would be useful to compare statistics on mean square errors to develop guidelines for the use of design-based vs. uncorrected estimates.

In sum, complex samples in latent variable SEMs applications are common. Although the SEMs are more general than other statistical models discussed in the survey sampling literature, there is a striking similarity in the questions and problems that face SEMs and those faced with more widely studied techniques such as regression models.

References

Asparouhov, T. (2005). Sampling weights in latent variable modeling. Structural equation modeling, 12(3), 411-434.

Asparouhov, T., & Muthen, B. (2005). Multivariate statistical modeling with survey data. In Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference. http://www. fcsm. gov/events/papers05. html.

Asparouhov, T., & Muthen, B. (2007). Testing for informative weights and weights trimming in multivariate modeling with survey data. In Proceedings of the 2007 JSM meeting, Section on Survey Research Methods, Salt Lake City, Utah.

Asparouhov, T., & Muthen, B. (2006). Comparison of estimation methods for complex survey data analysis. Mplus Web Notes.

Asparouhov, T., & Muthen, B. (2006). Multilevel modeling of complex survey data. Proceedings of the American Statistical Association, Seattle, WA: American Statistical Association.

du Toit, S. (2012). Analysis of structural equation models based on a mixture of continuous and ordinal random variables in the case of complex survey data. in: Edwards, M. and MacCallum, R.C. (eds.). Current Topics in the Theory and Application of Latent Variable Modeling. New York, NY: Routledge Academic.

Fuller W. (2009). Sampling Statistics. John Wiley & Sons, New York.

Hahs-Vaughn, D. L., & Lomax, R. G. (2006). Utilization of sample weights in single-level structural equation modeling. The Journal of experimental education, 74(2), 161-190.

Kaplan, D., & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. Structural Equation Modeling: A Multidisciplinary Journal, 6(4), 305-321.

Kim J, Brick J, Fuller W, Kalton G (2006). "On the Bias of the Multiple-Imputation Variance Estimator in Survey Sampling." Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(3), 509–521.

Kott P (1995). "A Paradox of Multiple Imputation." In: Proceedings of the Section on Survey Research Methods, pp. 384–389. American Statistical Association.

Muthén, B. (1989). Latent Variable Modeling In Heterogeneous Populations. Psychometrika, 54(4), 557-585

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review/Revue Internationale de Statistique*, 317-337.

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, *60*(1), 23-40.

Satorra A. & Bentler P. (1994). "Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis." In A von Eye, CC Clogg (eds.), Latent Variables analysis: applications to developmental research. Sage Publications, Inc.

Satorra, A., & Muthén, B. (1995). Complex sample data in structural equation modeling. Sociological methodology, 25, 267-316.

Satterthwaite F (1941). "Synthesis of Variance." Psychometrika, 6(5), 309–316.

Oberski, D. (forthcoming). lavaan.survey: An R Package for Complex Survey Analysis of Structural Equation Models. Journal of Statistical Software.

Skinner, C. J., Holt, D., & Smith, T. F. (1989). Analysis of complex surveys. John Wiley & Sons.

Skinner, C. J. (1986). Regression estimation and post-stratification in factor analysis. Psychometrika, 51(3), 347-356.

Skinner, C. J. (1986). Design effects of two-stage sampling. Journal of the Royal Statistical Society. Series B (Methodological), 89-99.

Stapleton, L. M. (2008). Variance estimation using replication methods in structural equation modeling with complex sample data. Structural Equation Modeling, 15(2), 183-210.

Stapleton, L. M. (2006). An assessment of practical solutions for structural equation modeling with complex sample data. Structural Equation Modeling, 13(1), 28-58.

Stapleton, L. M. (2006). Using multilevel structural equation modeling techniques with complex sample data. Structural equation modeling: A second course, 345-383.

Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. Structural Equation Modeling, 9(4), 475-502.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. Multivariate behavioral research, 44(6), 711-740.

Yuan, K., and Bentler, P.M.. "Normal theory based test statistics in structural equation modelling." *British Journal of Mathematical and Statistical Psychology* 51.2 (1998): 289-309.

Vieira, M. D.T. and Skinner, C. J. (2006) *Estimating models for panel survey data under complex sampling.* Southampton, UK, University of Southampton, Southampton Statistical Sciences Research Institute (S3RI Methodology Working Papers, M06/17). Retrieved from http://eprints.soton.ac.uk/42001/.

Wu, J. Y., & Kwok, O. M. (2012). Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches. Structural Equation Modeling: A Multidisciplinary Journal, 19(1), 16-35.