HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling

Ellen Roemer
Department of Business Administration and Economics, Hochschule Ruhr West, Mulheim an der Ruhr, Germany
Florian Schuberth
Department of Design, Production and Management, Universiteit Twente, Enschede, The Netherlands, and
Jörg Henseler
Department of Design, Production and Management, Universiteit Twente, Enschede, The Netherlands and Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisbon, Portugal

Received 8 February 2021
Revised 23 June 2021
Accepted 1 July 2021

Abstract

Purpose – One popular method to assess discriminant validity in structural equation modeling is the heterotrait-monotrait ratio of correlations (HTMT). However, the HTMT assumes tau-equivalent measurement models, which are unlikely to hold for most empirical studies. To relax this assumption, the authors modify the original HTMT and introduce a new consistent measure for congeneric measurement models: the HTMT2.

Design/methodology/approach – The HTMT2 is designed in analogy to the HTMT but relies on the geometric mean instead of the arithmetic mean. A Monte Carlo simulation compares the performance of the HTMT and the HTMT2. In the simulation, several design factors are varied such as loading patterns, sample sizes and inter-construct correlations in order to compare the estimation bias of the two criteria.

Findings – The HTMT2 provides less biased estimations of the correlations among the latent variables compared to the HTMT, in particular if indicators loading patterns are heterogeneous. Consequently, the HTMT2 should be preferred over the HTMT to assess discriminant validity in case of congeneric measurement models.

Research limitations/implications – However, the HTMT2 can only be determined if all correlations between involved observable variables are positive.

Originality/value – This paper introduces the HTMT2 as an improved version of the traditional HTMT. Compared to other approaches assessing discriminant validity, the HTMT2 provides two advantages: (1) the ease of its computation, since HTMT2 is only based on the indicator correlations, and (2) the relaxed assumption of tau-equivalence. The authors highly recommend the HTMT2 criterion over the traditional HTMT for assessing discriminant validity in empirical studies.

Keywords Structural equation modeling (SEM), Monte Carlo simulation, Discriminant validity, Fornell-Larcker criterion, Heterotrait-monotrait (HTMT), Measurement model assessment

Paper type Research paper

Jörg Henseler acknowledges a financial interest in ADANCO and its distributor, Composite Modeling.

Industrial Management & Data Systems
Vol. 121 No. 12, 2021
pp. 2637-2650
Emerald Publishing Limited
0263-5577
DOI 10.1108/IMDS-02-2021-0082

Introduction

The assessment of discriminant validity is of particular importance for the empirical study of relationships between theoretical concepts (Bagozzi and Phillips, 1982; Henseler *et al.*, 2015; Hair *et al.*, 2017; Voorhees *et al.*, 2016; Franke and Sarstedt, 2019; Rönkkö and Cho, 2020; Henseler, 2021). For the operationalization of these theoretical concepts it is vital that the employed measurement models actually measure what they are supposed to measure (Campbell and Fiske, 1959), thus establishing construct validity (Peter and Churchill, 1986). Construct validity comprises different forms, including discriminant validity (Netemeyer *et al.*, 2003). Discriminant validity in turn is defined as "the degree to which two measures designed to measure similar, but conceptually different, constructs are related. A low to moderate correlation is often considered evidence of discriminant validity" (Netemeyer *et al.*, 2003, p. 13).

The methodological literature provides different approaches to assess discriminant validity. Among others, the constrained phi approach (Jöreskog, 1971), the Fornell–Larcker criterion (Fornell and Larcker, 1981) and the comparison of cross-loadings (Chin, 1998) have been suggested to assess discriminant validity. Recently, Henseler *et al.* (2015) suggested the heterotrait–monotrait ratio of correlations (HTMT) to assess discriminant validity. Due to its good performance and straightforward application, the HTMT has found widespread application and dissemination, making Henseler *et al.* (2015) one of the most frequently cited papers in business research. Although the HTMT was originally proposed for models estimated by partial least squares path modeling (Wold, 1982), it also finds its application in structural equation modeling (Voorhees *et al.*, 2016). To assess discriminant validity using the HTMT, two strategies have been proposed: (1) comparison of the HTMT to predetermined thresholds and (2) constructing confidence intervals for the HTMT. Considering the former, heuristic rules are applied for the HTMT. For instance, the HTMT is compared to 0.85 to judge whether discriminant validity is violated (Henseler *et al.*, 2015). Following the latter, statistical inference is made by means of bootstrap confidence intervals, i.e. it is investigated whether the correlation between two latent variables is significantly different from 1. This approach has been very effective in detecting discriminant validity issues (Henseler *et al.*, 2015; Franke and Sarstedt, 2019).

In comparison to other approaches, the HTMT's main advantage is that it is relatively easy to calculate. For its computation, only the indicator correlation matrix is required (Henseler *et al.*, 2015). Consequently, the HTMT is not affected by the employed estimator and can be computed without estimating a model in advance. Despite this important advantage, a conceptual issue emerges since the HTMT assumes tau-equivalence (Henseler *et al.*, 2015; Rönkkö and Cho, 2020) and thus is likely to be biased in empirical cases, in which this assumption rarely holds (e.g. McNeish, 2018).

In order to relax the assumption of tau-equivalence and to offer a superior method to assess discriminant validity in the case of congeneric measurement models (i.e. with heterogeneous loading patterns), we develop a modification of the HTMT criterion, which we call HTMT2. Specifically, we compose a new formula for the HTMT2 criterion by replacing the arithmetic means applied in the HTMT's computation by geometric means. In doing so, we show that the HTMT2 is a consistent estimator for the inter-construct correlation in the case of congeneric measurement models. Therefore, the HTMT2 is highly recommendable to be used in empirical studies, if the involved correlations among observable variables are positive. To evaluate the HTMT2's finite sample behavior, we conduct a computational experiment in form of a Monte Carlo simulation that identifies the conditions, under which the HTMT2 outperforms the traditional HTMT.

