

Key Concepts : ML, AGLS, Robust, Bootstrap method

Diagnostic Study of Problems under Asymptotically Generalized Least Squares Estimation of Physical Health Model*

Jung Hee Kim¹

ABSTRACT

This study examined those problems noticed under the Asymptotically Generalized Least Squares estimator in evaluating a structural model of physical health. The problems were highly correlated parameter estimates and high standard errors of some parameter estimates. Separate analyses of the endogenous part of the model and of the metric of a latent factor revealed a highly skewed and kurtotic measurement indicator as the focal point of the manifested problems. Since the sample sizes are far below that needed to produce adequate AGLS estimates in the given modeling conditions, the adequacy of the Maximum Likelihood estimator is further examined with the robust statistics and the bootstrap method.

These methods demonstrated that the ML methods were unbiased and statistical decisions based upon the ML standard errors remained almost the same. Suggestions are made for future studies adopting structural equation modeling technique in terms of selecting of a reference indicator and adopting those statistics corrected for nonnormality.

I. INTRODUCTION

The maximum likelihood (ML) method has been predominantly used in estimating structural equation models. The ML method chooses best estimates maximizing the likelihood that the discrepancy between estimates and observed covariances could have arisen as a mere sampling fluctuation (Hayduk, 1987). The ML estimators hold desirable asymptotic properties in large samples. In particular, they

are unbiased, consistent, and efficient by having a smaller sampling variance than any other estimation (Bollen, 1989; Long, 1983). Another advantage is their approximately normal distribution, and this nature leads to a significance test of each parameter estimate based on the z-distribution (Bollen, 1989).

The ML method assumes multivariate normal distribution of the observed variables. The multivariate normal distribution further assumes that each variable has zero skewness and zero

* This is a part of dissertation at the University of Michigan.

¹ RN, PhD, Assistant Professor, Department of Nursing, Inha University, Incheon, Korea

kurtosis (Anderson and Gerbing, 1988). The effects of violating this multivariate normal assumption are still not very well understood. Cuttance (1987), summarizing the studies on this issue, states that ML parameter estimates are "relatively robust against skewness for a wide range of applications in the social and behavioral sciences" (p. 268), where robustness refers to the degree of producing valid estimates and inferences by the model under violations of the methodological and statistical assumptions.

When the data set includes severely skewed or highly kurtotic variables, Cuttance (1987) recommends the asymptotically distribution-free (ADF) method. In the same line, Bollen (1989) suggests estimating the model with both the ML and the ADF methods and comparing the results when a significant kurtosis occurs in the data set. Following these suggestions, I estimated the proposed model with a second method, the Arbitrary Generalized Least Squares (AGLS), which is one of the ADF estimators. While the AGLS method relaxes the multinormality assumption, it is computationally demanding (Anderson and Gerbing, 1988; Bollen, 1989; Satorra, 1990) and lacks robustness when sample size is small to moderate (Satorra, 1990). Hence, the AGLS method should be cautiously applied to take advantage of its nature of asymptotically distribution-free. In the present study, I report the problems noticed under the AGLS method for evaluating a structural model of physical health and investigate those problems for the underlying reasons. I then examine the adequacy of the ML estimates that were chosen as the primary estimator in evaluating the model.

II. STATEMENT OF THE PROBLEMS

1. Model Specification

The present study adopts three dimensions of physical health as the constructs of the

proposed model. Figure 1 illustrates the proposed model of physical health. The model is composed of four latent exogenous (ξ s, ξ s) and three latent endogenous factors (η s, η s). The ϕ s (ϕ s) are covariances between the latent exogenous factors. The β s (β s) and the γ s (γ s) refer to structural coefficients, which represent direct effects between η s and ξ s, respectively. The ζ s (ζ s) denote the residuals in the structural equations. The x s and y s are observed indicators. The λ s (λ s) are the factor loadings of the observed indicators on the latent factors. The δ s (δ s) and the ϵ s (ϵ s) are the measurement errors for the observed indicators the x s and the y s, respectively.

Their corresponding variances and covariances are denoted by θ_{δ} 's (θ_{δ} 's) and θ_{ϵ} 's (θ_{ϵ} 's).

