DSSR

Discussion Paper No.101

Construct Validation for a Nonlinear Measurement Model in Marketing and Consumer Behavior Research

Toshikuni Sato

August, 2019

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Construct Validation for a Nonlinear Measurement Model in Marketing and Consumer Behavior Research

Toshikuni Sato*

**Graduate School of Economics and Management, Tohoku University*

Abstract

This study proposes a method to evaluate the construct validity for a nonlinear measurement model. Construct validation is required when applying measurement and structural equation models to measurement data from consumer and related social science research. However, previous studies have not sufficiently discussed the nonlinear measurement model and its construct validation. This study focuses on convergent and discriminant validation as important processes to check whether estimated latent variables represent defined constructs. To assess the convergent and discriminant validity in the nonlinear measurement model, previous methods are extended and new indexes are investigated by simulation studies. Empirical analysis is also provided, which shows that a nonlinear measurement model is better than linear model in both fitting and validity. Moreover, a new concept of construct validation is discussed for future research: it considers the interpretability of machine learning (such as neural networks) because construct validation plays an important role in interpreting latent variables.

Keywords: Construct validation, Nonlinear measurement model, Reliability coefficient, Convergent validity, Discriminant validity

1. Introduction

The psychological scale, known as the "marketing scale" in marketing and consumer behavior research, is an instrument used to measure latent psychological constructs by applying factor analysis as measurement model. Assuming some constructs for consumer psychologies and behaviors, structural equation modeling (SEM) is often used with these constructs specified by the measurement model. Before estimating by SEM, we usually evaluate reliability and validity to check the accuracy of the estimated constructs. Hence, construct validity is an important topic to estimate the causal relationship among constructs in consumer research.

Construct validity has been discussed by a number of researchers (e.g., Cronbach & Meehl 1955; Campbell & Fiske 1959; Bagozzi et al. 1991; Anderson & Gerbing 1992; Messick 1995; Edwards 2001; 2003; Hughes 2018), and the modern concepts have been established by Messick (1995). Because we deal with uncertain and unobserved variables, researches are concerned about reliability and validity of latent variables; from not only a theoretical but also an empirical perspective. Therefore, some statistical methods of construct validation have been discussed and developed uniquely in the marketing area (Hair et al. 2009; Bagozzi & Yi 1988; Fornel & Lacker 1981).

The measurement model and validation for the constructs have a strong relationship with classical test theory (CTT). Although most researchers have not mentioned this relationship in practical research, CTT is a very important subject in psychometrics. In addition, the relationship between CTT and item response theory (IRT) is given Turker (1946) and Lord and Novick (1968); thus, IRT model is recognized as one kind of nonlinear CTT model in psychometrics (Lewis 2006).

In consumer research, however, CTT is always assumed implicitly when using the measurement model with questionnaires. Besides, methods related to measuring constructs have been extended with a linear CTT assumption; that is, observed scores are linearly rerated to true scores. Although this assumption makes it easier to measure true scores and to estimate reliability, it is necessary to consider the possibility of measuring error problem caused by choosing an inappropriate functional relationship between the observed and true scores.

The purpose of this study is to discuss a nonlinear measurement model and its construct validation in consumer research. First, we review the linear measurement model and the construct validation. Second, we discuss effective construct validation methods for a nonlinear measurement model. Third, the results of several simulation studies and empirical analysis using SEVQLAL (PZB 1985; 1988) are provided. Finally, we discuss the importance of construct validation and its extension to interpretable machine learning.

2. Linear Measurement Model and Construct Validation

2.1. Linear Factor Analysis Model and CTT

CTT is a traditional psychological measurement theory based on the concept of a "true score" in psychometrics (e.g., Novick 1966; Traub 1997; Jones & Thissen 2006; Lewis 2006). In the most basic approach to the measurement model of CTT, the observed score Z is considered to be the sum of a true score T and a random error E:

$$Z = T + E. \tag{1}$$

The standard deviation of the errors E indicates a statement of the (rack of) precision, or standard error, of the observed score. We want to measure the true score T, but we can only obtain the observed score containing the measurement error. Because the true score can be regarded as a latent variable, factor analysis is a standard method used to estimate the true score T, called the "construct" or "latent trait."

There are mostly three kinds of definitions for the measurement model, depending on different parameter assumptions (Jöreskog 1971; Novick & Lewis 1967; Rajaratnam et al.

1965); see Figure 1. To explain the difference among the three measurement models with factor analysis, we define a general equation form for independent individual i $(i = 1, \cdots, n)$ and for item j $(j = 1, \cdots, p)$:

$$z_{ji} = \lambda_j t_i + \varepsilon_{ji}, \tag{2}$$

where z_{ji} is a observed or standardized observed variable, λ_j is a factor loading called the "discrimination parameter" (or "regression coefficient") for item j, t_i is a common factor or a latent variable corresponding to the construct as a true score, and ε_{ji} is the measurement error assumed to be distributed as a normal distribution. The assumptions of CTT are represented by (2) with the following equations:

$$E(t_i) = 0 \text{ for all } i, \tag{3}$$
$$Var(t_i) = 1 \text{ for all } i, \tag{4}$$
$$E(\varepsilon_{ji}) = 0 \text{ for any } j \text{ and all } i, \tag{5}$$
$$Var(\varepsilon_{ji}) = \psi_j \text{ for any } j \text{ and all } i, \tag{6}$$
$$Cov(\varepsilon_{ji}, \varepsilon_{si}) = 0 \text{ for any } j \neq s \text{ and all } i, \tag{7}$$
$$Cov(t_i, \varepsilon_{ji}) = 0 \text{ for any } j \text{ and all } i. \tag{8}$$

The first, parallel measurement model is that the construct has the same degree of discrimination for each item and that the precision for each item is common. Hence, the following restrictions are additionally assumed:

$$\lambda_1 = \lambda_2 = \cdots = \lambda_p, \tag{9}$$
$$\psi_1 = \psi_2 = \cdots = \psi_p. \tag{10}$$

The second, tau-equivalent measurement model, assumes that the construct has the same discrimination for each item, but that all the items have a different precision. Hence, we additionally assume restriction (9) and that ψ_j for any j is a parameter. The third, congeneric measurement model assumes that the construct has a different discrimination for each item and that each item has a different precision. Hence, λ_j and ψ_j for any j are treated as parameters.

Therefore, each model can be estimated by factor analysis model with setting above restrictions. In marketing and most the other social science areas, congeneric measurement model is a standard method to estimate constructs.

Figure 1: Three different measurement equations

2.2. Misspecification between Reflective and Formative Models

Another kind of measurement model, the formative model, represents a principal component analysis (PCA) model specification. Although this model can be regarded as one kind of the factor analysis model specification from the view of probabilistic principal component analysis (PPCA), the refractive and formative Model are treated as different specifications (see Figure 2) in consumer behavior research. Jarvis et al. (2003) discussed the misspecification between refractive and formative models in consumer behavior research. They investigated the top journals related to Marketing (*Journal of Marketing Research, Journal of Marketing, Journal of Consumer Research, Marketing Science*,) and found some studies in even those top journals contain the misspecification. Because this misspecification provides a different estimate for the parameters in the structure model, it is important to clarify the assumptions between observable and latent variables when applying the measurement model.

Figure 2: Reflective and formative models

2.3. Linear Factor Analysis Model and Construct Validation

This section introduces different kinds of reliability coefficients and a method to evaluate the convergent and discriminant validity for construct validation.

2.3.1. Measurement Model and Reliability Coefficient

Reliability in CTT is defined as the proportion of observed score variance due to variance among individual true scores (Novic 1966; Lewis 2006; Webb et al. 2006). Coefficient alpha or Cronbach's alpha (Cronbach 1951) is most frequently used in the present methods (MacKenzie et al. 2011). From the composite measurement (Novic & Lewis 1967) aspect, we can obtain another expression of Cronbach's alpha in Eq. (11) and appendix A.1, and it is helpful to understand the relationship between the measurement model and the reliability coefficient. Equation (3) indicates that Cronbach's alpha represents a reliability coefficient when assuming the tau-equivalent test. In other words, this reliability estimates a coefficient to evaluating a measurement model with the condition that the factor ladings are equal for all observed variables. Therefore, when standard factor analysis is assumed, Cronbach's alpha is not suitable to evaluate the reliability for the measurements:

$$\alpha_t = \frac{p}{p-1}\left[1-\frac{\sum_{j=1}^{p} Var\left(z_{ji}\right)}{Var\left(Z\right)}\right] \Leftrightarrow \frac{p^2\lambda^2}{p^2\lambda^2+\sum_{j=1}^{p}\psi_j}. \tag{11}$$

Another well-known estimator for reliability is coefficient omega (McDonald 1978). As in the case of coefficient alpha (see Appendix A.2), coefficient omega can be expressed as Eq. (12). This is a reasonable estimator for the reliability of a congeneric test, which is a standard assumption of factor analysis. Moreover, the third entity in (12) was proposed for construct reliability (CR) by Fornell & Larcker (1981) in the marketing area (see also Hair et al. 2009; MacKenzie et al. 2011). This estimator is also valid for the parallel and tau-equivalent tests so that coefficient omega (or CR) is a generalization of the reliability estimator among the three basic test models:

$$\omega_t = 1-\frac{\sum_{j=1}^{p}\psi_j}{Var\left(Z\right)} \Leftrightarrow \frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2}{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j}. \tag{12}$$

2.3.2. Convergent and Discriminant Validity

Convergent validity is a confirmation that measures for the same construct have adequate relationships with each other, and the measures should be distinguished from that for other constructs. This is called "discriminant validity." Both validations are required for justification of a novel trait measure, validation of test interpretation and establishing construct validity (Campbell and Fiske 1959). Campbell and Fiske (1959) proposed multi trait method matrix (MTMM) to evaluate convergent and discriminant validity jointly. However, it is inconvenient for secondary users to prepare additional different measurement methods. Moreover, Bagozzi et al. (1991) showed that MTMM is not effective in several situations because of the limited assumptions.

Confirmatory factor analysis (CFA) also provides a method for convergent and discriminant validation (Anderson & Gerbing 1988; Bagozzi & Yi 1988 Bagozzi & Phillips 1982). In most situations, applying CFA results is useful to check construct validity. However, comparison between the fixed correlation (equal to 1) and the unfixed CFA models for discriminant validity is not effective because high correlation (equal to 0.9) can still produce significant differences in fit between the two models (Hair et al. 2009).

For effective judgment, average variance extracted (AVE), which was also produced by Fornell & Larcker (1981), can be applied to evaluate both convergent and discriminant validity (Fornell & Larcker 1981; Hair et al. 2009; MacKenzie et al. 2011). AVE is defined as Eq. (13) and is required to be > 0.5 for convergent validity. AVE can be regarded as an average of factor loadings (Hair et al. 2009) because the sum of standardized commonality and uniqueness is equal to 1. Compared with CR, AVE does not contain the cross terms of each factor loading because the square is inside the summation such that AVE indicates the average of the independent degree of the relationship between observed variables and a construct:

$$AVE_t = \frac{\sum_{j=1}^{p} \lambda_j^2}{\sum_{j=1}^{p} \lambda_j^2 + \sum_{j=1}^{p} \psi_j} \text{ or } \frac{\sum_{j=1}^{p} \lambda_j^2}{p}. \tag{13}$$

The criterion of discriminant validity is required so that each AVE is larger than the squared correlation among constructs.

In practice, we usually estimate the true score variance; thus, CR and AVE in these formulas are calculated by standardized factor loadings and uniqueness with converting $Var(t_i) = 1$. Otherwise, we use the following equations directly by replacing $Var(t_i)$ with an estimated value.

$$CR_t^* = \frac{\left(\sum_{j=1}^{p} \lambda_j\right)^2 Var(t_i)}{\left(\sum_{j=1}^{p} \lambda_j\right)^2 Var(t_i) + \sum_{j=1}^{p} \psi_j}. \tag{14}$$

$$AVE_t^* = \frac{\sum_{j=1}^{p} \lambda_j^2 Var(t_i)}{\sum_{j=1}^{p} \lambda_j^2 Var(t_i) + \sum_{j=1}^{p} \psi_j}. \tag{15}$$

2.3.3. Example for Problems of Invalidity

Here, we consider the insufficient convergent and discriminant validities (see Figure. 3). The first problem is unexpected small factor loading, hence, a small AVE. The equation of the relationship between t_1 and z_1 in Figure 3 can be expressed as follows:

$$z_{1,i} = 0.05 t_{1,i} + \varepsilon_{1,i}, \quad \varepsilon_{1,i} \sim N(0, 0.9975). \tag{16}$$

Because the measurement model represents a regression of observed variables on latent variables, this model cannot discriminate the answer in z_1. For example, we assume t_1 indicates "satisfaction." If $t_{1,i}$ takes 5 as strongly satisfied, then this model predicts $\hat{z}_{1,i} = 0.25$. If $t_{1,i}$ takes -5 as strongly dissatisfied, then this model predicts $\hat{z}_{1,i} = -0.25$. Hence, this model expresses that both satisfied and dissatisfied consumers will answer very close score in z_1 even if they have different degrees of potential satisfaction. In addition, owing to the large measurement error, this model indicates that the scores in z_1 will be observed randomly rather than depending on the satisfaction.