The structure of this paper largely follows the suggestion by Gregor and Hevner (2013). In the next section, we outline the composition of the traditional HTMT criterion supported by a theoretical and a numerical example to pave the way for the introduction of the new criterion HTMT2. We show that the HTMT2 is a consistent estimator for the inter-construct correlation in the case of congeneric measurement models. Thereafter,

we conduct a simulation study to compare the performance of the HTMT to the HTMT2. Our simulation study shows that the HTMT2 outperforms the traditional HTMT approach in several situations. We discuss the results and conclude with avenues for future research.

The traditional HTMT

The HTMT was introduced by Henseler *et al.* (2015) as an estimator for the correlation between two latent variables. It is based on the multitrait-multimethod (MTMM) matrix, in which correlations are compared to assess discriminant validity (Campbell and Fiske, 1959). For a deeper understanding of discriminant validity assessment using the HTMT, we provide a theoretical and a numerical example.

For the theoretical example, we consider a simple model with two correlated constructs ξ_1 and ξ_2 (see Figure 1). The inter-construct correlation is denoted by φ. Each construct is measured by three indicators; ξ_1 is measured by x_{11} to x_{13}, and ξ_2 is measured by x_{21} to x_{23}, where λ_{11} to λ_{13} as well as λ_{21} to λ_{23} represent the respective loadings. The random measurement errors are referred to as ε_{11} to ε_{13} and ε_{21} to ε_{23}.

For the construction of the MTMM matrix and the HTMT, only the correlations among the indicators are required. Two types of correlations should be distinguished: monotrait-heteromethod correlations and heterotrait-heteromethod correlations (Campbell and Fiske, 1959). The former include the indicator correlations within one and the same construct. For ξ_1 this would be the correlations between the indicators x_{11}, x_{12} and x_{13}. The latter refer to the correlations between the indicators of two different constructs (Campbell and Fiske, 1959; Henseler *et al.*, 2015). In our example, the correlations between the indicators of ξ_1 (x_{11}, x_{12}, and x_{13}) and the indicators of ξ_2 (x_{21}, x_{22}, and x_{23}) are the heterotrait-heteromethod correlations.

Figure 2 shows the full MTMM matrix for the model depicted in Figure 1; its elements are the indicator correlations r. For our theoretical example, the correlations among the indicators x_{11} to x_{23} can be found in the lower triangle of the matrix. The monotrait-heteromethod correlations are framed by a solid line, whereas the heterotrait-heteromethod correlations are framed by a dashed line. Regarding the MTMM matrix, monotrait-heteromethod correlations should be larger than the heterotrait-heteromethod correlations to ensure that constructs can be discriminated in a model (Campbell and Fiske, 1959).

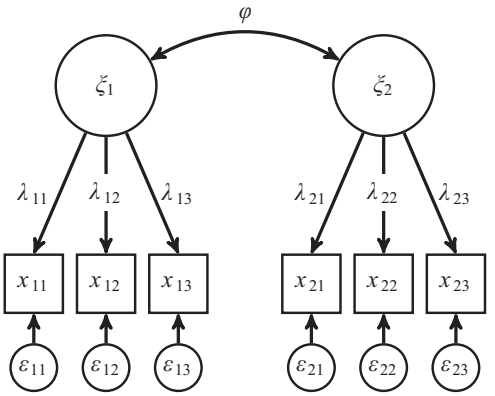

Figure 1.
A model with two constructs

Construct	Construct Indicator	ξ_1			ξ_2		
		x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}
ξ_1	x_{11}	1.0	0.7	0.7	0.5	0.5	0.5
	x_{12}	$r_{12,11}$	1.0	0.7	0.5	0.5	0.5
	x_{13}	$r_{13,11}$	$r_{13,12}$	1.0	0.5	0.5	0.5
ξ_2	x_{21}	$r_{21,11}$	$r_{21,12}$	$r_{21,13}$	1.0	0.4	0.4
	x_{22}	$r_{22,11}$	$r_{22,12}$	$r_{22,13}$	$r_{22,21}$	1.0	0.4
	x_{23}	$r_{23,11}$	$r_{23,12}$	$r_{23,13}$	$r_{23,21}$	$r_{23,22}$	1.0

— monotrait-heteromethod correlations

- - - heterotrait-heteromethod correlations

Figure 2.
MTMM matrix for the two factor model

Henseler *et al.* (2015) picked up this idea and translated it into a ratio of correlations. Specifically, the HTMT is the ratio of the arithmetic mean of the heterotrait–heteromethod correlations $r_{ig,jh}$ and the geometric mean of the arithmetic means of the monotrait-heteromethod correlations $r_{ig,ih}$ and $r_{jg,jh}$. In general, the HTMT can be calculated as follows:

$$\mathrm{HTMT}_{ij} = \frac{\overbrace{\frac{1}{K_i K_j}\sum_{g=1}^{K_i}\sum_{h=1}^{K_j} r_{ig,jh}}^{\substack{\text{arithmetic mean of indicator} \\ \text{correlations between } \xi_i \text{ and } \xi_j}}}{\sqrt{\left|\underbrace{\frac{2}{K_i(K_i-1)}\sum_{g=1}^{K_i-1}\sum_{h=g+1}^{K_i} r_{ig,ih}}_{\substack{\text{arithmetic mean of indicator} \\ \text{correlations within } \xi_i}} \cdot \underbrace{\frac{2}{K_j(K_j-1)}\sum_{g=1}^{K_j-1}\sum_{h=g+1}^{K_j} r_{jg,jh}}_{\substack{\text{arithmetic mean of indicator} \\ \text{correlations within } \xi_j}}\right|}} \tag{1}$$

where K_i and K_j denote the number of indicators belonging to construct ξ_i and ξ_j, respectively.