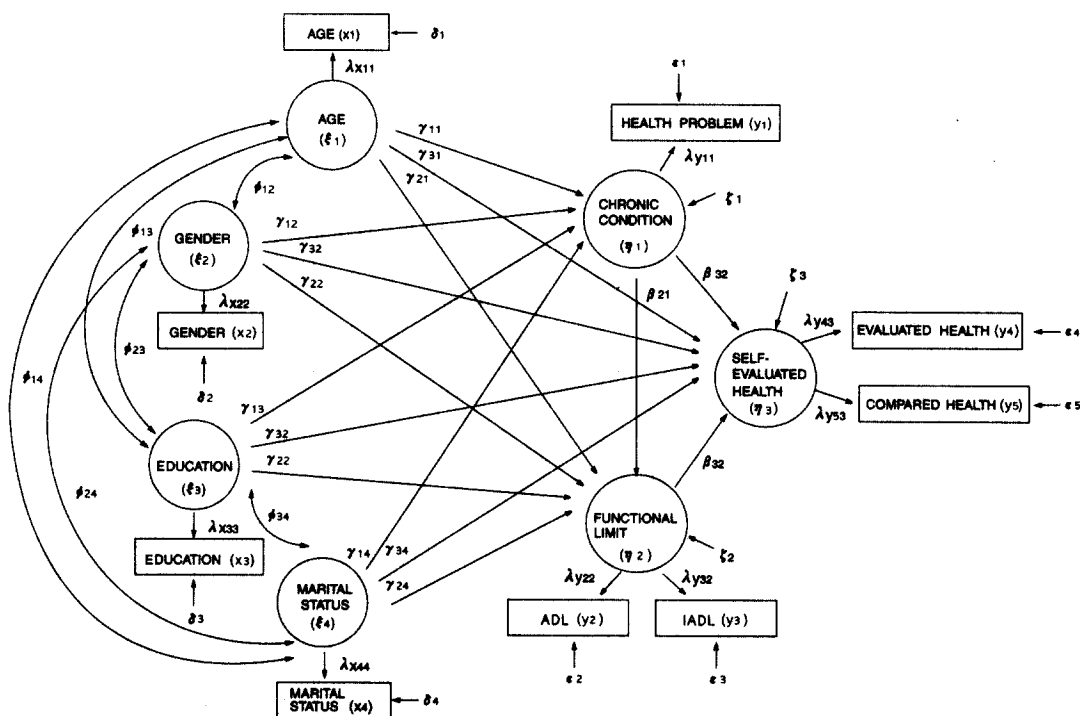
The η_1 is "chronic condition," η_2 is "functional limit," and η_3 is "self-evaluated health." Each of these constructs, respectively, represents the medical, functional, and self-evaluative dimensions. Chronic condition is measured with health problem (y_1). Health problem has three categories such as (1) experience of no health problem, (2) experience of only health problems not affecting daily activities, and (3) experience of only health problems affecting daily activities or of both types. Functional limit and self-evaluated health have two indicators each: ADL (activities of daily living, y_2) and IADL (instrumental activities of daily living, y_3) for the former factor and evaluated health (y_4) and compared health (y_5) for the latter one. ADL is the composite of subject's ability for eating, dressing, grooming, walking, in/out of bed, and bathing. Those activities such as getting to places, shopping, and handling money comprised IADL. Evaluated health is operationalized by subject's feeling about present health with 5 categories: (1) bad, (2) fairly bad, (3) average, (4) fairly good, and (5) very good. As for compared health, each subject compared health

with others of his or her own age with response categories of (1) worse, (2) about the same, and (3) better.

As sociodemographic factors, ξ_1 is "age," ξ_2 is "gender," ξ_3 is "education," and ξ_4 is "marital status." These factors are also measured with relevant single indicators. Age and education are operationalized by the actual age in years and the number of school years attended, respectively. Both gender and marital status have 2 response categories: (1) man and (2) woman for gender and (1) unmarried person and (2) married person for marital status.