The second problem is unexpected large correlation among constructs. In the model from Figure 3, $\widehat{\mathrm{AVE}_2} \cong 0.7$ is larger than $\hat{r}_{1,2}^2 = 0.64$ but $\widehat{\mathrm{AVE}_1} \cong 0.26$ is not. This example indicates that t_1 has a stronger relationship with t_2 than z_1, z_2, and z_3 even if one assumed the exact relationship between the observed variables and the construct. Therefore, this model cannot distinguish the difference between t_1 and t_2; hence, these constructs can be regarded as almost the same construct.

Figure 3: The problem of a small factor lading and a large correlation

For instance, a price indicates the price exactly; however, the items of measurement are defined by the researcher with some assumptions and theories. Hence, evaluating convergent and discriminant validity is important for the interpretation and explanation of each construct, especially in consumer research when treating very similar constructs.

3. Nonlinear Measurement Model and Its Construct Validation

This section discusses a nonlinear measurement model and its construct validation considering a nonlinear process in consumers' evaluation and decision making. In Section 2, we discussed that the measurement model represents a generating process of observed scores so that the true score assumed to appear linearly by adding random errors. Several researches establish a model while assuming the respondents consistently understand the questions, and are able and willing to answer them (Fowler & Cannell 1996). However, the answering questions sometimes involves complex thinking, and it then causes "Rater Errors" (see Mathis & Jackson 2010, pp.347-349). Although one expects the respondent to answer honesty, in most cases the answer might depend on individual standards or experiences. Respondents may determine which information they ought to provide by relying on relative previously formed attitudes or judgements from their memories, or whatever relevant accessible information, when they answer the questions (Schwarz 2007).

3.1. Nonlinear Measurement Model

Focusing on only linearity in the generating process of observable scores may produce improper estimates for the true scores. In addition, construct validation may lead to incorrect results because the previous method is based on the linear measurement model. Therefore, we consider the following nonlinear measurement model and its construct validation:

$$z_{ji} = \lambda_j f\left(t_i\right) + \varepsilon_{ji}, \tag{17}$$

This model uses one kind of nonlinear specification that enables extension to IRT model because IRT model regards the observed score as probability and is specified by a logistic function or cumulative normal distribution function. In addition, a basic IRT model has an exact relationship with linear categorical factor analysis (Lewis 2006). Although above model is extended in line with CTT, several kinds of functions can be specified in this model. The estimation of the above nonlinear measurement model can be replaced to nonlinear factor analysis (e.g., Zhu & Lee 1999).

3.2. Construct Validation for the Nonlinear Measurement Model

In Section 2, we introduced CR for reliability and AVE for convergent and discriminant validity, which are important indexes in construct validation. Therefore, we propose CR and AVE for the nonlinear measurement model. The reliability coefficient can be regarded as a unit slope for the regression of observed scores on true scores (Novic 1966). Hence, we may replace the estimation of the reliability coefficient with an estimation of marginal effects of true scores on the observed scores. However, it is required to evaluate the true score variance with a functional transformation so that CR and AVE for Eq. (17) are approximated by the following equation with Taylor series approach:

$$\begin{aligned} CR_t' &= \frac{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2 Var\left\{f\left(t_i\right)\right\}}{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2 Var\left\{f\left(t_i\right)\right\}+\sum_{j=1}^{p}\psi_j} \\ &\approx \frac{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2 \left\{f'\left(E\left(t_i\right)\right)\right\}^2 Var\left(t_i\right)}{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2 \left\{f'\left(E\left(t_i\right)\right)\right\}^2 Var\left(t_i\right)+\sum_{j=1}^{p}\psi_j}, \end{aligned} \tag{18}$$

$$\begin{aligned} AVE_t' &= \frac{\sum_{j=1}^{p}\lambda_j^2 Var\left\{f\left(t_i\right)\right\}}{\sum_{j=1}^{p}\lambda_j^2 Var\left\{f\left(t_i\right)\right\}+\sum_{j=1}^{p}\psi_j} \\ &\approx \frac{\sum_{j=1}^{p}\lambda_j^2 \left\{f'\left(E\left(t_i\right)\right)\right\}^2 Var\left(t_i\right)}{\sum_{j=1}^{p}\lambda_j^2 \left\{f'\left(E\left(t_i\right)\right)\right\}^2 Var\left(t_i\right)+\sum_{j=1}^{p}\psi_j}, \end{aligned} \tag{19}$$

where $f'\big(E(t_i)\big) = \frac{df(t_i)}{dt_i}\Big]_{t_i=E(t_i)}$ and $f'\big(E(t_i)\big) \neq 0$.

These estimators produce the same results of original CR and AVE in linear measurement model and the detail of these indexes are explained in Appendix B. In practice, Eq. (18) and (19) can be used by replacing $E(t_i) = 0$ and $Var(t_i) = \sigma_t^2$, because we usually assume $t_i \sim N(0, \sigma_t^2)$.

4. Simulation Study

To investigate the performance of CR′ and AVE′, we prepared the following common settings for simulation studies. The dataset is generated with a sample size of n = 300 from a nonlinear measurement model defined as

$$\begin{aligned} &\mathbf{z} = \Lambda \boldsymbol{F}\left(\mathbf{t}\right)+\boldsymbol{\varepsilon}, \\ &\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \Psi\right), \end{aligned} \tag{20}$$

with six observed variables that are related to two basic latent variables $\big(\boldsymbol{t}_{(1)}, \boldsymbol{t}_{(2)}\big)$, and a nonlinear function $F\big(\boldsymbol{t}_{(1)}, \boldsymbol{t}_{(2)}\big)$. The factor loadings are given by

$$\Lambda^T = \begin{bmatrix} 1 & \lambda_{2,1} & \lambda_{3,1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{4,2} & \lambda_{5,2} \end{bmatrix}, \tag{21}$$

where the 1s and 0s are treated as known fixed parameters, and the $\lambda_{j,k}$ are unknown parameters. The true population values of the unknown parameters are given by $\lambda_{j,k} = 1$ for all j and k as specified in Λ. The variance covariance matrix of latent variables $\boldsymbol{t}$ is given by $\big(\phi_{11}, \ \phi_{12}, \ \phi_{22}\big) = (1, 0.5, 1)$. The variance of each measurement error is given by $\psi_{jj} = 1.5$

for all $j = 1, \cdots, 6$. Bayesian estimation is adopted to obtain estimates for the parameters (see Appendix D).

4.1. Study 1: Logistic Function

In the first example, consider a logistic function defined as,

$$f\left(t_{k,i}\right)=\left\{\frac{1}{1+\exp\left(t_{k,i}\right)}-\frac{1}{2}\right\}C, \tag{22}$$

where $C = 7$ so that (22) takes -3.5 and 3.5 as the minimum and maximum values of the curve, respectively, and $f(0) = 0$. Hence, CR′ and AVE′ are given by

$$\underset{\langle logistic\rangle}{CR'_k} = \frac{\left\{\sum_{j=1}^{p}\lambda_{j,k}\right\}^2 \left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)}{\left\{\sum_{j=1}^{p}\lambda_{j,k}\right\}^2 \left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)+\sum_{j=1}^{p}\psi_j} \tag{23}$$

$$\underset{\langle logistic\rangle}{AVE'_k} = \frac{\sum_{j=1}^{p}\lambda_{j,k}^2 \left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)}{\sum_{j=1}^{p}\lambda_{j,k}^2 \left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)+\sum_{j=1}^{p}\psi_j} \tag{24}$$

Table 1 shows the result of study 1 and indicates that each HPDI for the bias between the parameter and the bias contains 0 so that the estimates by proposed CR′ and AVE′ were close to true settings.

Table 1: Results of the logistic function

However, the maximum and minimum values of a curve are unknown in practice; hence, we replace function (22) as shown below:

$$f\left(t_{k,i}\right)=\left\{\frac{1}{1+\exp\left(t_{k,i}\right)}-\frac{1}{2}\right\}z^*, \tag{25}$$

where $z^* = max(\mathbf{z}^*) - min(\mathbf{z}^*)$ represents a range of standardized dataset $\mathbf{z}^*$. We used the dataset generated from (22) with common settings whereas the model was specified (25) with $z^* = 6.018$. To compare the estimates with true parameters, we calculated the standardized parameters and estimates shown in Table 2. The results show that CR′ and AVE′ were estimated nearly unbiased by proposed method.

Table 2: Results of the logistic function in practice

4.2. Study 2: Quadratic Function

For the second example, consider the following quadratic function:

$$f\left(t_{k,i}\right)=\left\{I\left(t_{k,i}\geq 0\right)-I\left(t_{k,i}<0\right)\right\}t_{k,i}^{2}, \tag{26}$$

where I is an indicator function that takes the value 1 if the condition is satisfied and 0 otherwise. Therefore, the model can also be expressed as

$$z_{ji}=\lambda_{j,k}I\left(t_{k,i}\geq 0\right)t_{k,i}^{2}-\lambda_{j,k}I\left(t_{k,i}<0\right)t_{k,i}^{2}+\varepsilon_{ji}. \tag{27}$$

In this case, it is not so difficult to derive the variance of $t_{k,i}^2$ because of the well-known relationship between normal distribution and chi-squared distribution. Because $y_i^2 \sim \chi^2(1)$ with $E(y_i^2)=1$ and $Var(y_i^2)=2$ when $y_i \sim N(0,1)$ and $\sqrt{\sigma^2}y_i = t_i \sim N(0,\sigma^2)$, we obtain $Var(t_i{}^2)=Var\left\{\left(\sqrt{\sigma^2}y_i\right)^2\right\}=\sigma^4 Var(y_i^2)=2\sigma^4$. Hence, CR′ and AVE′ are defined as follows:

$$\begin{aligned}\underset{\langle quadratic\rangle}{CR'_k}&=\frac{2\left\{Var\left(t_{k,i}\right)\right\}^2\tilde{V}}{2\left\{Var\left(t_{k,i}\right)\right\}^2\tilde{V}+n\sum_{j=1}^{p}\psi_j}\\&=\frac{2\left\{Var\left(t_{k,i}\right)\right\}^2\left\{\sum_{j=1}^{p}\lambda_{k,j}\right\}^2}{2\left\{Var\left(t_{k,i}\right)\right\}^2\left\{\sum_{j=1}^{p}\lambda_{k,j}\right\}^2+\sum_{j=1}^{p}\psi_j},\end{aligned} \tag{28}$$

where

$$\begin{aligned}\tilde{V}&=\sum_{i=1}^{n}\left[\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i}\geq 0\right)-\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}\right]^2\\&=\sum_{i=1}^{n}\left[\left\{\sum_{j=1}^{p}\lambda_{j,k}I\left(t_{k,i}\geq 0\right)\right\}^2+\left\{\sum_{j=1}^{p}-\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}^2\right]\\&=\sum_{i=1}^{n}\left[\left\{\sum_{j=1}^{p}\lambda_{j,k}I\left(t_{k,i}\geq 0\right)\right\}^2+\left\{\sum_{j=1}^{p}\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}^2\right]\\&=n\left\{\sum_{j=1}^{p}\lambda_{j,k}\right\}^2,\end{aligned} \tag{29}$$

and

$$\begin{aligned}\underset{\langle quadratic\rangle}{AVE'_k}&=\frac{2\left\{Var\left(t_{k,i}\right)\right\}^2\bar{V}}{2\left\{Var\left(t_{k,i}\right)\right\}^2\bar{V}+n\sum_{j=1}^{p}\psi_j}\\&=\frac{2\left\{Var\left(t_{k,i}\right)\right\}^2\sum_{j=1}^{p}\lambda_j^2}{2\left\{Var\left(t_{k,i}\right)\right\}^2\sum_{j=1}^{p}\lambda_{k,j}^2+\sum_{j=1}^{p}\psi_j},\end{aligned} \tag{30}$$

where

$$\begin{aligned}\bar{V} &= \sum_{i=1}^{n}\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i}\geq 0\right)-\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}^{2} \\ &= \sum_{i=1}^{n}\left[\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i}\geq 0\right)\right\}^{2}+\sum_{j=1}^{p}\left\{-\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}^{2}\right] \\ &= \sum_{i=1}^{n}\left[\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i}\geq 0\right)\right\}^{2}+\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i}<0\right)\right\}^{2}\right] \\ &= n\sum_{j=1}^{p}\lambda_{j,k}^{2}.\end{aligned} \tag{31}$$

Table 3 shows the results of study 2 and indicates that CR′ and AVE′ were estimated closely to true settings by proposed method.