To establish discriminant validity, the HTMT value should be different from 1 because the HTMT is an estimator for the inter-construct correlation; if the correlation between two constructs is 1, they cannot be discriminated properly (Henseler *et al.*, 2015). For this judgment, the proposition was to compare the HTMT to a pre-defined threshold value (Henseler *et al.*, 2015; Voorhees *et al.*, 2016; Franke and Sarstedt, 2019). Recommended threshold values range from 0.85, which is considered a conservative benchmark (Henseler *et al.*, 2015; Voorhees *et al.*, 2016), to a more liberal cut-off value of 0.9 (Henseler *et al.*, 2015; Franke and Sarstedt, 2019) or higher. The choice of the threshold level should, however, be made against the background of how conservative the researcher wants to be in assessing discriminant validity and how confident (s)he is regarding the uniqueness of the constructs (Henseler *et al.*, 2015; Franke and Sarstedt, 2019).

In addition, the HTMT can be exposed to statistical inference. Specifically, Henseler *et al.* (2015) and Franke and Sarstedt (2019) suggested to investigate whether the upper bound of the 90% bootstrap confidence interval is larger than 1 warranting a type I error rate of 5%. If the value of 1, i.e. the two constructs are perfectly correlated, is larger than the upper bound of the bootstrap confidence interval, it can be concluded that the construct correlation is

significantly smaller than 1 (Henseler *et al.*, 2015; Franke and Sarstedt, 2019). Whereas Henseler *et al.* (2015) found that statistical inference about the HTMT is the most liberal way of assessing discriminant validity, Franke and Sarstedt (2019, p. 441) clearly advocate that "researchers should prefer inferential tests over simple cutoff values."

To further illustrate the logic of the HTMT criterion, we use a numerical example based on the correlations in the upper triangle of the MTMM matrix in Figure 2. The mean value of the heterotrait–heteromethod correlations is 0.5, whereas the geometric mean of the mean monotrait–heteromethod correlations (0.7 and 0.4) equals 0.5291. Taken together, the HTMT is computed as follows:

$$\text{HTMT} = \frac{0.5}{\sqrt{0.7 \cdot 0.4}} = 0.945. \tag{2}$$

Applying heuristic rules for the HTMT, the value of 0.945 in Equation (2) is then compared to a threshold value, e.g. 0.85 for more conservative and 0.90 for more liberal assessments. The HTMT value of 0.945 clearly exceeds even the more liberal HTMT level of 0.90. Hence, based on the heuristic rules, the HTMT value above the two threshold values indicates a lack of discriminant validity.

Moreover, we construct a 90% bootstrap confidence interval around the HTMT based on 999 bootstrap runs (Henseler *et al.*, 2015; Franke and Sarstedt, 2019) using the percentile bootstrap approach (Aguirre-Urreta and Rönkkö, 2018). The upper bound of the 90% percentile bootstrap confidence interval is 0.9899. Consequently, the value of 1 is not covered by the confidence interval showing that the HTMT value of 0.945 is significantly different from 1.

Recent studies have demonstrated a good performance of the HTMT (Henseler *et al.*, 2015; Voorhees *et al.*, 2016; Franke and Sarstedt, 2019). Another major advantage of the criterion is that its computation is hardly demanding, because only the indicators' correlation matrix is required as input for simple calculations (Henseler *et al.*, 2015; Voorhees *et al.*, 2016). In contrast to other approaches, no model estimates are required for its computation (Franke and Sarstedt, 2019). Consequently, it does not suffer from Heywood cases–a phenomenon not untypical for factor analysis (see Krijnen *et al.*, 1998). Hence, it is an easily applicable method to assess discriminant validity independent of the employed estimator (Voorhees *et al.*, 2016; Franke and Sarstedt, 2019).

Noteworthily, the HTMT does not come without disadvantages. Specifically, the HTMT makes the rather rigid assumption of tau-equivalent measurement models (Henseler *et al.*, 2015; Rönkkö and Cho, 2020). To illustrate this, Figure 3 depicts two different types of measurement models: Figure 3a represents a tau-equivalent measurement model, which assumes that all loadings (λ) are equal, i.e. all covariances among the indicators are equal (Lord and Novick, 1968). In contrast, Figure 3b shows a congeneric measurement model, in which this assumption is relaxed and indicator loadings (λ_1, λ_2 and λ_3) can vary (Jöreskog, 1971).

Transferred to our example in Figure 1, the HTMT assumes that all indicator loadings of one construct are equal, i.e. $\lambda_{11} = \lambda_{12} = \lambda_{13}$ and $\lambda_{21} = \lambda_{22} = \lambda_{23}$. This assumption, however, is unlikely to hold in most empirical research settings. As McNeish (2018, p. 414) puts it: "Tau equivalence tends to be unlikely for most scales that are used in empirical research–some items strongly relate to the construct while some are more weakly related." Relaxing the assumption of tau-equivalence and assessing discriminant validity in the case of congeneric measurement models, in which indicator loadings may differ from each other (i.e. $\lambda_{11} \neq \lambda_{12} \neq \lambda_{13}$ and $\lambda_{21} \neq \lambda_{22} \neq \lambda_{23}$), requires a revision of the traditional HTMT.

The HTMT2

We propose a new version of the HTMT, which we name HTMT2–an idea initially sketched by Henseler (2021). The purpose of the HTMT2 is to relax the HTMT's assumption of

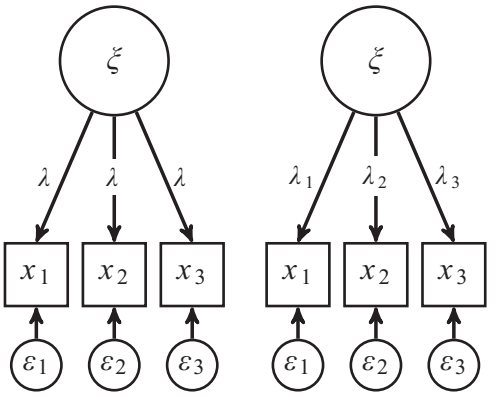

Figure 3.
Measurement models

tau-equivalence and thus to allow for assessing discriminant validity in the context of congeneric measurement models. In doing so, HTMT's principle of comparing heterotrait–heteromethod correlations to monotrait–heteromethod correlations in the MTMM matrix remains untouched. Hence, like the HTMT, the HTMT2 is based on correlations as a measure of linear dependency among variables. In contrast to the HTMT, for the calculation of the HTMT2, the geometric mean instead of the arithmetic mean is used for calculating the average indicator correlation. The use of the geometric mean is inspired by the fact that the variance-covariance matrix implied by reflective measurement models is based on products of loadings, i.e. the loadings are linked in multiplicative and not an additive way. Against this background, the use of the geometric mean appears more suitable. Moreover, as we will show in this section, the use of the geometric mean draws the HTMT2 into a consistent estimator for the inter-construct correlation in the case of congeneric measurement models.