In handling those single indicators, some parameters need to be constrained in order to achieve model identification. For the endogenous part of the model, both the factor loading (λ_{y11}) and the measurement error

variance (θ_i 's) of health problem are constrained. The factor loadings (λ_i 's) and the measurement error variances (θ_i 's) of the indicators are constrained for the exogenous part. A factor loading is determined by multiplying the square root of reliability and the square root of variance of the corresponding indicator. A measurement error variance is obtained by multiplying the variance of the indicator by its unreliability. The model relies on several assumptions (Long, 1983): (1) the latent exogenous factors (ξ s) and the residuals (ζ s) are uncorrelated; (2) the latent factors including ξ s and η s and the measurement errors including δ s and ϵ s are uncorrelated; (3) the measurement errors (δ s) of the exogenous indicators are uncorrelated with those (ϵ s) of the endogenous indicators.



<Figure 1> Proposed Structural Model of Physical Health

2. Problems under AGLS Estimation

The data set for model evaluation comes from the "Aging in the Western Pacific Study" (Andrews, et al., 1986). The samples were selected from the noninstitutional populations aged 60 and over in Peninsula Malaysia. The total sample size for the present analysis is 970 elders and they were randomly divided into two subsamples for replication, which leads to a check on conclusions derived from a significance test and further insight into the variability of a result (Finifter, 1972). The two subsamples (475 elders for the Sample 1 and 495 elders for the Sample 2) are very similar to each other in the distributions of age, gender, education, and marital status.

Regarding descriptive statistics for the indicators, ADL has very large, positive skewness (ranging from 6.94 to 7.46) and kurtosis (ranging from 3.74 to 7.27) in the samples. The majority of the subjects have no limit in ADL. Education has a skewed (ranging from 1.98 to 2.32) and kurtotic (ranging from 3.74 to 7.27) distribution.

Replications of the model with the AGLS method produced two types of problems in the samples. These were: (1) high correlations of parameter estimates and (2) high standard errors leading to insignificance of some estimates with large magnitudes. With regard to the first problem, some of the correlations of parameter estimates were greater than .90 in absolute values, signifying an identification problem (Bentler, 1989; Hayduk, 1987; Joreskog & Sorbom, 1989). Those highly correlated parameter estimates were particularly related with functional limit (η_2). The second problem of high standard errors was found in the factor loading (λ_{32}) of IADL in the Sample 1 and the residual error variance (Ψ_{22}) of functional limit and the measurement error variance (θ_{412}) of ADL in all the samples. These parameter estimates had high magnitudes ranging from .653 to .970 in their

standardized forms. Yet, their standard errors were too high to reject the null hypothesis at the .05 level, suggesting that these high values are not significantly different from zero.

III. METHODS

The apparent problems were investigated with two major strategies: they were to examine the endogenous part of the model and to change the metric of functional limit (η_2). The former strategy shows whether the endogenous part is the problematic area of the model by excluding the influence of the exogenous part. Based upon the result of the former strategy, the metric of functional limit was changed by changing its reference indicator with IADL, which is much less skewed and kurtotic than is ADL, the original reference indicator. Since the AGLS method turned out to be improper for the present data set, the adequacy of the ML method which was taken as the primary estimator was assessed with the robust statistics including Satorra-Bentler scaled test statistic and robust standard errors. The bootstrap method was further adopted to clarify some ambiguities found in the robust standard errors. These procedures were performed using EQS and LISREL programs.

IV. RESULTS and DISCUSSION

1. Comparisons of the Full and Endogenous Models

Examining only the endogenous part of the model by separating from the full model would narrow down the problematic areas by isolating the effect of the exogenous part. This approach is reasonable in view of the concentration of the AGLS estimation problems on the endogenous part. Under the AGLS method, the correlations between the factor loading (λ_{32}) of IADL and the residual error variance (Ψ_{22}) of functional

limit were greater than .95 in the endogenous model in all the samples. On the other hand, the corresponding correlations of the ML estimates in the endogenous model were high, but lower than .90 in most cases.