Table 3: Results of the quadratic function

4.3. Study 3: Asymmetric Function

Set the following factor ladings so that the model contains asymmetry.

$$\Lambda^{T}=\begin{bmatrix}1 & \lambda_{21} & \lambda_{31} & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & \lambda_{52} & \lambda_{62}\\ \lambda_{13} & \lambda_{23} & \lambda_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & \lambda_{44} & \lambda_{54} & \lambda_{64}\end{bmatrix}, \tag{32}$$

where the 1s and 0s are treated as known fixed parameters, and the $\lambda_{j,k}$ are unknown parameters given by $\lambda_{j,k}=1$ for $k=1,2$ and by $\lambda_{j,k}=1.5$ for $k=3,4$ as specified in Λ as true population values.

Consider the following asymmetric linear function and asymmetric logistic function:

$$f\left(t_{k,i}\right)=\left\{I\left(t_{i}\geq 0\right)+I\left(t_{i}<0\right)\right\}t_{k,i}\,, \tag{33}$$

$$f\left(t_{k,i}\right)=\left\{I\left(t_{i}\geq 0\right)+I\left(t_{i}<0\right)\right\}\left\{\frac{1}{1+\exp\left(t_{k,i}\right)}-\frac{1}{2}\right\}C\,. \tag{34}$$

where $C = 7$. CR′ and AVE′ for each measurement model are given by

$$\underset{\left\langle \begin{smallmatrix}asymmetric\\ \text{-}\,linear\end{smallmatrix}\right\rangle}{CR_{k}'}=\frac{Var\left(t_{k,i}\right)\tilde{W}}{Var\left(t_{k,i}\right)\tilde{W}+n\sum_{j=1}^{p}\psi_{j}}, \tag{35}$$

$$\underset{\left\langle \begin{smallmatrix}asymmetric\\ \text{-}\,linear\end{smallmatrix}\right\rangle}{AVE_{k}'}=\frac{Var\left(t_{k,i}\right)\bar{W}}{Var\left(t_{k,i}\right)\bar{W}+n\sum_{j=1}^{p}\psi_{j}}, \tag{36}$$

and

$$\underset{\left\langle\substack{asymmetric\\ -logistic}\right\rangle}{CR'_k} = \frac{\left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)\tilde{W}}{\left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)\tilde{W} + n\sum_{j=1}^{p}\psi_j}, \tag{37}$$

$$\underset{\left\langle\substack{asymmetric\\ -logistic}\right\rangle}{AVE'_k} = \frac{\left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)\bar{W}}{\left\{\frac{C\exp(0)}{\{1+\exp(0)\}^2}\right\}^2 Var\left(t_{k,i}\right)\bar{W} + n\sum_{j=1}^{p}\psi_j}, \tag{38}$$

where

$$\tilde{W} = \sum_{i=1}^{n}\left[\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i} \geq 0\right) + \lambda_{j,k+2}I\left(t_{k,i} < 0\right)\right\}\right]^2, \tag{39}$$

and

$$\bar{W} = \sum_{i=1}^{n}\sum_{j=1}^{p}\left\{\lambda_{j,k}I\left(t_{k,i} \geq 0\right) + \lambda_{j,k+2}I\left(t_{k,i} < 0\right)\right\}^2. \tag{40}$$

Table 4 shows the results of the asymmetric linear measurement model. Table 5 shows the results of estimates by the asymmetric logistic function defined in (34), and Table 6 shows the results by replacing C in function (34) in the same way as in study 1 with $z^* = 5.636$. $P(\mathrm{E})$ in the tables indicates the probability of event E; thus the relationship of asymmetry was estimated almost certainly. The results indicate that the biases of estimates by proposed method are close to 0 in all settings

Table 4: Results of the asymmetric linear function
Table 5: Results of the asymmetric logistic function
Table 6: Results of the asymmetric logistic function in practice

5. Empirical Analysis

We investigate nonlinear SERVQUAL model (PZB 1985; 1988; Figure 4) and its construct validation. SERVQUAL is a famous scale used in marketing to measure perceived service quality as the difference between consumers' expectation and actual perception (PZB 1985; 1988; 1993; 1994a; 1994b). Although a number of researchers conclude that the validity of SERVQUAL scale and model is not sufficient (e.g., Babakus & Boller 1992; Brown et al. 1993; Carman 1990; Cronin & Taylor 1992; 1994), they have discussed the validity under linear assumptions. Because consumers' perceived service quality follows a value function according to prospect theory (Kahneman &Tversky 1979; Sivakumar et al. 2014), it is reasonable to assume a nonlinear process in the measurement model for SERVQUAL.

The dataset (n = 300) was compiled from two companies in three industries through a Japanese research company. We estimate a linear measurement model with quadratic (QM), logistic (LGM), and their asymmetric measurement model (ALM, AQM, ALGM) by Bayesian estimation. To compare these models, we calculate WAIC (Watanabe 2010a; Watanabe 2010b; Gelman 2013) and WBIC (Watanabe 2013) shown in Tables 7 and 8, which represent information criteria for model selection in terms of prediction and logarithm of Bayes marginal likelihood, respectively. We also produce the logarithm of the Bayes factor (Lee 2007; Song & Lee 2012) in Table 9.

Figure 4: SERVQUAL model
Table 7: WAIC
Table 8: WBIC
Table 9: Logarithm of the Bayes factor (double scale)

WAIC and WBIC in Tables 7 and 8 select the same model in each company except Hotel B and Retail A. The bold and italic numbers in Table 9 show the acceptable model H1 compared with H0 and the best model (see also Lee 2007, p.114), respectively, in each company; thus the logarithm of the Bayes factor indicates that the most nonlinear measurement models are supported strongly in each company.

Table 10 and 11 report the estimated CR and AVE in each company. The bold and italic numbers show that the estimated CR and AVE are less than the criterion 0.7 for CR and 0.5 for AVE. The quadratic model is the best model in most companies; however, some estimated CR and AVE do not achieve the criterion. Moreover, the estimated CR and AVE tend to get worse compared with the linear model. On the contrary, we find that the logistic and asymmetric logistic model improves CR and AVE compared with the other models.

Table 10: CR (reliability coefficient)
Table 11: AVE (convergent validity)

Tables 12 to 17 report a judgment of discriminant validity in each company. In each lower triangular matrix, diagonal elements show estimated AVEs and nondiagonal elements show squared estimated correlations among five factors. The bold and italic numbers indicate that the nondiagonal element is lower than the diagonal element so that the squared correlation is lower than AVE, meaning insufficient discriminant validity. We find that discriminant validities are satisfied in the logistic and asymmetric logistic model, whereas the other model does not achieve sufficient validity, in almost all cases.

6. Concluding Remarks

In this paper, we discussed a construct validation for a nonlinear measurement model. Two indexes, CR′ and AVE′, were developed as an alternative to CR and AVE, which were introduced in marketing area by Fornell & Larcker (1981). Simulation studies showed the performance of these new indexes and the several illustrations to derivate CR′ and AVE′.

We also provided a reassessment of the validity of the SERVQUAL model proposed by PZB (1985; 1988) to measure perceived service quality in marketing research. Five nonlinear SERVQUAL models were investigated in empirical analyses, including the linear model. We found that the logistic and asymmetric logistic model are robust among all of the industries in terms of construct validity. Our results indicate that observed perceived service quality is associated nonlinearly and asymmetrically with latent true perceived service quality following the prospect theory (Kahneman &Tversky 1979; Sivakumar et al. 2014).

In future research, it might be possible to adopt the concept of construct validation to create interpretable machine learning with a latent variable such as a neural network model. Because the machine learning model, or the algorithm known as “Black Box” (Ribeiro et al. 2016a; 2016b), in many cases, results in a reasonable interpretation from these methods, it is an important task in the social science area (Park 2012). Construct validation has been discussed to provide a certain validity and interpretation of latent variables estimated by factor analysis as a measurement model with item scales. We believe that construct validation connects the knowledge of establishing a model between social science and machine learning in terms of better prediction with reasonable interpretation.