The HTMT2 is given by Equation (3).

$$\mathrm{HTMT2}_{ij} = \frac{\overbrace{\sqrt[K_i \cdot K_j]{\prod_{g=1}^{K_i}\prod_{h=1}^{K_j} r_{ig,jh}}}^{\text{geometric mean of indicator correlations between } \xi_i \text{ and } \xi_j}}{\sqrt{\underbrace{\sqrt[\frac{K_i^2-K_i}{2}]{\prod_{g=1}^{K_i-1}\prod_{h=g+1}^{K_j} r_{jg,jh}}}_{\text{geometric mean of indicator correlations within } \xi_i} \cdot \underbrace{\sqrt[\frac{K_j^2-K_j}{2}]{\prod_{g=1}^{K_j-1}\prod_{h=g+1}^{K_j} r_{jg,jh}}}_{\text{geometric mean of indicator correlations within } \xi_j}}} \tag{3}$$

In the numerator of Equation (3), the geometric mean of the heterotrait–heteromethod correlations is calculated, while the denominator is composed of the geometric means of the two geometric means of the monotrait–heteromethod correlations. In contrast to the arithmetic mean, the geometric mean is only defined for strictly positive values. For

the HTMT2, this means that all indicator correlations used for calculating the HTMT2 must be greater than zero. This requirement is also known from the calculation of Cronbach's alpha which also requires that all indicators are positively correlated (Sijtsma, 2009).

Subsequently, we show that the HTMT2 is a consistent estimator for the inter-construct correlation φ. Let x_{i1} to x_{iK_i} be the K_i indicators of constructs ξ_i and x_{j1} to x_{jK_j} the K_J indicators of construct ξ_j. The empirical correlation matrix $\mathbf{S}$ of the indicators is generally given as follows:

$$\mathbf{S} = \begin{pmatrix} 1 & r_{i1,i2} & \dots & r_{i1,iK_i} & r_{i1,j1} & r_{i1,j2} & \dots & r_{i1,jK_j} \\ r_{i2,i1} & 1 & \dots & r_{i2,iK_i} & r_{i2,j1} & r_{i2,j2} & \dots & r_{i2,jK_j} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{iK_i,i1} & r_{iK_i,i2} & \dots & 1 & r_{iK_i,j1} & r_{iK_i,j2} & \dots & r_{iK_i,jK_j} \\ r_{j1,i1} & r_{j1,i2} & \dots & r_{j1,iK_i} & 1 & r_{j1,j2} & \dots & r_{j1,jK_j} \\ r_{j2,i1} & r_{j2,i2} & \dots & r_{j2,iK_i} & r_{j2,j1} & 1 & \dots & r_{j2,jK_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{jK_j,i1} & r_{jK_j,i2} & \dots & r_{jK_j,iK_i} & r_{jK_j,j1} & r_{jK_j,j2} & \dots & 1 \end{pmatrix} \tag{4}$$

Furthermore, we assume that the empirical correlation matrix $\mathbf{S}$ is a consistent estimate of $\mathbf{\Sigma}$, i.e. plim $\mathbf{S} = \mathbf{\Sigma}$, where plim is the probability limit, and $\mathbf{\Sigma}$ is the correlation matrix implied by the reflective measurement models:

$$\Sigma = \begin{pmatrix} 1 & \lambda_{i1}\lambda_{i2} & \dots & \lambda_{i1}\lambda_{iK_i} & \varphi_{ij}\lambda_{i1}\lambda_{j1} & \varphi_{ij}\lambda_{i1}\lambda_{j2} & \dots & \varphi_{ij}\lambda_{i1}\lambda_{jK_j} \\ \lambda_{i2}\lambda_{i1} & 1 & \dots & \lambda_{i2}\lambda_{iK_i} & \varphi_{ij}\lambda_{i2}\lambda_{j1} & \varphi_{ij}\lambda_{i2}\lambda_{j2} & \dots & \varphi_{ij}\lambda_{i2}\lambda_{jK_j} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{iK_i}\lambda_{i1} & \lambda_{iK_i}\lambda_{i2} & \dots & 1 & \varphi_{ij}\lambda_{iK_i}\lambda_{j1} & \varphi_{ij}\lambda_{iK_i}\lambda_{j2} & \dots & \varphi_{ij}\lambda_{iK_i}\lambda_{jK_j} \\ \varphi_{ij}\lambda_{j1}\lambda_{i1} & \varphi_{ij}\lambda_{j1}\lambda_{i2} & \dots & \varphi_{ij}\lambda_{j1}\lambda_{iK_i} & 1 & \lambda_{j1}\lambda_{j2} & \dots & \lambda_{j1}\lambda_{jK_j} \\ \varphi_{ij}\lambda_{j2}\lambda_{i1} & \varphi_{ij}\lambda_{j2}\lambda_{i2} & \dots & \varphi_{ij}\lambda_{j2}\lambda_{iK_i} & \lambda_{j2}\lambda_{j1} & 1 & \dots & \lambda_{j2}\lambda_{jK_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_{ij}\lambda_{jK_j}\lambda_{i1} & \varphi_{ij}\lambda_{jK_j}\lambda_{i2} & \dots & \varphi_{ij}\lambda_{jK_j}\lambda_{iK_i} & \lambda_{jK_j}\lambda_{j1} & \lambda_{jK_j}\lambda_{j2} & \dots & 1 \end{pmatrix}, \tag{5}$$

where λ_{ik} denotes the factor loading of indicator x_{ik} on construct ξ_i , and φ_{ij} is the correlation between constructs ξ_i and ξ_j. Considering the HTMT2:

$$\text{plim HTMT2}_{ij} = \text{plim} \frac{\sqrt[K_i \cdot K_j]{\prod_{g=1}^{K_i} \prod_{h=1}^{K_j} r_{ig,jh}}}{\sqrt{\sqrt[\frac{K_i^2 - K_i}{2}]{\prod_{g=1}^{K_i - 1} \prod_{h=g+1}^{K_i} r_{ig,ih}} \cdot \sqrt[\frac{K_j^2 - K_j}{2}]{\prod_{g=1}^{K_j - 1} \prod_{h=g+1}^{K_j} r_{jg,jh}}}} = \tag{6}$$

$$\frac{\sqrt[K_i \cdot K_j]{\prod_{g=1}^{K_i} \prod_{h=1}^{K_j} \varphi_{ij}\lambda_{ig}\lambda_{jh}}}{\sqrt{\sqrt[\frac{K_i^2 - K_i}{2}]{\prod_{g=1}^{K_i - 1} \prod_{h=g+1}^{K_i} \lambda_{ig}\lambda_{ih}} \cdot \sqrt[\frac{K_j^2 - K_j}{2}]{\prod_{g=1}^{K_j - 1} \prod_{h=g+1}^{K_j} \lambda_{jg}\lambda_{jh}}}} = \tag{7}$$

$$\varphi_{ij}\frac{\sqrt[K_i K_j]{\prod_{g=1}^{K_i}\prod_{h=1}^{K_j}\lambda_{ig}\lambda_{jh}}}{\sqrt[K_i^2-K]{\prod_{g=1}^{K_i-1}\prod_{h=g+1}^{K_i}\lambda_{ig}\lambda_{ih}}\cdot\sqrt[K_i^2-K]{\prod_{g=1}^{K_j-1}\prod_{h=g+1}^{K_j}\lambda_{jg}\lambda_{jh}}}= \tag{8}$$

$$\varphi_{ij}\frac{\sqrt[K_i K_j]{\prod_{g=1}^{K_i}\lambda_{ig}^{K_j}\cdot\prod_{g=1}^{K_j}\lambda_{jg}^{K_i}}}{\sqrt[K_i^2-K_i]{\prod_{g=1}^{K_i}\lambda_{ig}^{K_i-1}}\cdot\sqrt[K_j^2-K_j]{\prod_{g=1}^{K_j}\lambda_{jg}^{K_j-1}}}= \tag{9}$$

$$\varphi_{ij}\frac{\prod_{g=1}^{K_i}\lambda_{ig}^{\frac{1}{K_i}}\cdot\prod_{g=1}^{K_j}\lambda_{jg}^{\frac{1}{K_j}}}{\prod_{g=1}^{K_i}\lambda_{ig}^{\frac{1}{K_i}}\cdot\prod_{g=1}^{K_j}\lambda_{jg}^{\frac{1}{K_j}}}=\varphi_{ij} \tag{10}$$

Consequently, plim $\mathrm{HTMT2}_{ij} = \varphi_{ij}$ (q.e.d.). It is noted that the HTMT2 does not make any specific distributional assumption about the indicators. To show consistency of the HTMT2, it is sufficient that $\mathbf{S}$ is a consistent estimate of $\mathbf{\Sigma}$. However in the case of ordinal categorical indicators, the Pearson correlation may be replaced by a correlation measure that takes the scale of such indicators into account such as the polychoric or polyserial correlation coefficient (Olsson *et al.*, 1982). Similar has been suggested for the calculation of Cronbach's alpha (Zumbo *et al.*, 2007).

Evaluating the HTMT2: a Monte Carlo simulation

To investigate the performance of the new criterion, we ran a Monte Carlo simulation, in which we compared the HTMT2 to the traditional HTMT. The aim was to explore the HTMT2's finite sample behavior as well as its relative performance when the assumption of tau-equivalence is sequentially relaxed. To judge the performance of the two measures, we examined their estimated bias for the inter-construct correlation φ, i.e. the difference between the mean of the estimated and true inter-construct correlation.

For the simulation, we considered a model containing two constructs each measured by three indicators. To relax the assumption of tau-equivalence, we increased the heterogeneity in the loading patterns from homogeneous patterns (1) to substantially heterogeneous patterns (6). To do so, we increased and decreased the loadings in one block by 0.05 from pattern to pattern. The loading patterns used for the simulations are displayed in Table 1.

Loading Pattern	Heterogeneity	λ_{11}	λ_{12}	λ_{13}	λ_{21}	λ_{22}	λ_{23}
1	no	0.70	0.70	0.70	0.70	0.70	0.70
2		0.65	0.70	0.75	0.65	0.70	0.75
3		0.60	0.70	0.80	0.60	0.70	0.80
4		0.55	0.70	0.85	0.55	0.70	0.85
5		0.50	0.70	0.90	0.50	0.70	0.90
6	high	0.45	0.70	0.95	0.45	0.70	0.95

Table 1.
Loading patterns (first experimental factor)

Since the HTMT assumes tau-equivalent measurement models, we expect an increasing bias for more heterogeneous loading patterns. In contrast, the HTMT2 is expected to be less biased, i.e. it produces an average inter-construct correlation close to the population counterpart. As an additional experimental factor, we varied the sample size; the factor levels were 100, 250 and 500 observations per sample (in analogy to Henseler *et al.*, 2015) in order to investigate the finite sample behavior of the two measures. Since the HTMT2 is a consistent estimator for the inter-construct correlation, we expected that the higher the sample size, the smaller the bias. We presumed a similar behavior for the HTMT in the case of tau-equivalent measurement models. Moreover, we applied both the HTMT and the HTMT2 to the indicators' population correlation matrix. As a final experimental factor, we considered four different levels of inter-construct correlations, i.e. $\varphi = 0.75, 0.85, 0.90$ and 1.00 (in analogy to Henseler *et al.*, 2015; Franke and Sarstedt, 2019). We refrained from studying medium and lower degrees of inter-construct correlations, because in these cases discriminant validity infringements become less likely. In total, we studied 6 (loading patterns) $\times$ 4 (3 different sample sizes + population correlation matrix) $\times$ 4 (inter-construct correlations) = 96 conditions.