The problematic AGLS coefficients, the standard errors, and their z-statistics were also compared between the full and endogenous models (Tables 1). The residual error variances ($\sigma_{\epsilon_{22}}$) of functional limit in the endogenous model were consistently larger than those of the full model because of fewer explanatory variables for functional limit (η_2) in the endogenous model. Yet, their greater z-statistics indicate smaller standard errors than those in the full model in proportion to their respective coefficients. These smaller standard errors than those in the full model were also found in other

problematic estimates such as the factor loading of IADL and the measurement error variance of ADL.

Accordingly, all the problematically insignificant estimates in the full model turned out to be significant in the endogenous model with one exception, which was the residual error variance ($\sigma_{\epsilon_{22}}$) of functional limit in the Sample 2. However, its z-statistic (1.88) was close to the 1.96 cutoff point for statistical significance.

Based upon the disappearance of problematically insignificant AGLS estimates in the above, it is speculated that the introduction of the exogenous part of the model have produced some strain in the estimation. This speculation, however, is not sustained because of the appearance of very

<Table 1> AGLS Estimates of Full and Endogenous Models

Items	Sample 1		Sample 2		Full Sample	
	Full ¹	Endog ²	Full	Endog	Full	Endog
λ_{y2} Factor loading of IADL						
Coefficient (standardized)	.970	.795	.883	.816	.933	.803
Coefficient (unstandardized)	9.055	1.733	7.194	2.698	6.472	2.069
Standard Error	4.858	.419	2.851	.707	2.328	.265
z-statistic	1.864	4.136	2.523	3.816	2.780	7.808
$\sigma_{\epsilon_{22}}$ Residual error variance of functional limit						
Coefficient (standardized)	.757	.900	.707	.886	.748	.895
Coefficient (unstandardized)	.013	.324	.016	.137	.025	.221
Standard Error	.012	.161	.012	.073	.015	.083
z-statistic	1.083	2.012	1.333	1.877	1.667	2.663
$\sigma_{\epsilon_{22}}$ Measurement error variance of ADL						
Coefficient (standardized)	.653	.692	.797	.706	.714	.715
Coefficient (unstandardized)	.032	.809	.091	.370	.082	.622
Standard Error	.101	.243	.078	.136	.085	.148
z-statistic	.317	3.329	1.167	2.721	.965	4.203

¹ Full model

² Endogenous model: a model consisted of the endogenous part of the full model.

high correlations of the AGLS estimates in the endogenous model estimation. That is, an identification problem noted under the AGLS estimator was apparent even before bringing in the exogenous part into the model. Adding the exogenous model contributed to exposing estimation problems embedded in the endogenous model in the present analyses. This was, perhaps, caused by increased strain in the model through the added complexity from the exogenous model. Therefore, the problematic areas can be localized to the endogenous part of the model.

2. Change of the Metric of Functional Limit

Within the endogenous model, the problems with the AGLS estimation revolve around functional limit (η_2). In the model specification, a

unit change in functional limit is scaled to be the same as a unit change in ADL, which has very high skewness and kurtosis in the samples. Hence, we need to see the findings when the metric of functional limit (η_2) is constrained to be the same as that of IADL (y_3), which is much less skewed (ranging from 1.52 to 1.61) and kurtotic (ranging from 1.64 to 2.07) than is ADL (y_2). These findings are presented in Table 2.

When IADL (y_3) was used as the reference indicator for setting the metric of functional limit (η_2), the AGLS estimator produced no correlations of parameter estimates greater than .90 in any of the samples. Instead, all the correlations of parameter estimates greater than .90 under the original metric of functional limit fell below .60. Thus, the identification problem which was manifested through high correlations of estimates disappeared under the new metric of functional limit.