Figures and Tables

Figure 1: Three different measurement equations

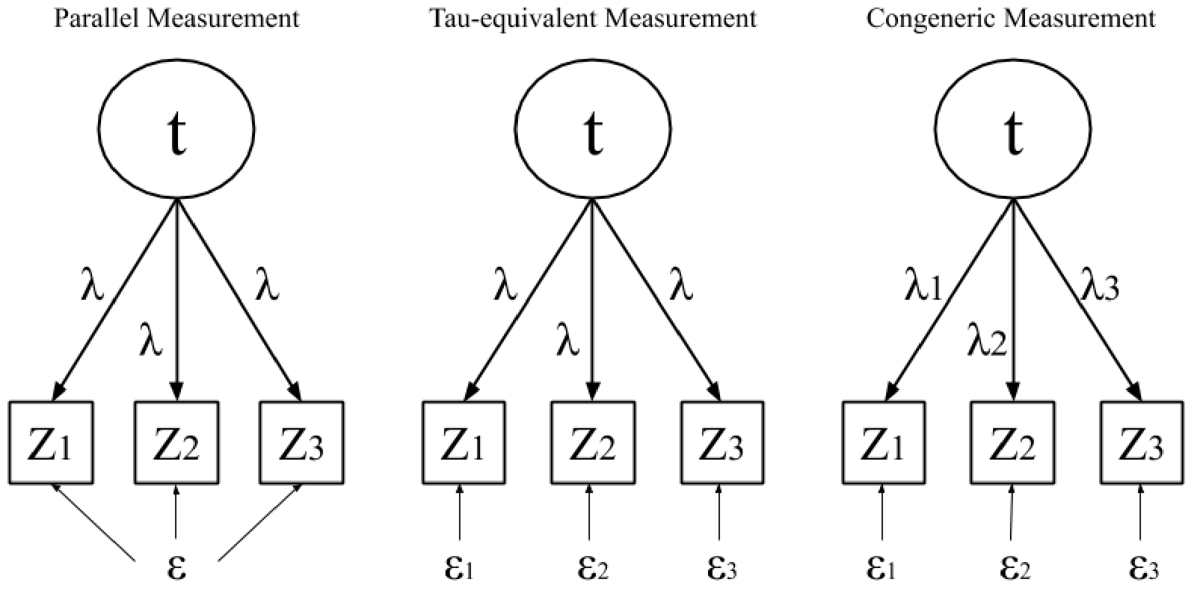

Figure 2: Reflective and formative models

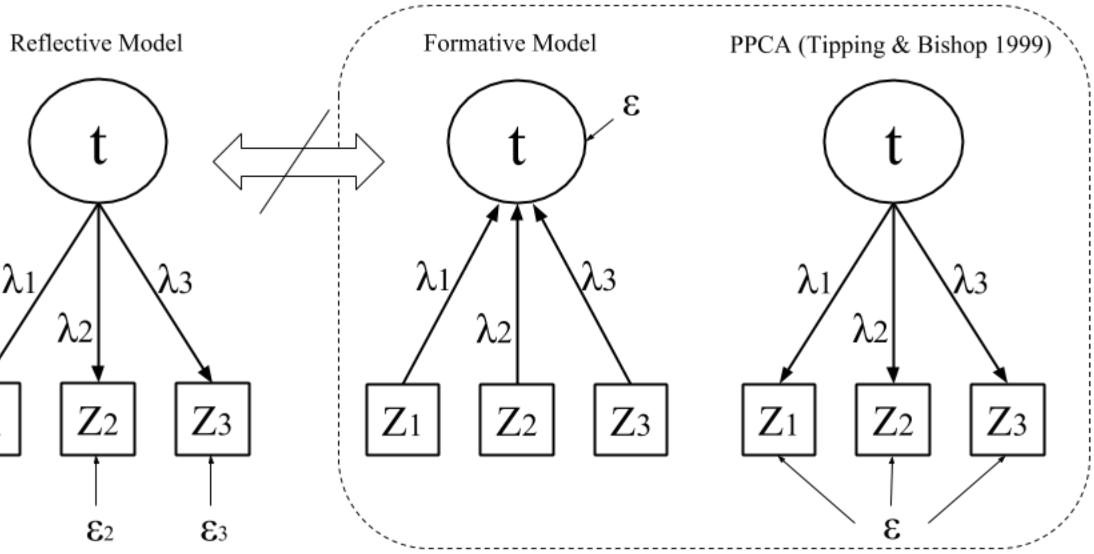

Figure 3: The problem of a small factor lading and a large correlation

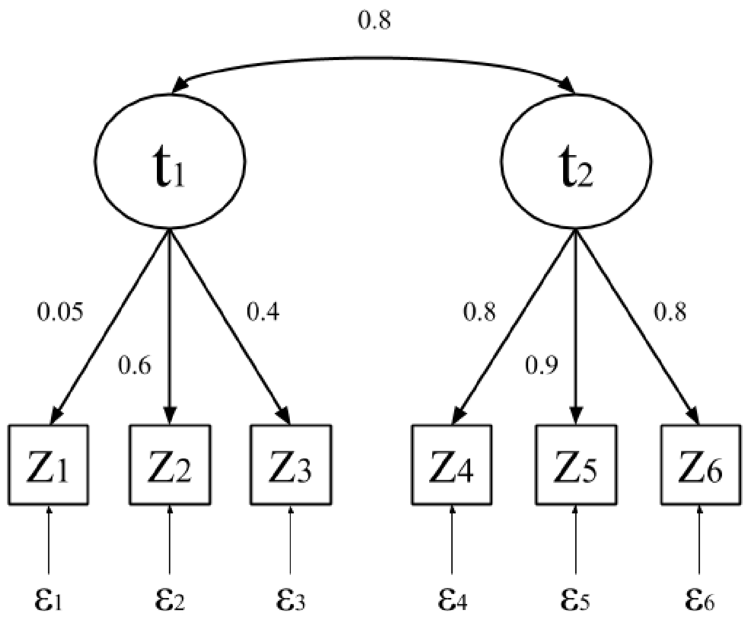

Figure 4: SERVQUAL model

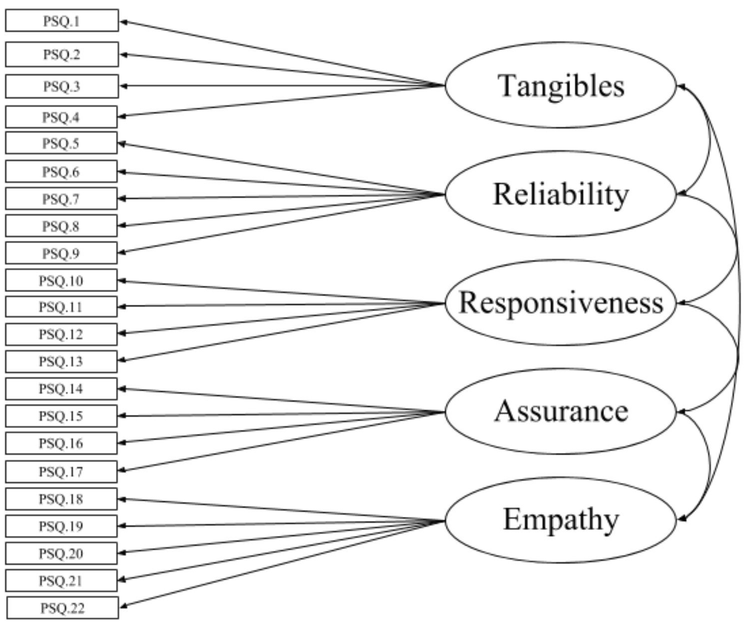

Table 1: Results of the logistic function

Logistic	Setting	Bias	SE	95%HPDI				
psi1	1.500	0.025	0.176	[	-0.293	,	0.396	]
psi2	1.500	-0.183	0.192	[	-0.525	,	0.208	]
psi3	1.500	0.168	0.211	[	-0.203	,	0.600	]
psi4	1.500	0.052	0.179	[	-0.289	,	0.404	]
psi5	1.500	0.075	0.201	[	-0.300	,	0.502	]
psi6	1.500	-0.052	0.198	[	-0.398	,	0.348	]
lam2	1.000	0.028	0.082	[	-0.125	,	0.184	]
lam3	1.000	0.035	0.083	[	-0.107	,	0.211	]
lam5	1.000	0.096	0.081	[	-0.063	,	0.254	]
lam6	1.000	0.059	0.087	[	-0.100	,	0.228	]
Phi11	1.000	-0.076	0.141	[	-0.320	,	0.197	]
Phi22	1.000	-0.109	0.134	[	-0.354	,	0.145	]
Phi12	0.500	-0.053	0.074	[	-0.186	,	0.088	]
CR'1	0.860	-0.007	0.017	[	-0.041	,	0.023	]
CR'2	0.860	-0.006	0.016	[	-0.035	,	0.029	]
AVE'1	0.671	-0.011	0.030	[	-0.069	,	0.043	]
AVE'2	0.671	-0.009	0.029	[	-0.059	,	0.056	]

Table 2: Results of the logistic function in practice

Logistic2	Setting	std	Bias	SE	95%HPDI				
psi1	1.500	0.329	0.012	0.045	[	-0.066	,	0.105	]
psi2	1.500	0.329	-0.053	0.048	[	-0.140	,	0.041	]
psi3	1.500	0.329	0.016	0.052	[	-0.085	,	0.120	]
psi4	1.500	0.329	0.007	0.054	[	-0.088	,	0.111	]
psi5	1.500	0.329	-0.030	0.042	[	-0.115	,	0.053	]
psi6	1.500	0.329	-0.037	0.041	[	-0.116	,	0.042	]
lam11	1.000	0.819	-0.008	0.028	[	-0.063	,	0.043	]
lam21	1.000	0.819	0.031	0.028	[	-0.025	,	0.081	]
lam31	1.000	0.819	-0.011	0.032	[	-0.077	,	0.050	]
lam42	1.000	0.819	-0.005	0.034	[	-0.071	,	0.052	]
lam52	1.000	0.819	0.018	0.025	[	-0.033	,	0.067	]
lam62	1.000	0.819	0.022	0.025	[	-0.026	,	0.068	]
Phi12	0.500	0.500	0.004	0.056	[	-0.108	,	0.108	]
CR'1	0.860	0.860	0.004	0.016	[	-0.028	,	0.033	]
CR'2	0.860	0.860	0.010	0.016	[	-0.019	,	0.042	]
AVE'1	0.671	0.671	0.008	0.030	[	-0.049	,	0.063	]
AVE'2	0.671	0.671	0.020	0.030	[	-0.036	,	0.079	]

Table 3: Results of the quadratic function

Quadratic	Setting	Bias	SE	95%HPDI				
psi1	1.500	-0.160	0.153	[	-0.457	,	0.153	]
psi2	1.500	-0.038	0.149	[	-0.313	,	0.243	]
psi3	1.500	0.178	0.182	[	-0.135	,	0.553	]
psi4	1.500	0.057	0.175	[	-0.300	,	0.387	]
psi5	1.500	0.070	0.166	[	-0.258	,	0.377	]
psi6	1.500	-0.031	0.153	[	-0.322	,	0.255	]
lam12	1.000	-0.094	0.057	[	-0.208	,	0.012	]
lam13	1.000	-0.017	0.068	[	-0.148	,	0.112	]
lam25	1.000	0.067	0.067	[	-0.052	,	0.203	]
lam26	1.000	0.031	0.067	[	-0.107	,	0.151	]
Phi11	1.000	0.026	0.100	[	-0.183	,	0.195	]
Phi22	1.000	0.012	0.093	[	-0.165	,	0.197	]
Phi12	0.500	0.062	0.075	[	-0.078	,	0.209	]
CR'1	0.800	-0.006	0.031	[	-0.073	,	0.050	]
CR'2	0.800	0.008	0.029	[	-0.044	,	0.068	]
AVE'1	0.571	-0.007	0.046	[	-0.110	,	0.074	]
AVE'2	0.571	0.014	0.045	[	-0.072	,	0.104	]

Table 4: Results of the asymmetric linear function

A-L	Setting	Bias	SE	95%HPDI				
psi1	1.500	-0.220	0.182	[	-0.548	,	0.169	]
psi2	1.500	0.222	0.182	[	-0.164	,	0.551	]
psi3	1.500	0.142	0.188	[	-0.197	,	0.557	]
psi4	1.500	-0.140	0.169	[	-0.475	,	0.159	]
psi5	1.500	-0.095	0.197	[	-0.452	,	0.276	]
psi6	1.500	0.104	0.166	[	-0.200	,	0.439	]
lam21	1.000	0.153	0.178	[	-0.174	,	0.505	]
lam31	1.000	0.105	0.182	[	-0.257	,	0.450	]
lam52	1.000	0.343	0.241	[	-0.055	,	0.834	]
lam62	1.000	0.042	0.200	[	-0.339	,	0.436	]
lam13	1.500	0.192	0.237	[	-0.306	,	0.575	]
lam23	1.500	-0.029	0.233	[	-0.440	,	0.444	]
lam33	1.500	-0.273	0.213	[	-0.681	,	0.109	]
lam44	1.500	-0.170	0.235	[	-0.558	,	0.348	]
lam54	1.500	0.084	0.296	[	-0.428	,	0.648	]
lam64	1.500	-0.162	0.256	[	-0.642	,	0.318	]
Phi11	1.000	-0.150	0.211	[	-0.467	,	0.278	]
Phi22	1.000	-0.164	0.236	[	-0.583	,	0.291	]
Phi12	0.500	-0.183	0.085	[	-0.341	,	-0.020	]
CR'1	0.766	-0.041	0.028	[	-0.093	,	0.017	]
CR'2	0.763	-0.035	0.026	[	-0.086	,	0.012	]
AVE'1	0.521	-0.048	0.035	[	-0.118	,	0.019	]
AVE'2	0.518	-0.041	0.032	[	-0.098	,	0.024	]

E	P(E)
lam11 < lam13	1.000
lam21 < lam23	0.907
lam31 < lam33	0.719
lam42 < lam44	0.937
lam52 < lam54	0.860
lam62 < lam64	0.913

Table 5: Results of the asymmetric logistic function

A-LG1	Setting	Bias	SE	95%HPDI				
psi1	1.500	-0.045	0.179	[	-0.367	,	0.328	]
psi2	1.500	-0.095	0.181	[	-0.404	,	0.290	]
psi3	1.500	0.167	0.189	[	-0.238	,	0.511	]
psi4	1.500	0.070	0.174	[	-0.243	,	0.435	]
psi5	1.500	0.097	0.190	[	-0.272	,	0.482	]
psi6	1.500	-0.095	0.168	[	-0.411	,	0.233	]
lam21	1.000	0.038	0.100	[	-0.139	,	0.255	]
lam31	1.000	0.103	0.106	[	-0.089	,	0.312	]
lam52	1.000	0.052	0.099	[	-0.124	,	0.247	]
lam62	1.000	0.154	0.099	[	-0.050	,	0.331	]
lam13	1.500	0.164	0.140	[	-0.093	,	0.443	]
lam23	1.500	0.086	0.131	[	-0.148	,	0.347	]
lam33	1.500	0.095	0.139	[	-0.161	,	0.371	]
lam44	1.500	-0.103	0.123	[	-0.340	,	0.134	]
lam54	1.500	0.070	0.133	[	-0.174	,	0.341	]
lam64	1.500	-0.133	0.123	[	-0.367	,	0.103	]
Phi11	1.000	-0.165	0.147	[	-0.440	,	0.122	]
Phi22	1.000	-0.005	0.193	[	-0.333	,	0.389	]
Phi12	0.500	-0.064	0.078	[	-0.204	,	0.101	]
CR'1	0.907	-0.005	0.012	[	-0.028	,	0.018	]
CR'2	0.907	-0.004	0.012	[	-0.028	,	0.018	]
AVE'1	0.764	-0.010	0.025	[	-0.055	,	0.041	]
AVE'2	0.765	-0.008	0.025	[	-0.057	,	0.041	]

E	P(E)
lam11 < lam13	1.000
lam21 < lam23	1.000
lam31 < lam33	1.000
lam42 < lam44	1.000
lam52 < lam54	1.000
lam62 < lam64	0.966