The complete simulation was conducted in the statistical programming environment R (R Core Team, 2020, Version 4.0.2). For each condition, we drew 1,000 data sets from a multivariate normal distribution with means of zero and a variance-covariance matrix equal to the model-implied correlation matrix of the corresponding condition. To generate the data sets, we used the mvrnorm function of the MASS package (Venables and Ripley, 2002). To compute HTMT and HTMT2 values, we used own R implementations. A total of 72,024 values were calculated for each of the HTMT and the HTMT2.

Finally, for a fair comparison, we removed data sets for which at least one negative indicator correlation was observed. Although the HTMT can be technically calculated in this case, its results are not trustworthy because negative and positive correlations cancel out. Similarly, the HTMT2 cannot be calculated in this case, as the geometric mean is not defined. Such cases occurred most frequently under the condition of small sample size in combination with a low inter-construct correlation and a substantial heterogeneous loading pattern. In this situation 7.2% of the data sets were removed. In all other situations, the share of removed data sets was below 5%.

Results and discussion

The simulation results are shown in Figure 4 [1]. The results for the different sample sizes including the indicators' population correlation matrix are visualized in the rows of Figure 4. In the columns, the four considered values of the inter-construct correlations are displayed. The vertical lines in each cell represent the six different loading patterns ranging from "no heterogeneity" to "high heterogeneity" (see Table 1). Finally, the black solid line with circles in each cell represents the results for the HTMT, whereas the dashed line with triangles represents the results for the HTMT2.

The results clearly indicate that the HTMT2 outperforms the HTMT in several situations. As expected, HTMT2 results are rather unaffected by the loading patterns when it comes to estimating the inter-construct correlation. In contrast, the HTMT's estimated bias increases when loading patterns become more heterogeneous. In other words, the HTMT becomes more distorted the more heavily indicator loadings deviate from each other, i.e. the more a measurement model diverges from tau-equivalence. In these cases, the HTMT may lead to erroneous conclusions regarding discriminant validity assessments of measurement models.

Considering the size of inter-construct correlations, both HTMT2 and HTMT results remain unaffected. However, larger inter-construct correlations increase HTMT's bias, particularly in case of a substantially heterogeneous loading pattern.

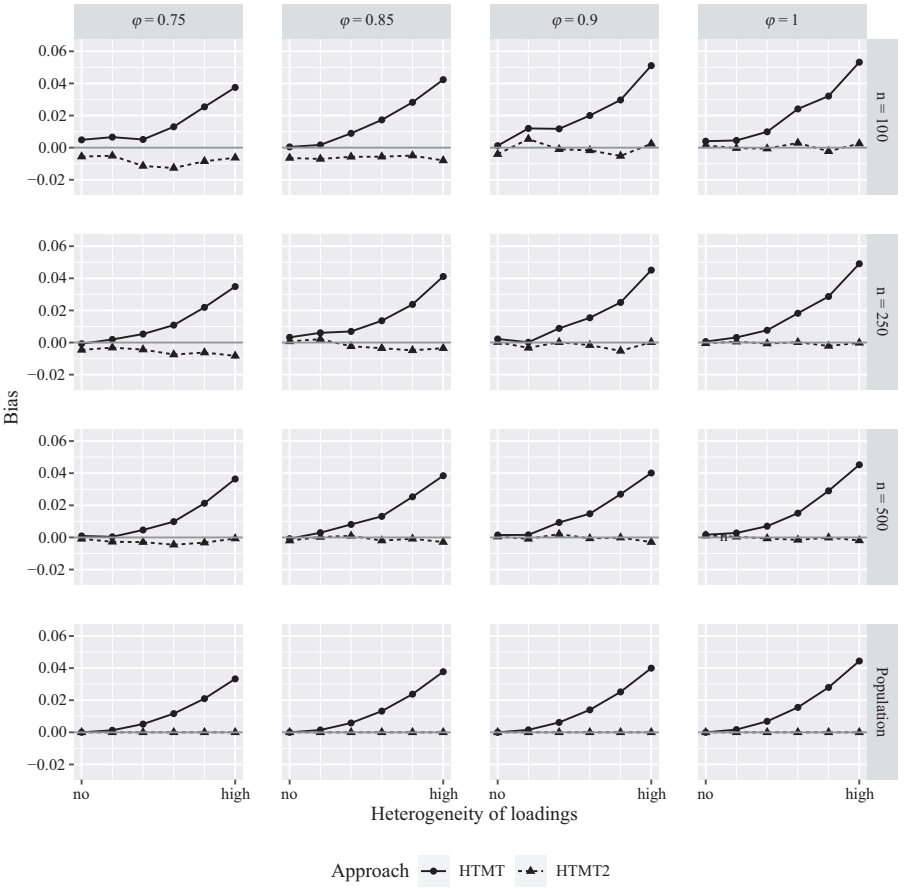

Figure 4.
Simulation results

Finally, with regard to sample size, HTMT2's bias disappears for large sample sizes as expected. In case of small sample sizes ($n \leq 100$), the HTMT2 is slightly downward biased, particularly for smaller inter-construct correlations ($\varphi \leq 0.85$). In the case that the HTMT and the HTMT2 are calculated based on the indicators' population correlation matrix, the HTMT2 shows no bias. In contrast, the HTMT is biased if the assumption about tau-equivalence is violated.

As a result, the HTMT2 is a suitable measure to assess discriminant validity in the case of congeneric measurement models as it provides consistent estimates of inter-construct correlations. Since the HTMT2 is based on the MTMM matrix and follows the same logic as the HTMT, it benefits from the same advantages as the HTMT. In particular, the HTMT2 can be easily calculated based on the indicator correlations without performing any further estimations. Hence, the HTMT2 can be calculated without time consuming procedures even for large sample sizes making it an attractive measure in the era of big data. To conclude, the HTMT2 includes HTMT's advantages, while overcoming HTMT's drawback of being only consistent for tau-equivalent measurement models. Consequently, researchers and practitioners should prefer the HTMT2 over the HTMT in situations, in which indicator loadings deviate from each other.