<Table 2> AGLS Estimates with Different Metrics of Functional Limit

Items	Sample 1		Sample 2		Full Sample	
	Constrained		Constrained		Constrained	
	ADL	IADL	ADL	IADL	ADL	IADL
λ_{y_2} , λ_{y_3} Factor loading of	IADL	ADL	IADL	ADL	IADL	ADL
Coefficient (standardized)	.970	.588	.883	.451	.933	.535
Coefficient (unstandardized)	9.055	.111	7.194	.140	6.472	.154
Standard Error	4.858	.059	2.851	.055	2.328	.056
z-statistic	1.864	1.881	2.523	2.545	2.780	2.750
σ_{η_2} Residual error variance of functional limit						
Coefficient (standardized)	.757	.757	.707	.707	.748	.748
Coefficient (unstandardized)	.013	1.070	.016	.851	.025	1.028
Standard Error	.012	.411	.012	.317	.015	.264
z-statistic	1.083	2.603	1.333	2.685	1.667	3.894
θ_{e_2} Measurement error variance of ADL						
Coefficient (standardized)	.653	.654	.797	.796	.714	.714
Coefficient (unstandardized)	.032	.033	.091	.092	.082	.082
Standard Error	.101	.101	.078	.078	.085	.085
z-statistic	.317	.327	1.167	1.179	.965	.965

With regard to the problematic AGLS estimates with high standard errors, their standardized coefficients remained the same across the samples. The large, yet insignificant, residual error variances ($\sigma_{\epsilon_{22}}$) in all the samples became significant, exhibiting reasonable statistical decisions for their large magnitudes. When the ML estimator was applied, the new metric of functional limit produced almost the same z-statistics of the residual error variance ($\sigma_{\epsilon_{22}}$) in the samples.

The z-statistics of the factor loading ($\lambda_{y_{22}}$) and the measurement error variances ($\theta_{\epsilon_{22}}$) of ADL were nearly identical between the two metrics of functional limit, providing the same statistical decisions. Since the measurement error variable (ϵ_2) of ADL has no linkages to any other component in the model, no changes occurred in its error variances ($\theta_{\epsilon_{22}}$). The unchanged feature under the new metric basically emanated from unchanged distributions of ADL in the samples.

These findings reveal that the metric of functional limit (η_2) based upon ADL(y_2) was responsible for the problems of the high correlations of parameter estimates and the insignificant residual error variances ($\sigma_{\epsilon_{22}}$) noted under the AGLS method. Remaining problems with the factor loading ($\lambda_{y_{22}}$) and the measurement error variances ($\theta_{\epsilon_{12}}$) of ADL confirm the extremely kurtotic distribution of ADL (y_2) as the focal point of the problems noted in the AGLS estimation.

3. Adequacy of Maximum Likelihood Method

Since the AGLS method turned out to be improper for the present data set, the adequacy of the ML method which was taken as the primary estimator remains to be assessed. This is more so because the present data set did not satisfy the multivariate normality assumption for the ML estimator. The ML method is known to be robust against a certain violation of the

multivariate normality assumption. Yet, it is not clear whether the degree of departures of the indicators from normality in the present data set could produce correct χ^2 statistics and standard errors. In order to examine this issue, robust statistics under the ML estimator were obtained through the EQS program. The examination further included the bootstrap method to clarify some ambiguities found in the robust standard errors.

1) The Robust Statistics

The robust statistics include both "Satorra-Bentler scaled" test statistic ("scaled test statistic" in the following) and robust standard errors. According to Bentler (1989), the test statistic is "designed to have a distribution that is more closely approximated by χ^2 than the usual test statistic" (p. 47) and the robust standard errors are correct in large samples even if the distributional assumption is violated.

The z-statistics of the free parameter estimates derived from the ML standard errors and from the robust standard errors were compared for the samples. In particular, the standard errors of the variances of gender and marital status became very small. Their z-statistics (the robust standard errors) of the variances range from 30 (.033) to 77 (.013). Since these standard errors approaching zero may reflect an estimation problem rather than high precision (Bentler, 1989), those very small values of the robust standard errors need to be taken with caution. These ambiguities from the robust standard errors will be examined in the following section.