Table 6: Results of the asymmetric logistic function in practice

A-LG2	Setting	std	Bias	SE	95%HPDI				
psi1	1.500	0.131	0.005	0.023	[	-0.042	,	0.047	]
psi2	1.500	0.131	-0.010	0.021	[	-0.046	,	0.035	]
psi3	1.500	0.131	0.004	0.023	[	-0.036	,	0.052	]
psi4	1.500	0.131	0.010	0.023	[	-0.029	,	0.057	]
psi5	1.500	0.131	-0.020	0.019	[	-0.058	,	0.012	]
psi6	1.500	0.131	-0.021	0.018	[	-0.053	,	0.013	]
lam11	1.000	0.517	-0.068	0.029	[	-0.123	,	-0.011	]
lam21	1.000	0.517	-0.004	0.036	[	-0.072	,	0.068	]
lam31	1.000	0.517	0.015	0.042	[	-0.063	,	0.101	]
lam42	1.000	0.517	0.011	0.026	[	-0.041	,	0.059	]
lam52	1.000	0.517	0.010	0.035	[	-0.054	,	0.075	]
lam62	1.000	0.517	0.096	0.034	[	0.032	,	0.158	]
lam13	1.500	0.776	0.037	0.020	[	-0.007	,	0.072	]
lam23	1.500	0.776	0.008	0.027	[	-0.040	,	0.060	]
lam33	1.500	0.776	-0.014	0.028	[	-0.065	,	0.044	]
lam44	1.500	0.776	-0.015	0.022	[	-0.057	,	0.026	]
lam54	1.500	0.776	0.005	0.024	[	-0.043	,	0.050	]
lam64	1.500	0.776	-0.060	0.028	[	-0.112	,	-0.002	]
Phi12	0.500	0.500	-0.005	0.055	[	-0.116	,	0.091	]
CR'1	0.907	0.907	0.003	0.012	[	-0.021	,	0.027	]
CR'2	0.907	0.907	0.005	0.011	[	-0.018	,	0.026	]
AVE'1	0.764	0.764	0.007	0.026	[	-0.052	,	0.052	]
AVE'2	0.765	0.765	0.012	0.024	[	-0.036	,	0.058	]

E	*P* (E)
lam11 < lam13	1.000
lam21 < lam23	1.000
lam31 < lam33	1.000
lam42 < lam44	1.000
lam52 < lam54	1.000
lam62 < lam64	0.964

Table 7: WAIC

WAIC	original	QM	LGM	ALM	AQM	ALGM	result
Hotel B	14,000.70	13,864.31	13881.07	14019.27	13949.86	13948.76	QM
Hotel A	13,536.59	13494.81	13,438.16	13,546.74	13,501.80	13,499.15	LGM
Bank B	14,366.11	13,085.80	14,282.70	14,393.41	14,115.11	14,339.73	QM
Bank A	14,607.09	13,510.48	14,561.77	14,687.57	13,718.13	14,657.97	QM
Retail B	14,321.25	11,849.23	14,292.65	14,336.49	14,193.31	14,349.40	QM
Retail A	13,603.49	13,375.52	13,495.42	13,623.07	13,418.68	13,588.92	QM

Table 8: WBIC

WBIC	original	QM	LGM	ALM	AQM	ALGM	result
Hotel B	6,623.40	6,590.95	6,555.86	6,625.20	6,600.61	6,574.88	LGM
Hotel A	6,410.11	6394.379	6,373.68	6,420.75	6,416.31	6,383.15	LGM
Bank B	6,801.78	6,241.22	6740.022	6,818.56	6,706.92	6,783.17	QM
Bank A	6,928.36	6,442.77	6,877.27	6,903.85	6,511.96	6,875.95	QM
Retail B	6,772.41	5,607.02	6,745.82	6,758.65	6,744.35	6,769.94	QM
Retail A	6,466.98	6,385.94	6,399.78	6,444.04	6,379.74	6,420.63	AQM

Table 9: Logarithm of the Bayes factor (double scale)

H1 \ H0	Original	QM	LGM	ALM	AQM	Original	QM	LGM	ALM	AQM
			Hotel B					Hotel A		
QM	**64.90**					**31.47**				
LGM	***135.07***	70.17				***72.88***	41.41			
ALM	-3.60	-68.51	-138.68			-21.28	-52.75	-94.15		
AQM	**45.58**	-19.32	-89.49	49.18		-12.40	-43.87	-85.28	8.88	
ALGM	**97.05**	32.15	-38.02	100.65	51.47	**53.93**	22.46	-18.94	75.21	66.33
			Bank B					Bank A		
QM	***1,121.11***					***971.18***				
LGM	**123.51**	-997.60				**102.17**	-869.01			
ALM	-33.58	-1,154.69	-157.08			**49.01**	-922.17	-53.15		
AQM	**189.71**	-931.40	66.21	223.29		**832.79**	-138.39	730.62	783.77	
ALGM	**37.20**	-1,083.91	-86.30	70.78	-152.51	**104.81**	-866.37	2.64	55.79	-727.98
			Retail B					Retail A		
QM	***2,330.79***					**162.06**				
LGM	**53.19**	-2,277.60				**134.39**	-27.67			
ALM	**27.53**	-2,303.26	-25.67			**45.87**	-116.20	-88.53		
AQM	**56.13**	-2,274.66	2.94	28.61		***174.47***	12.41	40.08	128.61	
ALGM	**4.94**	-2,325.85	-48.25	-22.59	-51.19	**92.68**	-69.38	-41.71	46.82	-81.79

Table 10: CR (reliability coefficient)

CR	original	QM	LGM	ALM	AQM	ALGM	original	QM	LGM	ALM	AQM	ALGM
			Hotel B						Hotel A			
Tangibles	0.732	***0.680***	0.770	0.752	***0.693***	0.781	0.739	***0.685***	0.772	0.745	0.705	0.774
Reliability	0.733	***0.646***	0.771	0.731	***0.650***	0.772	0.826	0.768	0.849	0.825	0.772	0.849
Responsiveness	0.793	0.746	0.821	0.798	0.749	0.828	0.857	0.806	0.876	0.850	0.811	0.878
Assurance	0.757	***0.684***	0.792	0.760	***0.684***	0.797	0.848	0.799	0.871	0.849	0.805	0.869
Empathy	0.861	0.822	0.874	0.862	0.823	0.879	0.863	0.822	0.883	0.870	0.841	0.886
			Bank A						Bank A			
Tangibles	0.735	***0.684***	0.763	0.741	***0.681***	0.769	0.821	0.731	0.842	0.821	0.740	0.845
Reliability	***0.695***	***0.606***	0.745	***0.699***	***0.593***	0.740	0.774	***0.672***	0.813	0.773	***0.692***	0.815
Responsiveness	0.763	***0.665***	0.803	0.758	***0.659***	0.792	0.852	0.735	0.881	0.854	0.744	0.883
Assurance	0.709	***0.601***	0.736	0.704	***0.642***	0.745	0.802	0.739	0.854	0.828	0.761	0.859
Empathy	0.813	0.723	0.841	0.814	0.727	0.836	0.882	0.780	0.897	0.878	0.798	0.899
			Retail A						Retail A			
Tangibles	0.732	***0.638***	0.764	0.742	***0.689***	0.764	0.764	***0.683***	0.799	0.753	***0.694***	0.786
Reliability	0.771	***0.698***	0.797	0.762	***0.691***	0.789	0.810	0.762	0.836	0.812	0.765	0.837
Responsiveness	0.737	***0.674***	0.782	0.735	***0.667***	0.773	0.808	0.742	0.839	0.805	0.740	0.835
Assurance	0.745	***0.676***	0.783	0.759	***0.669***	0.788	0.833	0.760	0.858	0.833	0.768	0.861
Empathy	0.802	0.753	0.836	0.817	0.756	0.839	0.858	0.813	0.879	0.865	0.826	0.885

Table 11: AVE (convergent validity)

AVE	original	QM	LGM	ALM	AQM	ALGM	original	QM	LGM	ALM	AQM	ALGM
			Hotel B						Hotel A			
Tangibles	***0.418***	***0.368***	***0.477***	***0.451***	***0.368***	***0.493***	***0.418***	***0.357***	***0.464***	***0.429***	***0.380***	***0.468***
Reliability	***0.360***	***0.273***	***0.409***	***0.360***	***0.276***	***0.413***	***0.492***	***0.406***	0.534	***0.492***	***0.412***	0.536
Responsiveness	***0.492***	***0.428***	0.538	0.504	***0.436***	0.556	0.603	0.514	0.641	0.592	0.522	0.647
Assurance	***0.443***	***0.357***	***0.499***	***0.449***	***0.356***	0.504	0.587	0.508	0.636	0.593	0.517	0.634
Empathy	0.558	***0.486***	0.584	0.560	***0.489***	0.597	0.563	***0.490***	0.608	0.582	0.523	0.618
			Bank B						Bank A			
Tangibles	***0.415***	***0.364***	***0.457***	***0.432***	***0.360***	***0.470***	0.536	***0.410***	0.573	0.538	***0.423***	0.581
Reliability	***0.321***	***0.250***	***0.380***	***0.331***	***0.232***	***0.379***	***0.410***	***0.297***	***0.469***	***0.416***	***0.321***	***0.479***
Responsiveness	***0.453***	***0.341***	0.511	***0.449***	***0.331***	***0.497***	0.592	***0.413***	0.652	0.597	***0.426***	0.658
Assurance	***0.391***	***0.310***	***0.428***	***0.392***	***0.326***	***0.444***	0.528	***0.454***	0.626	0.581	***0.481***	0.638
Empathy	***0.475***	***0.359***	0.525	***0.480***	***0.355***	0.520	0.606	***0.423***	0.642	0.598	***0.451***	0.647
			Retail B						Retail A			
Tangibles	***0.435***	***0.367***	***0.484***	***0.457***	***0.373***	***0.486***	***0.453***	***0.357***	0.501	***0.439***	***0.367***	***0.484***
Reliability	***0.405***	***0.320***	***0.444***	***0.400***	***0.314***	***0.439***	***0.464***	***0.397***	0.512	***0.473***	***0.403***	0.518
Responsiveness	***0.420***	***0.352***	***0.480***	***0.432***	***0.350***	***0.482***	0.515	***0.422***	0.568	0.511	***0.420***	0.565
Assurance	***0.428***	***0.356***	***0.487***	***0.454***	***0.345***	***0.495***	0.558	***0.449***	0.610	0.562	***0.460***	0.616
Empathy	***0.464***	***0.401***	0.527	***0.493***	***0.395***	0.533	0.552	***0.471***	0.596	0.567	***0.491***	0.611

Table 12: Discriminant validity in Hotel B

Hotel B	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.418				
Reliability	0.244	0.360			
Responsiveness	0.323	***0.423***	0.492		
Assurance	0.316	***0.380***	***0.666***	0.443	
Empathy	0.252	0.250	0.396	0.433	0.558
QM					
Tangibles	0.368				
Reliability	0.254	0.273			
Responsiveness	***0.325***	***0.446***	0.428		
Assurance	0.319	***0.457***	***0.691***	0.357	
Empathy	0.260	0.227	0.329	***0.418***	0.486
LGM					
Tangibles	0.477				
Reliability	0.161	0.409			
Responsiveness	0.232	0.276	0.538		
Assurance	0.221	0.244	0.438	0.499	
Empathy	0.190	0.173	0.297	0.308	0.584
ALM					
Tangibles	0.451				
Reliability	0.311	0.360			
Responsiveness	0.374	***0.442***	0.504		
Assurance	0.327	***0.422***	***0.687***	0.449	
Empathy	0.258	0.277	0.394	0.434	0.560
AQM					
Tangibles	0.368				
Reliability	***0.356***	0.276			
Responsiveness	***0.383***	***0.478***	0.436		
Assurance	0.322	***0.500***	***0.701***	0.356	
Empathy	0.257	0.273	0.352	***0.435***	0.489
ALGM					
Tangibles	0.493				
Reliability	0.233	0.413			
Responsiveness	0.303	0.344	0.556		
Assurance	0.256	0.313	0.518	0.504	
Empathy	0.205	0.216	0.324	0.339	0.597

Table 13: Discriminant validity in Hotel A

Hotel A	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.418				
Reliability	***0.419***	0.492			
Responsiveness	0.357	***0.639***	0.603		
Assurance	0.354	***0.525***	***0.711***	0.587	
Empathy	0.267	***0.521***	0.522	0.513	0.563
QM					
Tangibles	0.357				
Reliability	***0.446***	0.406			
Responsiveness	***0.351***	***0.624***	0.514		
Assurance	***0.353***	***0.501***	***0.715***	0.508	
Empathy	0.237	***0.451***	0.439	0.431	0.490
LGM					
Tangibles	0.464				
Reliability	0.340	0.534			
Responsiveness	0.313	***0.557***	0.641		
Assurance	0.296	0.473	***0.650***	0.636	
Empathy	0.235	0.464	0.488	0.481	0.608
ALM					
Tangibles	0.429				
Reliability	***0.456***	0.492			
Responsiveness	0.375	***0.651***	0.592		
Assurance	0.361	***0.534***	***0.710***	0.593	
Empathy	0.261	***0.542***	0.543	0.526	0.582
AQM					
Tangibles	0.380				
Reliability	***0.468***	0.412			
Responsiveness	0.367	***0.615***	0.522		
Assurance	***0.374***	***0.494***	***0.712***	0.517	
Empathy	0.250	***0.497***	0.500	0.479	0.523
ALGM					
Tangibles	0.468				
Reliability	0.386	0.536			
Responsiveness	0.362	***0.593***	0.647		
Assurance	0.342	0.484	***0.655***	0.634	
Empathy	0.259	0.486	0.507	0.475	0.618