In line with prior research on the traditional HTMT, we recommend using statistical inference for the HTMT2 to detect discriminant validity problems (Franke and Sarstedt, 2019; Rönkkö and Cho, 2020). As for the HTMT, this can be done by constructing bootstrap confidence intervals for the HTMT2 to investigate whether the confidence interval covers the value 1 (Henseler *et al.*, 2015) [2]. If this is the case, a researcher has found no empirical evidence against a construct correlation of 1, which raises doubts about discriminant validity.

Conclusion and future research

In this paper, we have introduced the HTMT2 as an improved version of the HTMT criterion to assess discriminant validity in structural equation modeling (Henseler *et al.*, 2015). We have proved that the HTMT2 is a consistent estimator for the inter-construct correlation in the case of congeneric measurement models and thus outperforms the HTMT. Since the HTMT2 is equally based on the MTMM matrix comprising the correlations between the indicators in the measurement model, it benefits from the same advantages as the traditional HTMT criterion, i.e. its computation is straightforward without any need for estimation procedures. In contrast to the HTMT, the HTMT2 relaxes the rigid assumption of tau-equivalent measurement models thus equalizing one of HTMT's main disadvantages.

To limit the scope of our study, we only focused on the application of the HTMT2 in the case of reflective measurement models. However, in empirical studies researchers also deal with formative measurement models. Currently, the literature does not provide a unique definition about formative measurement (Diamantopoulos and Winklhofer, 2001). Formative measurement can either refer to the causal-formative measurement model (e.g. Bollen, 1984; Bollen and Lennox, 1991) or the composite model (Fornell and Bookstein, 1982). For composite models, the application of the HTMT and HTMT2 is of little value because the correlations among the indicators of one construct are unconstrained by the composite model and thus do not depend on the loadings (Henseler *et al.*, 2014; Dijkstra, 2017; Schuberth *et al.*, 2018). As a consequence, the HTMT2 will not converge in probability to the construct correlation because the loadings do not cancel each other out. However, the HTMT2 can be applied to constructs embedded in causal-formative measurement models if additional reflective measures are specified for their identification such as in the multiple indicators multiple causes (MIMIC) model (Joreskog and Goldberger, 1975). In this case, only the reflective indicators should be used for calculating the HTMT2. Applying the HTMT2 to formative indicators is problematic, because neither the monotrait–heteromethod nor the heterotrait–heteromethod correlations of formative indicators are informative about discriminant validity.

Our research paves the way for future avenues of research. First and foremost, the HTMT2 is not defined for cases, in which indicator correlations are negative. Monotrait–heteromethod correlations may be negative, if one or more indicators are negatively worded to measure the associated construct. In this case, reverse coding of the indicator(s) may represent a way to solve this issue. Future research may particularly investigate cases, in which heterotrait–heteromethod correlations become negative. For these cases, solutions to deal with negative heterotrait–heteromethod correlations for the HTMT2 need to be developed.

Second, further research is needed to identify appropriate threshold values indicating discriminant validity infringements. Even though we share the prevailing opinion on preferring statistical inference over applying heuristic rules (Franke and Sarstedt, 2019; Rönkkö and Cho, 2020), there may be some groups of users, for whom it is more convenient to compare the HTMT2 statistics to a predetermined cut-off value. Therefore, future simulation studies may identify threshold values to identify discriminant validity infringements.

Third, future research should investigate the performance of the various bootstrap confidence intervals for the HTMT2 and their suitability for statistical inference in this context.

Finally, to evaluate the HTMT2 in more detail, its performance could be compared to other methods of assessing discriminant validity (e.g., the well-known Fornell–Larcker criterion (Fornell and Larcker, 1981) or the constrained phi approach (Jöreskog, 1971). Simulation studies as well as empirical studies could serve as a foundation for this direction of research. Both cut-off values and statistical inference should be taken into account.

Notes

1. A table with the numerical simulation results can be received upon e-mail request to Jörg Henseler.
2. For the computation of the HTMT2, users may refer to the following website: www.henseler.com. The bootstrapping procedure for the HTMT2 is implemented in ADANCO 2.3. Interested users may consult www.composite-modeling.com for further information. In addition, the HTMT2 is implemented in the R package cSEM (Rademaker and Schuberth, 2020).

References

Aguirre-Urreta, M.I. and Rönkkö, M. (2018), "Statistical inference with PLSc using bootstrap confidence intervals", *MIS Quarterly*, Vol. 42 No. 3, pp. 1001-1020.

Bagozzi, R.P. and Phillips, L.W. (1982), "Representing and testing organizational theories: a holistic construal", *Administrative Science Quarterly*, Vol. 27 No. 3, pp. 459-489.

Bollen, K.A. (1984), "Multiple indicators: internal consistency or no necessary relationship?", *Quality and Quantity*, Vol. 18 No. 4, pp. 377-385.

Bollen, K. and Lennox, R. (1991), "Conventional wisdom on measurement: a structural equation perspective", *Psychological Bulletin*, Vol. 110 No. 2, p. 305.

Campbell, D.T. and Fiske, D.W. (1959), "Convergent and discriminant validation by the multitrait-multimethod matrix", *Psychological Bulletin*, Vol. 56 No. 2, pp. 81-105.

Chin, W.W. (1998), "The partial least squares approach for structural equation modeling", in Marcoulides, G.A. (Ed.), *Modern Methods for Business Research*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 295-336.

Diamantopoulos, A. and Winklhofer, H.M. (2001), "Index construction with formative indicators: an alternative to scale development", *Journal of Marketing Research*, Vol. 38 No. 2, pp. 269-277.

Dijkstra, T.K. (2017), "A perfect match between a model and a mode", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, Springer, Cham, pp. 55-80.

Fornell, C. and Bookstein, F.L. (1982), "Two structural equation models: LISREL and PLS applied to consumer exit-voice theory", *Journal of Marketing Research*, Vol. 19 No. 4, pp. 440-452.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Franke, G. and Sarstedt, M. (2019), "Heuristics versus statistics in discriminant validity testing: a comparison of four procedures", *Internet Research*, Vol. 29 No. 3, pp. 430-447.