Tables 3 shows the estimation from ML and robust methods of the problematic estimates under the AGLS method. The z-statistics derived from the robust standard errors were consistently lower than those from the ML standard errors, indicating that the ML method underestimated their standard errors. The z-statistics derived from the robust standard errors

<Table 3> Estimates from ML and Robust Methods

Items	Sample 1		Sample 2		Full Sample	
	ML	ROBUST	ML	ROBUST	ML	ROBUST
λ_{y2} Factor loading of IADL						
Coefficient (standardized)	.980		.892		.940	
Coefficient (unstandardized)	2.575		3.142		2.830	
Standard Error	.449	.663	.509	.966	.342	.559
z-statistic	5.735	3.884	6.173	3.253	8.275	5.063
$\sigma^2_{\epsilon 2}$ Residual error variance of functional limit						
Coefficient (standardized)	.757		.711		.741	
Coefficient (unstandardized)	.188		.097		.139	
Standard Error	.041	.097	.021	.052	.022	.052
z-statistic	4.585	1.938	4.619	1.865	6.318	2.673
$\theta_{\epsilon 2}$ Measurement error variance of ADL						
Coefficient (standardized)	.792		.789		.714	
Coefficient (unstandardized)	.951		.511		.730	
Standard Error	.072	.309	.038	.148	.039	.174
z-statistic	13.208	3.078	13.447	3.453	18.718	4.195
Chi-square Test statistics	27.29	26.72	15.76	16.0	36.80	35.90
Probability	.004	.005	.150	.141	.000	.000

conclude statistical significance except for the residual error variance ($\sigma^2_{\epsilon 2}$) of functional limit in the Sample 1 and in the Sample 2. Yet, their z-statistics (1.94 and 1.87) were very close to the 1.96 cutoff point at 5% significance level.

Given these differences in the estimated standard errors, it is important to know whether other statistical tests for the ML estimates could be sustainable when using robust standard errors. Comparisons of a z-statistic for each parameter estimate in the model revealed that the statistical decisions from the ML standard errors and robust standard errors did not change in most estimates for all the samples. In addition to the residual error variance ($\sigma^2_{\epsilon 2}$) of functional limit mentioned above, there were two exceptional estimates. The estimate of the effect (y_{23}) of education on functional limit

became insignificant with its robust standard error in the Sample 1. In this exception, the z-statistics of the insignificant estimates were close to the 1.96 cutoff point. In terms of the overall fit of the model, the scaled χ^2 values were generally smaller than the χ^2 values derived from the ML estimator. Yet, these changes led to few changes in the significance level through the p-values, indicating the improvements in the χ^2 values were nominal.

2) Application of the Bootstrap Method

Very small values of the robust standard errors of gender and marital status mentioned above may indicate either a high precision in estimates or some noise in estimation. Accordingly, the accuracy of the standard errors obtained from the ML method remains to be

answered. The bootstrap method was adopted in order to handle these issues.

As a general methodology to answer the accuracy of an unknown parameter estimator (Efron and Tibshirani, 1986), the bootstrap generates standard errors and confidence

intervals "that are typically better than alternatives that rely on untested assumptions" (Stine, 1989). The bootstrap does not assume a particular distribution of the variables, while it assumes comparability between the bootstrap sampling distribution of the estimates and the

<Table 4> Unstandardized Free Parameter Estimates and Standard Errors from ML, Robust, and Bootstrap Methods for the Full Sample (N=970)

Parameters	Original ML		Robust ¹	Bootstrap ²	
	Estimate	SE ³	SE	Avg. Est ⁴	SE
Variances of the latent exogenous factors (ϕ s)					
ϕ_{11} Age	1.000	.049	.056	1.000	.003
ϕ_{22} Gender	.998	.045	.013	1.000	.003
ϕ_{33} Education	1.000	.053	.092	1.001	.012
ϕ_{44} Marital Status	1.005	.076	.021	1.001	.004
Covariances of the latent exogenous factors (ϕ s)					
ϕ_{12} Age, Gender	-.120	.034	.033	-.124	.033
ϕ_{13} Age, ion	-.468	.038	.034	-.465	.032
ϕ_{24} Gender, Marital Status	-.546	.045	.036	-.546	.033
ϕ_{34} Education, Marital Status	.368	.046	.041	.371	.038
Factor loadings of the latent endogenous factors (λ y's)					
λ_{y32} IADL	2.830	.342	.559	2.920	.647
λ_{y53} Compared health	.481	.034	.031	.481	.033
Effects associated with endogenous factors (β s)					
η_1 Chronic condition on:					
η_2 Functional limit	.117	.020	.031	.120	.035
η_3 Self-evaluated health	-.455	.032	.033	-.455	.035
η_2 Functional limit on:					
η_3 Self-evaluated health	-.478	.084	.106	-.493	.118
Effects associated with exogenous factors (cs)					
ξ_1 Age on:					
η_1 Chronic condition	-.032	.041	.042	-.034	.040
η_2 Functional limit	.136	.022	.035	.140	.038
η_3 Self-evaluated health	-.001	.033	.035	.001	.037