Table 14: Discriminant validity in Bank B

Bank B	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.415				
Reliability	0.264	0.321			
Responsiveness	0.077	***0.389***	0.453		
Assurance	0.066	***0.371***	***0.489***	0.391	
Empathy	0.073	0.258	***0.456***	0.298	0.475
QM					
Tangibles	0.364				
Reliability	0.217	0.250			
Responsiveness	0.052	***0.336***	0.341		
Assurance	0.017	***0.257***	***0.361***	0.310	
Empathy	0.080	0.213	***0.375***	0.167	0.359
LGM					
Tangibles	0.457				
Reliability	0.193	0.380			
Responsiveness	0.061	0.253	0.511		
Assurance	0.051	0.242	0.346	0.428	
Empathy	0.059	0.183	0.322	0.223	0.525
ALM					
Tangibles	0.432				
Reliability	0.313	0.331			
Responsiveness	0.088	***0.393***	0.449		
Assurance	0.069	***0.348***	***0.505***	0.392	
Empathy	0.075	0.249	***0.466***	0.306	0.480
AQM					
Tangibles	0.360				
Reliability	0.324	0.232			
Responsiveness	0.080	***0.360***	0.331		
Assurance	0.071	***0.381***	***0.543***	0.326	
Empathy	0.080	***0.222***	***0.463***	0.327	0.355
ALGM					
Tangibles	0.470				
Reliability	0.222	0.379			
Responsiveness	0.075	0.272	0.497		
Assurance	0.056	0.260	0.359	0.444	
Empathy	0.063	0.183	0.350	0.239	0.520

Table 15: Discriminant validity in Bank A

Bank A	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.536				
Reliability	0.523	0.410			
Responsiveness	0.334	***0.644***	0.592		
Assurance	0.333	***0.589***	***0.691***	0.528	
Empathy	0.238	***0.466***	***0.643***	***0.551***	0.606
QM					
Tangibles	0.410				
Reliability	***0.519***	0.297			
Responsiveness	0.307	***0.525***	0.413		
Assurance	0.192	***0.378***	***0.417***	0.454	
Empathy	0.174	0.284	***0.485***	0.294	0.423
LGM					
Tangibles	0.573				
Reliability	0.407	0.469			
Responsiveness	0.277	***0.519***	0.652		
Assurance	0.283	***0.498***	0.596	0.626	
Empathy	0.209	0.414	0.574	0.514	0.642
ALM					
Tangibles	0.538				
Reliability	0.519	0.416			
Responsiveness	0.356	***0.666***	0.597		
Assurance	0.351	***0.600***	***0.715***	0.581	
Empathy	0.250	***0.478***	***0.665***	***0.567***	0.598
AQM					
Tangibles	0.423				
Reliability	***0.515***	0.321			
Responsiveness	0.325	***0.519***	0.426		
Assurance	0.215	***0.398***	0.424	0.481	
Empathy	0.176	***0.294***	***0.510***	0.300	0.451
ALGM					
Tangibles	0.581				
Reliability	0.444	0.479			
Responsiveness	0.308	***0.555***	0.658		
Assurance	0.317	***0.532***	0.638	0.638	
Empathy	0.223	0.430	0.604	0.537	0.647

Table 16: Discriminant validity in Retail B

Retail B	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.435				
Reliability	0.111	0.405			
Responsiveness	0.126	0.287	0.420		
Assurance	0.216	0.157	***0.668***	0.428	
Empathy	0.170	0.145	0.380	0.395	0.464
QM					
Tangibles	0.367				
Reliability	0.033	0.320			
Responsiveness	0.084	0.365	0.352		
Assurance	0.105	0.245	***0.707***	0.356	
Empathy	0.143	0.118	0.317	***0.376***	0.401
LGM					
Tangibles	0.484				
Reliability	0.095	0.444			
Responsiveness	0.102	0.196	0.480		
Assurance	0.162	0.120	0.435	0.487	
Empathy	0.137	0.117	0.273	0.294	0.527
ALM					
Tangibles	0.457				
Reliability	0.151	0.400			
Responsiveness	0.141	0.273	0.432		
Assurance	0.222	0.184	***0.731***	0.454	
Empathy	0.189	0.174	0.402	0.417	0.493
AQM					
Tangibles	0.373				
Reliability	0.182	0.314			
Responsiveness	0.164	0.306	0.350		
Assurance	0.209	0.210	***0.687***	0.345	
Empathy	0.235	0.136	***0.392***	***0.415***	0.395
ALGM					
Tangibles	0.486				
Reliability	0.124	0.439			
Responsiveness	0.130	0.203	0.482		
Assurance	0.186	0.146	***0.534***	0.495	
Empathy	0.161	0.134	0.316	0.324	0.533

Table 17: Discriminant validity in Retail A

Retail.A	Tangibles	Reliability	Responsiveness	Assurance	Empathy
original					
Tangibles	0.453				
Reliability	***0.463***	0.464			
Responsiveness	0.303	***0.614***	0.515		
Assurance	0.318	***0.610***	***0.715***	0.558	
Empathy	0.149	0.327	0.482	***0.570***	0.552
QM					
Tangibles	0.357				
Reliability	***0.506***	0.397			
Responsiveness	0.333	***0.634***	0.422		
Assurance	0.346	***0.612***	***0.697***	0.449	
Empathy	0.125	0.273	0.408	***0.502***	0.471
LGM					
Tangibles	0.501				
Reliability	0.331	0.512			
Responsiveness	0.228	0.477	0.568		
Assurance	0.235	0.477	0.552	0.610	
Empathy	0.117	0.265	0.376	0.440	0.596
ALM					
Tangibles	0.439				
Reliability	***0.466***	0.473			
Responsiveness	0.315	***0.600***	0.511		
Assurance	0.339	***0.598***	***0.705***	0.562	
Empathy	0.171	0.349	0.498	***0.588***	0.567
AQM					
Tangibles	0.367				
Reliability	***0.526***	0.403			
Responsiveness	0.354	***0.620***	0.420		
Assurance	0.378	***0.602***	***0.691***	0.460	
Empathy	0.167	0.327	***0.462***	***0.546***	0.491
ALGM					
Tangibles	0.484				
Reliability	0.124	0.518			
Responsiveness	0.130	0.203	0.565		
Assurance	0.186	0.146	0.534	0.616	
Empathy	0.161	0.134	0.316	0.324	0.611

Appendix

A. Relationship Between Measurement Model and Reliability Coefficient

1. Coefficient alpha and tau-equivalent test

Consider a composite measure for the tau-equivalent measurement model as follows:

$$Z=\sum_{j=1}^{p} z_{ji}=\sum_{j=1}^{p}\left(\lambda t_i+\varepsilon_{ji}\right)=p\lambda t_i+\sum_{j=1}^{p}\varepsilon_{ji}=T+E. \qquad \text{(A.1)}$$

Hence,

$$\begin{aligned}\alpha_t^{*}&=\frac{Var(T)}{Var(Z)}=\frac{Var(T)}{Var(T+E)}=\frac{Var(p\lambda t_i)}{Var\left(p\lambda t_i+\sum_{j=1}^{p}\varepsilon_{ji}\right)}\\&=\frac{p^2\lambda^2 Var(t_i)}{p^2\lambda^2 Var(t_i)+\sum_{j=1}^{p}Var(\varepsilon_{ji})}=\frac{p^2\lambda^2}{p^2\lambda^2+\sum_{j=1}^{p}\psi_{ji}}.\end{aligned} \qquad \text{(A.2)}$$

Coefficient alpha can be expressed as the following equation, assuming the tau-equivalent test:

$$\begin{aligned}\alpha_t&=\frac{p}{p-1}\left[1-\frac{\sum_{j=1}^{p}Var(z_{ji})}{Var(Z)}\right]=\frac{p}{p-1}\left[1-\frac{\sum_{j=1}^{p}Var(\lambda t_i+\varepsilon_{ji})}{Var\left(p\lambda t_i+\sum_{j=1}^{p}\varepsilon_{ji}\right)}\right]\\&=\frac{p}{p-1}\left[1-\frac{p\lambda^2+\sum_{j=1}^{p}\psi_j}{p^2\lambda^2+\sum_{j=1}^{p}\psi_j}\right]\\&=\frac{p}{p-1}\left[\frac{p^2\lambda^2+\sum_{j=1}^{p}\psi_j-p\lambda^2-\sum_{j=1}^{p}\psi_j}{p^2\lambda^2+\sum_{j=1}^{p}\psi_j}\right]\\&=\frac{p}{p-1}\left[\frac{p\lambda^2(p-1)}{p^2\lambda^2+\sum_{j=1}^{p}\psi_j}\right]=\frac{p^2\lambda^2}{p^2\lambda^2+\sum_{j=1}^{p}\psi_j}=\alpha_t^{*}.\end{aligned} \qquad \text{(A.3)}$$

2. Coefficient omega/CR and congeneric test

Consider a composite measure for the following congeneric measurement model:

$$Z=\sum_{j=1}^{p} z_{ji}=\sum_{j=1}^{p}\left(\lambda_j t_i+\varepsilon_{ji}\right)=\sum_{j=1}^{p}\lambda_j t_i+\sum_{j=1}^{p}\varepsilon_{ji}=T+E. \qquad \text{(A.4)}$$

Hence,

$$\begin{aligned}\omega_t^{*}&=\frac{Var(T)}{Var(Z)}=\frac{Var(T)}{Var(T+E)}=\frac{Var\left(\sum_{j=1}^{p}\lambda_j t_i\right)}{Var\left(\sum_{j=1}^{p}\lambda_j t_i+\sum_{j=1}^{p}\varepsilon_{ji}\right)}\\&=\frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2 Var(t_i)}{\left(\sum_{j=1}^{p}\lambda_j\right)^2 Var(t_i)+\sum_{j=1}^{p}Var(\varepsilon_{ji})}=\frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2}{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j}\end{aligned} \qquad \text{(A.5)}$$

Coefficient omega can be expressed as the following equation, assuming a congeneric test:

$$\begin{aligned}\omega_t &= 1-\frac{\sum_{j=1}^{p}\psi_j}{Var(Z)} = 1-\frac{\sum_{j=1}^{p}\psi_j}{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j} \\ &= \frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j-\sum_{j=1}^{p}\psi_j}{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j} = \frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2}{\left(\sum_{j=1}^{p}\lambda_j\right)^2+\sum_{j=1}^{p}\psi_j} = \omega_t^{*}.\end{aligned} \tag{A.6}$$

B. Proposed Estimators for CR and AVE in the Nonlinear Measurement Model
Consider a composite measure for the following nonlinear measurement model:

$$Z=\sum_{j=1}^{p} z_{ji}=\sum_{j=1}^{p}\left(\lambda_j f(t_i)+\varepsilon_{ji}\right)=\sum_{j=1}^{p}\lambda_j f(t_i)+\sum_{j=1}^{p}\varepsilon_{ji}=T+E. \tag{B.1}$$

In the same way as for (A.5), reliability in the nonlinear measurement model is given by

$$CR' = \frac{Var(T)}{Var(Z)} = \frac{Var(T)}{Var(T+E)} = \frac{\left(\sum_{j=1}^{p}\lambda_j\right)^2 Var\{f(t_i)\}}{\left(\sum_{j=1}^{p}\lambda_j\right)^2 Var\{f(t_i)\}+\sum_{j=1}^{p}Var(\varepsilon_{ji})} \tag{B.2}$$

In practice, it is necessary to evaluate the $Var\{f(t_i)\}$ from the estimated variance of t_i. Therefore, adopting a linear Taylor series approximation with $E(t_i)$ as expansion point (see Green 2011, Serfling 1980), we obtain

$$\begin{aligned} f(t_i) &\approx f(E(t_i))+\frac{f'(E(t_i))}{1!}(t_i-E(t_i))+ higher-order\ terms \\ &\approx \left[f(E(t_i))-f'(E(t_i))E(t_i)\right]+f'(E(t_i))t_i. \end{aligned} \tag{B.3}$$

Then,

$$\begin{aligned} Var(f(t_i)) &= \int_{t_i}\{f(t_i)-E(f(t_i))\}^2 p(t_i)dt_i \\ &\approx Var\left\{\left[f(E(t_i))-f'(E(t_i))E(t_i)\right]+f'(E(t_i))t_i\right\} \\ &= \{f'(E(t_i))\}^2 Var(t_i). \end{aligned} \tag{B.4}$$

where $p(\ \ )$ is a probability distribution function. Hence, $Var\{f(t_i)\}$ can be approximated by using the estimated mean and variance of t_i and by calculating the first derivative of the nonlinear function, we obtain the following CR′:

$$CR_t' \approx \frac{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2\{f'(E(t_i))\}^2 Var(t_i)}{\left\{\sum_{j=1}^{p}\lambda_j\right\}^2\{f'(E(t_i))\}^2 Var(t_i)+\sum_{j=1}^{p}\psi_j}. \tag{B.5}$$

For AVE′, assuming $\lambda_j\lambda_s = 0$ for any $j \neq s$ at (B.2) and (B.5), we obtain

$$\begin{aligned} AVE_t' &= \frac{\sum_{j=1}^{p}\lambda_j^2 Var\{f(t_i)\}}{\sum_{j=1}^{p}\lambda_j^2 Var\{f(t_i)\}+\sum_{j=1}^{p}Var(\varepsilon_{ji})} \\ &\approx \frac{\sum_{j=1}^{p}\lambda_j^{\,2}\{f'(E(t_i))\}^2 Var(t_i)}{\sum_{j=1}^{p}\lambda_j^{\,2}\{f'(E(t_i))\}^2 Var(t_i)+\sum_{j=1}^{p}\psi_j} \end{aligned} . \tag{B.6}$$

If the measurement model is the linear model so that $f(t_i) = t_i$, then $CR_t' = CR_t$ and $AVE_t' = AVE_t$ because $f'(t_i) = 1$ at any point.