Gregor, S. and Hevner, A.R. (2013), "Positioning and presenting design science research for maximum impact", *MIS Quarterly*, Vol. 37 No. 2, pp. 337-355.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2017), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd ed., Sage Publications, Thousand Oaks, London, New Delhi.

Henseler, J. (2021), *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables*, Guilford Press, New York.

Henseler, J., Dijkstra, T.K., Sarstedt, M., Ringle, C.M., Diamantopoulos, A., Straub, D.W., Ketchen, D.J., Hair, J.F., Hult, G.T.M. and Calantone, R.J. (2014), "Common beliefs and reality about PLS:

comments on Rönkkö and Evermann (2013)", *Organizational Research Methods*, Vol. 17 No. 2, pp. 182-209.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 1-21.

Jöreskog, K.G. (1971), "Statistical analysis of sets of congeneric tests", *Psychometrika*, Vol. 36 No. 2, pp. 109-133.

Joreskog, K.G. and Goldberger, A.S. (1975), "Estimation of a model with multiple indicators and multiple causes of a single latent variable", *Journal of the American Statistical Association*, Vol. 70 No. 351, p. 631.

Krijnen, W.P., Dijkstra, T.K. and Gill, R.D. (1998), "Conditions for factor (in)determinacy in factor analysis", *Psychometrika*, Vol. 63 No. 4, pp. 359-367.

Lord, F.M. and Novick, M.R. (1968), *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, MA.

McNeish, D. (2018), "Thanks coefficient alpha, we'll take it from here", *Psychological Methods*, Vol. 23 No. 3, pp. 412-433.

Netemeyer, R.G., Bearden, W.O. and Sharma, S. (2003), *Scaling Procedures: Issues and Applications*, Sage Publications, Thousand Oaks, London, New Delhi.

Olsson, U., Drasgow, F. and Dorans, N.J. (1982), "The polyserial correlation coefficient", *Psychometrika*, Vol. 47 No. 3, pp. 337-347.

Peter, J.P. and Churchill, G.A. Jr (1986), "Relationships among research design choices and psychometric properties of rating scales: a meta-analysis", *Journal of Marketing Research*, Vol. 23 No. 1, pp. 1-10.

R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, available at: https://www.R-project.org/.

Rademaker, M.E. and Schuberth, F. (2020), "cSEM: composite-based structural equation modeling", Package version: 0.4.0.9000, available at: https://m-e-rademaker.github.io/cSEM/.

Rönkkö, M. and Cho, E. (2020), "An updated guideline for assessing discriminant validity", *Organizational Research Methods*, pp. 1-42, doi: 10.1177/1094428120968614.

Schuberth, F., Henseler, J. and Dijkstra, T.K. (2018), "Confirmatory composite analysis", *Frontiers in Psychology*, Vol. 9 No. 2541, doi: 10.3389/fpsyg.2018.02541.

Sijtsma, K. (2009), "On the use, the misuse, and the very limited usefulness of cronbach's alpha", *Psychometrika*, Vol. 74 No. 1, pp. 107-142.

Venables, W.N. and Ripley, B.D. (2002), *Modern Applied Statistics with S*, 4th ed., Springer, New York.

Voorhees, C.M., Brady, M.K., Calantone, R. and Ramirez, E. (2016), "Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies", *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 119-134.

Wold, H. (1982), "Soft modeling: the basic design and some extensions", in Jöreskog, K.G. and Wold, H. (Eds), *Systems under Indirect Observation: Causality, Structure, Prediction Part II*, North-Holland, Amsterdam, pp. 1-54.

Zumbo, B.D., Gadermann, A.M. and Zeisser, C. (2007), "Ordinal versions of coefficients alpha and theta for Likert rating scales", *Journal of Modern Applied Statistical Methods*, Vol. 6 No. 1, p. 4.

About the authors

Ellen Roemer is Professor of Market Research and International Marketing at Hochschule Ruhr West, University of Applied Sciences, Mülheim an der Ruhr, Germany. Her research interests focus on the analysis of customer relationships, with special attention toward the adoption and acceptance of eco-innovations. She complements her theoretical work with empirical studies using qualitative and quantitative research designs. Noteworthily, she uses partial least squares (PLS) path-modeling or

experimental designs with eye-tracking to answer research questions. She has published in leading Journals such as Industrial Data and Management Systems, Industrial Marketing Management, Journal of Marketing Management and Journal of Strategic Marketing.

Florian Schuberth is Assistant Professor at the Chair of Product–Market Relations in the Faculty of Engineering Technology at the University of Twente, the Netherlands. He obtained his PhD Degree in Econometrics at the Faculty of Business Management and Economics of the University of Würzburg, Germany. His main research interests are structural equation modeling, in particular on composite-based estimators and their enhancement. He is also co-inventor of confirmatory composite analysis (CCA). Florian Schuberth is the corresponding author and can be contacted at: f.schuberth@utwente.nl

Jörg Henseler is full professor and holds the Chair of Product–Market Relations in the Faculty of Engineering Technology at the University of Twente, the Netherlands. Additionally, he is Visiting Professor at NOVA Information Management School, Universidade Nova de Lisboa, Portugal, and Distinguished Invited Professor in the Department of Business Administration and Marketing at the University of Seville, Spain. His broad-ranging research interests encompass empirical methods of marketing and design research, as well as the management of design, products, services and brands. He is co-inventor of consistent partial least squares (PLSc), the heterotrait–monotrait ratio of correlations (HTMT) and confirmatory composite analysis (CCA). He is a highly cited researcher according to Web of Science; his work has been published in European Journal of Information Systems, International Journal of Research in Marketing, Journal of the Academy of Marketing Science, Journal of Supply Chain Management, MIS Quarterly, and Organizational Research Methods, among others. He chairs the Scientific Advisory Board of ADANCO, software for composite-based structural equation modeling (https://www.composite-modeling.com).