<Table 4> Unstandardized Free Parameter Estimates and Standard Errors from ML, Robust, and Bootstrap Methods for the Full Sample (N=970)
(continued)

Parameters	Original ML		Robust ¹	Bootstrap ²	
	Estimate	SE ³	SE	Avg. Est	SE
ξ_2 Gender on:					
η_1 Chronic condition	-.053	.054	.054	-.053	.058
η_2 Functional limit	.048	.021	.024	.049	.025
η_3 Self-evaluated health	.102	.040	.041	.097	.039
ξ_3 Education on:					
η_1 Chronic condition	.045	.045	.048	.043	.050
η_2 Functional limit	-.050	.018	.019	-.048	.020
η_3 Self-evaluated health	.171	.033	.032	.167	.034
ξ_4 Marital status on:					
η_1 Chronic condition	-.155	.064	.064	-.157	.070
η_2 Functional limit	-.045	.025	.026	-.046	.027
η_3 Self-evaluated health	.009	.048	.049	.002	.051
Residual error variances in equations (Ψ 's)					
Ψ_{11} Chronic condition	.984	.054	.032	.977	.015
Ψ_{22} Functional limit	.139	.022	.052	.145	.057
Ψ_{33} Self-evaluated health	.381	.046	.048	.376	.049
Measurement error variances in indicators (θ 's)					
θ_{e22} ADL	.730	.039	.174	.730	.178
θ_{e33} IADL	.200	.158	.155	.202	.146
θ_{e44} Evaluated health	.250	.044	.047	.244	.048
θ_{e55} Compared health	.267	.016	.014	.268	.014

¹ Robust standard errors are derived from the Maximum Likelihood fitting function.

² The Bootstrap sample size (the number of replications) is 200.

³ Standard errors

⁴ Average estimate of 200 replications.

Note: In 31 replications, measurement error variance of IADL was constrained at lower bound.

empirical sampling distribution of the estimates (Bollen and Stine, 1990). Hence, the bootstrap is a proper method of evaluating the obtained standard errors in the current analyses given the nonnormal distribution of the variables in the present data set.

A bootstrap sample is a random sample

drawn with replacement from the actual sample. The number of resampling determines the size of a bootstrap sample. In general, a bootstrap sample size of 100 is recommended for estimating standard errors, while an adequate size ranges between 50 and 200 (Efron and Tibshirani, 1986; Stine, 1989). Because of the

large amount of work involved in pursuing the bootstrap with the model examined, the bootstrap approach was applied only for the Full sample. In each sampling from the Full sample with 970 cases, the same number (970) was sampled 200 times to form a bootstrap sample. The standard deviations of the estimates from the bootstrap sample of size 200 become the criteria to compare the original ML and the robust ML standard errors.

Table 4 presents the unstandardized parameter estimates and their standard errors from three different methods. With regard to the parameter estimates, the averages of the bootstrap estimates approximate the original ML estimates, suggesting that the ML estimates are unbiased. The standard errors were similar with some exceptions. These exceptions involve those variables with ADL which is highly kurtotic in the data set. When there is a discrepancy, the robust standard errors were much closer to the bootstrap standard errors than were the original ML standard errors. In particular, the bootstrap standard errors of age, gender, and marital status were smaller than the robust standard errors, eliminating the suspected identification problem emanating from very small standard errors in the robust statistics reported previously. Hence, it is concluded that the robust standard errors are very reasonable in the present study, whereas some of the ML standard errors are biased. A more important issue is whether the statistical decisions derived from the ML estimator could remain stable. As presented before, stability is empirically supported by almost identically made statistical decisions at the .05 significance level between the ML and the robust standard errors.