C. Additional Extension of CR and AVE in Heterogeneity

We also provide the CR and AVE in case for measurement model with heterogeneity (individual parameters). Consider a composite measure for all j and i:

$$\begin{aligned} Z &= \sum_{j=1}^{p}\sum_{i=1}^{n} z_{ji} = \sum_{j=1}^{p}\sum_{i=1}^{n}\left(\lambda_{ji}t_i + \varepsilon_{ji}\right) \\ &= \sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i + \sum_{j=1}^{p}\sum_{i=1}^{n}\varepsilon_{ji} = T + E. \end{aligned} \tag{C.1}$$

Hence, the reliability is given by

$$\frac{Var(T)}{Var(Z)} = \frac{Var(T)}{Var(T+E)} = \frac{Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i\right)}{Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i\right)+Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\varepsilon_{ji}\right)} \tag{C.2}$$

Because $\varepsilon_{ji} \perp \varepsilon_{jl}$ for any $i \neq l$ and $Var(\varepsilon_{ji}) = \psi_{ji}$ for any j and i,

$$Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\varepsilon_{ji}\right) = \sum_{i=1}^{n}\left\{\sum_{j=1}^{p}Var(\varepsilon_{ji})\right\} = \sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\psi_j\right\} \tag{C.3}$$

Then, because $t_i \perp t_l$ for any $i \neq l$ and $Var(t_i) = 1$ for all i,

$$\begin{aligned} Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i\right) &= \sum_{i=1}^{n}Var\left(\sum_{j=1}^{p}\lambda_{ji}t_i\right) \\ &= \sum_{i=1}^{n}\left[\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^2 Var(t_i)\right] = \sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^2 \end{aligned} \tag{C.4}$$

Hence, CR for the measurement model with individual parameters is given by

$$\underset{\langle heterogeneity\rangle}{CR_t} = \frac{Var(T)}{Var(Z)} = \frac{Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i\right)}{Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\lambda_{ji}t_i\right)+Var\left(\sum_{j=1}^{p}\sum_{i=1}^{n}\varepsilon_{ji}\right)} \tag{C.5}$$

$$=\frac{\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^{2}}{\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^{2}+\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\psi_{ji}\right\}}. \tag{C.5}$$

For AVE with heterogeneity, because AVE assumes $\lambda_{ji}\lambda_{si}=0$ for any $j\neq s$ at the second equation in (B.5,) we obtain the following:

$$\underset{\langle heterogeneity\rangle}{AVE_t}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\lambda_{ji}^{2}}{\sum_{i=1}^{n}\sum_{j=1}^{p}\lambda_{ji}^{2}+\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{ji}} \tag{C.6}$$

We use these results to derive the CR′ and AVE′ for asymmetric function. If $\psi_{ji}=\psi_{jl}=\psi_j$ and $\lambda_{ji}=\lambda_{jl}=\lambda_j$ for any $i\neq l$,

$$\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\psi_{j}\right\}=n\sum_{j=1}^{p}\psi_{j}\ , \tag{C.3'}$$

and

$$\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}=n\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}. \tag{C.4'}$$

Hence,

$$\begin{aligned}\frac{\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^{2}}{\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\lambda_{ji}\right\}^{2}+\sum_{i=1}^{n}\left\{\sum_{j=1}^{p}\psi_{ji}\right\}}&=\frac{n\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}}{n\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}+n\sum_{j=1}^{p}\psi_{j}}\\&=\frac{\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}}{\left\{\sum_{j=1}^{p}\lambda_{j}\right\}^{2}+\sum_{j=1}^{p}\psi_{j}}=CR_t,\end{aligned} \tag{C.5'}$$

and

$$\begin{aligned}\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\lambda_{ji}^{2}}{\sum_{i=1}^{n}\sum_{j=1}^{p}\lambda_{ji}^{2}+\sum_{i=1}^{n}\sum_{j=1}^{p}\psi_{ji}}&=\frac{n\sum_{j=1}^{p}\lambda_{j}^{2}}{n\sum_{j=1}^{p}\lambda_{j}^{2}+n\sum_{j=1}^{p}\psi_{j}}\\&=\frac{\sum_{j=1}^{p}\lambda_{j}^{2}}{\sum_{j=1}^{p}\lambda_{j}^{2}+\sum_{j=1}^{p}\psi_{j}}=AVE_t\end{aligned} \tag{C.6'}$$

The last equation in C.5′and C.6′ indicate the original CR and AVE so that C.5 and C.6 can be widely used in general cases.

D. MCMC Algorithm for Nonlinear Measurement Model

We introduce the MCMC algorithm according to Zhu and Lee (1999). Consider the following nonlinear factor analysis model for the $p \times 1$ manifest random vector $\boldsymbol{y}^T = \left(y^{(1)}, \cdots, y^{(p)}\right)$:

$$\boldsymbol{y} = \Lambda F(\omega) + \varepsilon, \tag{D.1}$$

where Λ is a $p \times r$ factor loading matrix, $\omega = \left(\omega^{(1)}, \cdots, \omega^{(q)}\right)$ is a random vector of latent factors with $q < p$, ε is a random vector of error measurements, and $F(\omega) = \left(f_1(\omega), \cdots, f_r(\omega)\right)^T$ with differentiable functions $f_1, \cdots, f_r$, and $q \leq r$. Similar to the usual assumptions for factor analysis, it is assumed that ω is distributed as $N[\mathbf{0}, \Phi]$ and ε is distributed as $N[\mathbf{0}, \Psi]$, where Ψ is a diagonal matrix and ω and ε are independent.

Let $\mathbf{Y} = \{y_i, \cdots, y_n\}$ be the observed data matrix corresponding to a random sample obtained from a population with model (D.1,): $\boldsymbol{\Omega} = \{\omega_1, \cdots, \omega_n\}$ is the matrix of latent factors, and $\mathbf{F} = \{F(\omega_1), \cdots, F(\omega_n)\}$. We set prior distributions as follows,

parameter	settings
$\Lambda_j \mid \psi_{jj} \sim N\left(\Lambda_{j,0}, \psi_{jj} H_{0j}\right)$	$\Lambda_{j,0} = \mathbf{0}$, $H_{0j} = I_r$
$\psi_{\varepsilon,j} \sim IG\left(\alpha_{0j}, \beta_{0j}\right)$	$\alpha_{0j} = 0.01$, $\beta_{0j} = 0.01$
$\Phi \sim IW_q\left(\mathbf{R}_0, \rho_0\right)$	$\mathbf{R}_0 = I_q$, $\rho_0 = q$

For posteriors, set $s(= 1, \cdots, S)$ as a number of MCMC iterations and generate $\xi_i | \Lambda, \Psi, \Phi, y_i$ as follows

[1]
$$\xi_i^s = \xi_i^{s-1} + \eta; \quad \eta \sim N\left(\mathbf{0}, \sigma_\xi^2 I_q\right). \tag{D.2}$$

The probability of acceptance is

$$\min\left[\frac{p\left(\xi_i^s \mid \Lambda, \Psi, \Phi, y_i\right)}{p\left(\xi_i^{s-1} \mid \Lambda, \Psi, \Phi, y_i\right)}\right], \tag{D.3}$$

where σ_ξ^2 is a step size parameter that is given such that each acceptance rate becomes approximately 0.25. For $j = 1, \cdots, p$,

[2]
$$\left[\boldsymbol{\Lambda}_j \mid \psi_{\varepsilon,j}, \boldsymbol{\Omega}, \mathbf{Y}\right] \sim N\left(\mathrm{a}_j, \psi_{jj} \mathrm{A}_j\right) \tag{D.4}$$

where $\mathrm{A}_j = \left(H_{0j}^{-1} + \mathbf{F}\mathbf{F}^T\right)^{-1}$, and $\mathrm{a}_j = \mathrm{A}_{\lambda,j}\left(H_{0j}^{-1}\boldsymbol{\Lambda}_{0j} + \mathbf{F}\mathbf{Y}_j\right)$.

[3]
$$\left[\psi_{jj} \mid \boldsymbol{\Omega}, \mathbf{Y}\right] \sim IG\left(\alpha_{0j} + \frac{n}{2}, \beta_j\right), \tag{D.5}$$

where $\beta_j = \beta_{0j} + \frac{1}{2}\left(\mathbf{Y}_j^T \mathbf{Y}_j - \mathrm{a}_j^T \mathrm{A}_j^{-1} \mathrm{a}_j + \boldsymbol{\Lambda}_{0j}^T H_{0j}^{-1} \boldsymbol{\Lambda}_{0j}\right)$.

[4]
$$[\Phi | \boldsymbol{\Omega}] \sim IW\left(\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \mathbf{R}_0, n + \rho_0\right). \tag{D.6}$$

The above results are valid for situations where all elements of Λ are free parameters. Here, consider that Λ_j^T, the jth row of Λ, contains fixed parameters. Let c_j be the $1 \times q$ row vector such that $c_{jk} = 0$ if λ_{jk} is a fixed parameter and $c_{jk} = 1$ if λ_{jk} is an unknown parameter for

$j = 1, \cdots, p$ and $k = 1, \cdots, q$; $r_j = c_{j1} + \cdots + c_{jq}$ be the number of unknown parameter in Λ_j^T; Λ_j^{*T} be a $1 \times r_j$ row vector that contains the only unknown parameters in Λ_j^T; $\mathbf{\Omega}_j^*$ be an $r_j \times n$ submatrix of $\mathbf{\Omega}$ such that all the rows corresponding to $c_{jk} = 0$ are deleted; and $\mathbf{Y}_j^{*T} = (y_{j,1}^*, \cdots, y_{j.n}^*)$ with

$$y_{j,i}^* = y_{j,i} - \sum_{k=1}^{q} \lambda_{j,k} f_k(\omega_i)(1 - c_{jk}). \tag{D.7}$$

The conditional distributions with $\mathbf{\Lambda}_j$, $\mathbf{Y}_j$, $\mathbf{\Omega}$ in part of $[\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon]$ must be replaced by $\mathbf{\Lambda}_j^*$, $\mathbf{Y}_j^*$, $\mathbf{\Omega}_j^*$.

For example, consider an asymmetric nonlinear measurement model with two latent factors as follows;

$$\begin{aligned} \boldsymbol{y} = \Lambda F(\omega) + \varepsilon &= \Lambda \begin{bmatrix} I(\omega_1 \geq 0) f(\omega_1) \\ I(\omega_2 \geq 0) f(\omega_2) \\ I(\omega_1 < 0) f(\omega_1) \\ I(\omega_2 < 0) f(\omega_2) \end{bmatrix} + \varepsilon \\ &= \begin{bmatrix} 1 & 0 & \lambda_{1,3} & 0 \\ \lambda_{2,1} & 0 & \lambda_{2,3} & 0 \\ \lambda_{3,1} & 0 & \lambda_{3,3} & 0 \\ 0 & 1 & 0 & \lambda_{4,4} \\ 0 & \lambda_{5,2} & 0 & \lambda_{5,4} \\ 0 & \lambda_{6,2} & 0 & \lambda_{6,4} \end{bmatrix} \begin{bmatrix} f_1(\omega_1) \\ f_2(\omega_2) \\ f_3(\omega_1) \\ f_4(\omega_2) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}, \end{aligned} \tag{D.8}$$

Hence, for $j = 1$ with $k = 1$ and $j = 4$ with $k = 2$,

[2'] $$\left[\lambda_{j,k} \mid \psi_{jj}, \mathbf{\Omega}, \mathbf{Y}\right] \sim N\left(\mathrm{a}_j, \psi_{jj} a_j\right) \tag{D.9}$$

where $a_j = \left(h_{0\,j(k+2,k+2)}^{-1} + f_{k+2}(\omega_k)^T f_{k+2}(\omega_k)\right)^{-1}$, and

$\mathrm{a}_j = a_j \left(h_{0\,j(k+2,k+2)}^{-1} \lambda_{0\,j,k+2} + f_{k+2}(\omega_{k+2})^T \left\{\mathbf{Y}_j - f_k(\omega_k)\right\}\right)$.

[3'] $$\left[\psi_{jj} \mid \mathbf{\Omega}, \mathbf{Y}\right] \sim IG\left(\alpha_{0\,j} + \frac{n}{2}, \beta_j\right), \tag{D.10}$$

where $$\begin{aligned} \beta_j = \beta_{0\,j} + \frac{1}{2}\Big(&\left\{\mathbf{Y}_j - f_k(\omega_k)\right\}^T \left\{\mathbf{Y}_j - f_k(\omega_k)\right\} \\ &- \mathrm{a}_j^2 a_j^{-1} + \lambda_{0\,j,k+2}^2 h_{0\,j(k+2,k+2)}^{-1}\Big) \end{aligned},$$

and for $j = 2, 3$ with $k = 1$ and $j = 5, 6$ with $k = 2$

[2''] $$\left[\left\{\lambda_{jk}\right\} \mid \psi_{jj}, \mathbf{\Omega}, \mathbf{Y}\right] \sim N\left(\mathrm{a}_j, \psi_{jj} A_j\right) \tag{D.11}$$

where $\mathrm{A}_j = \left(\begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{1,1} & h_{2,2} \end{bmatrix}_{0j}^{-1} + \begin{bmatrix} f_k(\omega_k) \\ f_{k+2}(\omega_k) \end{bmatrix} \begin{bmatrix} f_k(\omega_k) \\ f_{k+2}(\omega_k) \end{bmatrix}^T \right)^{-1}$, and

$\mathrm{a}_j = \mathrm{A}_j \left(\begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{1,1} & h_{2,2} \end{bmatrix}_{0j}^{-1} \begin{bmatrix} \lambda_{j,k} \\ \lambda_{j,k+2} \end{bmatrix}_0 + \begin{bmatrix} f_k(\omega_k) \\ f_{k+2}(\omega_k) \end{bmatrix} \mathbf{Y}_j \right)$.

[3”] $$\left[\psi_{jj} \mid \mathbf{\Omega}, \mathbf{Y}\right] \sim IG\left(\alpha_{0j} + \frac{n}{2}, \beta_j\right), \tag{D.12}$$

where $\beta_j = \beta_{0j} + \frac{1}{2}\left(\mathbf{Y}_j^T \mathbf{Y}_j - \mathrm{a}_j^T \mathrm{A}_j^{-1} \mathrm{a}_j + \mathbf{\Lambda}_{0j}^T H_{0j}^{-1} \mathbf{\Lambda}_{0j}\right)$.

For improper solutions in Λ, it is possible to fix the left side parameters 1s in Λ as follows,

$$\Lambda = \begin{bmatrix} 1 & 0 & \lambda_{13} & 0 \\ 1 & 0 & \lambda_{23} & 0 \\ 1 & 0 & \lambda_{33} & 0 \\ 0 & 1 & 0 & \lambda_{44} \\ 0 & 1 & 0 & \lambda_{54} \\ 0 & 1 & 0 & \lambda_{64} \end{bmatrix} \tag{D.13}$$

or to assume that λ_{jk} follows positive truncated normal distribution with above restriction.

We take 1,000 MCMC samples after the algorithm converged in 500 for all simulation studies and take 3,000 MCMC samples after the algorithm converged in 2,000 for all models in the empirical analysis.

References

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, 103 (3), 411-423.

Anderson, James C. and David W. Gerbing (1992), "Assumptions and Comparative Strengths of the Two-Step Approach," *Sociological Methods & Research*, 20 (3), 321-333.

Babakus, Emin and Gregory W. Boller (1992), "An Empirical Assessment of the SERVQUAL Scale," *Journal of Business Research*, 24(3), 253-268.

Bagozzi, Richard P., and Lynn W. Phillips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly*, 27 (3), 459-489.

Bagozzi, Richard P., and Youjae Yi (1988), "On the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 16(1), 74-94.

Bagozzi, Richard P., Youjae Yi, and Lynn W. Phillips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36(3), 421-458.

Brown, Tom J., Gilbert A. Churchill, Jr., and J. Paul Peter (1993), "Research Note: Improving the Measurement of Service Quality," *Journal of Retailing*, vol. 69(1), 127-147.

Campbell, Donald T., and Donald W. Fiske (1959), "Convergent and Discriminate Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56 (2). 81-105.

Carman, James M. (1990), "Consumer Perceptions of Service Quality: an Assessment of the SERVQUAL Dimensions," *Journal of Retailing,* 66(2), 33-55.

Cronbach, Lee J., and Paul E. Meehl (1955), "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52(4), 281-302.

Cronbach, Lee J. (1951), "Coefficient Alpha and The Internal Structure of Tests," *Psychometrika*, 16(3), 297-334.

Cronbach, L. C., Peter Schönemann, and Douglas McKie (1965), "Alpha Coefficient for Stratified-Parallel Tests," *Educational and Psychological Measurement*, 25(2), 291-312.

Cronin, Joseph J. and Steven A. Taylor (1992), "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56(3), 55-68.

Cronin, Joseph J. and Steven A. Taylor (1994), "SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *Journal of Marketing*, 58(1), 125-131.

Edwards, Jeffery R. (2001), "Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework," *Organizational Research Methods*, 4(2), 144-192.

Edwards, Jeffery R. (2003), "Construct Validation in Organizational Behavior Research," J. Greenberg (Ed.), *Organizational behavior: The state of the science (2nd ed)*, 327-371.

Fornell, Claes, and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18(1), 39-50.

Fornell, Claes, and Youjae, Yi. (1992), "Assumptions of the Two-Step Approach to Latent Variable Modeling," *Sociological Methods & Research*, 20 (3), 291-320.

Fowler, Floyd Jackson, Jr. and Charles F. Cannell (1996), "Using Behavioral Coding to Identify Cognitive Problems with Survey Questions," Norbert Schwarz and Seymour Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in survey research*, Jossey-Bass, 15-36.

Gelman, Andrew, Gareth O. Roberts, and Walter R. Gilks (1995), "Efficient Metropolis jumping rules," J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith (Eds.), *Bayesian statistics 5*, Oxford: Oxford University Press, 599 -607.

Green, William H. (2011), *Econometric Analysis 7th edition*, Pearson.

Hair Jr, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson (2009), *Multivariate Data Analysis 7th edition*, Pearson.

Hu, Li-tze, and Peter M. Bentler (1998), "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification," *Psychological Methods*, 3(4), 424-453.
Hu, Li-tze, and Peter M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," *Structural Equation Modeling*, 6(1), 1-55.
Hughes, David J. (2018), "Psychometric Validity Establishing the Accuracy and Appropriateness of Psychometric Measures," P. Irwing, T. Booth, and D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*, 751-779.
Jarvis, Cheryl Burke., Scott B. Mackenzie, and Philip M. Podsakoff (2003), "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research*, 30(2), 199-218.
Jones, Lyle V. and David Thisssen (2006), "A History and Overview of Psychometrics," *Handbook of Statistics vol.26: Psychometrics*, pp. 1-28.
Jöreskog, K. G. (1971), "Statistical Analysis of Sets of Congeneric Tests," *Psychometrika*, 36(2), 109-133.
Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263-292.
Lee, Sik-Yum (2007), *Structural Equation Modeling: A Bayesian Approach*, John Wiley & Sons.
Lewis, Charles (2006), "Selected Topics in Classical Test Theory," *Handbook of Statistics vol.26: Psychometrics*, pp. 29-43.
MacKenzie, Scott B., Phillip M. Podsakoff, and Nathan P. Podsakoff (2011), "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly*, 35(2), 293-334.
Mathis, Robert L. and John H. Jackson (2010), *Human Resource Management 13th Edition*, South-Western Cengage Learning.
Malhotora, Naresh K., and David F. Birks (2007), *Marketing Research: An Applied Approach. 3rd edition*, Pearson.
McDonald, Roderck P. (1978), "Generalizability in Factorable Domains: "Domain Validity and Generalizability"," *Educational and Psychological Measurement*, 38(1), 75-79.
Messick, Samuel (1995), "Validity of Psychological assessment: Validation of Inferences from Persons' and Performances as Scientific Inquiry Into Score Meaning," *American Psychologist*, 50(9), 741-749.
Novick, Melvin R. (1966), "The Axioms and Principal Results of Classical Test Theory," *Journal of Mathematical Psychology*, 3(1), 1-18.
Novick, Melvin R., and Charles Lewis (1967), "Coefficient Alpha and The Reliability of Composite Measurements," *Psychometrika*, 32(1), 1-13.
Parasuraman, A., Leonard L. Berry and Valarie A. Zeithaml (1993), "More on Improving Service Quality Measurement," *Journal of Retailing*, 69(1), 140-147.
Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, 49(4), 41-50.
Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64(1), 12-40.
Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1994a), "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Imprecations for Further Research," *Journal of Marketing*, 58(1), 111-124.
Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1994b), "Alternative Scale for Measuring Service Quality: A Comparative Assessment Based on Psychometric and Diagnostic Criteria," *Journal of Retailing*, 70(3), 201-230

Park, C. Whan (2012), “Two types of Atteractive Research: Cute Research and Beautiful Research,” *Journal of Consumer Psychology*, 22(3), 299-302.

Rajaratnam, Nageswari, and Lee J. Cronbach, and Goldine C. Gleser (1965), “Generalizability of Stratified-Parallel Tests,” *Psychometorika*, 30(1), 39-56.

Ribeiro, Marco Tulio, Sameer Singh and Carlos Guestrin (2016a), “Model-Agnostic Interpretability in Machine Learning,” *2016 ICML Workshop on Human Interpretability in Machine Learning*, 91-95.

Ribeiro, Marco Tulio, Sameer Singh and Carlos Guestrin (2016b), ““Why Should I Trust You?” Explaining the Predictions of Any Classifier,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.

Schwarz, Norbert (2007), “Cognitive Aspects of Survey Mythology,” *Applied Cognitive Psychology*, 21(2), 277-287.

Serfling, Robert J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons.

Song, Xin-Yuan and Sik-Yum Lee (2012), *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*, John Wiley & Sons.

Tipping, Michael E. and Christopher M. Bishop (1999), “Probabilistic Principal Component Analysis,” *Journal of the Royal Statistical Society. Series B*, 61(3), 611-622.

Traub, Ross E. (1997), “Classical Test Theory in Historical Perspective,” *Educational Measurement: Issues and Practice*, 16(4), 8-14.

Watanabe, Sumio (2010a), “Equations of States in Singular Statistical Estimation,” *Neural Networks*, 23(1), 20-34.

Watanabe, Sumio (2010b), “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory,” *Journal of Machine Learning Research*, 11(Dec), 3571-3594.

Watanabe, Sumio (2013), “A Widely Applicable Bayesian Information Criterion,” *Journal of Machine Learning Research*, 14(Mar), 867-897.

Webb, N. M., Shavelson, R. J., and Haertel (2006), “Reliablity Coefficients and Generalizability Theory,” *Handbook of Statistics vol.26: Psychometrics*, pp. 81-124.

Werts, Charles E., Darryl R. Rock, Robert L. Linn, and Karl G. Jöreskong (1978), “A General Method of Estimating the Reliability of a Composite,” *Educational and Psychological Measurement*, 38(4), 933-938.

Zhu, Hong-Tu, and Sik-Yum Lee (1999), “Statistical Analysis of Nonlinear Factor Analysis Models,” *British Journal of Mathematical and Statistical Psychology*, 52(2), 225-242.