V. CONCLUSIONS

The proposed model was primarily estimated with the ML method and replicated with the

AGLS method due to the violation of the data set for multinormality assumption of the ML method. However, the AGLS estimator was influenced by extremely nonnormal distribution of ADL, which is an indicator of the latent factor, functional limit, and thus, it behaved inadequately in the present modeling conditions. This observation suggests that there may be a threshold for the AGLS estimator in accepting nonnormal distributions of indicators in order to produce proper results given the present conditions.

The ADF method is known to have the merit of the asymptotically distribution free, which is insensitivity to the distributions of data. Despite having this merit, the method lacks robustness against small and moderate sample sizes because its asymptotic properties are proven to be true only for large samples (Anderson and Gerbing, 1988; Satorra, 1990). For example, an empirical study (Hu, Bentler, and Kano, 1992) required a sample size of 5000 to produce satisfactory χ^2 statistics for a given model. On the other hand, Boomsma (1987) reported robustness of the ML estimator against small samples if the size is greater than 200 for the model tested. With different modeling conditions, the sample sizes of the present data set are well above 200, with which the ML estimator worked well. However, it is hard to determine the adequate sample size for the AGLS estimator in the present model since there are no guidelines for such sample sizes for AGLS applications. Nevertheless, it is most likely that the present sample sizes are far below that needed to produce adequate AGLS estimates in the present modeling conditions such as the degree of nonnormal distributions of observed variables and the degree of model complexity.

The adequacy of the ML method for evaluating the proposed model was demonstrated through comparisons of the ML and the robust statistics in light of the bootstrap findings. The

ML estimates were unbiased. Although the robust standard errors were better than the original ML standard errors, statistical decisions based upon the original ML standard errors remained almost the same.

It is suggested that a reference indicator for determining the metric of a latent variable should not have high skewness and kurtosis. A safe practice would be to choose a reference indicator as the one with closest to normal distribution. Since the nonnormal distributions of variables produce biased standard errors and χ^2 statistics from the ML method, it is recommended to adopt the statistics corrected for the nonnormality such as the robust standard errors. In addition, the ADF estimators need to be used with caution, since the estimators tend to be unstable with high standard errors when the data set does not satisfy its requirement of a large sample size under the complexity of a model to be estimated.

REFERENCES

- Anderson, J. C. & Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin, 103(3), 411-423.
- Andrews, G. R., Esterman, A. J., Braunack-Mayer, A. J., & Rungie, C.M. (1986). Ageing in the Western Pacific. Manila: World Health Organization Regional Office for the Western Pacific.
- Bentler, P. M. (1989). EQS structural equations program manual. Los Angeles: BMDP Statistical Software.
- Bollen, K. A. & Stine, R. (1990). Direct and indirect effects: classical and bootstrap estimates of variability. in C.C. Clogg (Ed.), Sociological Methodology 1990, Volume 20. Oxford: Basil Blackwell.
- Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley & Sons.
- Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. in P. Cuttance & R. Ecob (Eds.), Structural modeling by examples: applications in educational, sociological, and behavioral research. New York: Cambridge University Press.
- Cuttance, P. (1987). Issues and problems in the application of structural equation models. in P. Cuttance & R. Ecob (Eds.) Structural modeling by example: applications in educational, sociological, and behavioral research. New York: Cambridge University Press.
- Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science, 1(1), 54-77.
- Finifter, B. M. (1972). The generation of confidence: evaluating research findings by random subsamples replication. in H. L. COSTNER (Ed.), Sociological methodology 1972. San Francisco: Jossey-Bass Inc., Publishers.
- Hayduk, L. A. (1987). Structural equation modeling with LISREL. Baltimore: The Johns Hopkins University Press.
- Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112(2), 351-362.
- Joreskog, K. G. & Sorbom, D. (1989). LISREL 7: a guide to the program and application. 2nd edition. Chicago: SPSS Inc.
- Long, J. S. (1983). Covariance structure models: an introduction to LISREL. Beverly Hills, CA: Sage Publications.
- Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. Quality and Quantity, 24, 367-386.
- Stine, R. (1989). An introduction to bootstrap methods: examples and ideas. Sociological Methods and Research, 18(2 & 3), 243-291.