



Published in final edited form as:

*Psychol Methods*. 2011 June ; 16(2): 116–126. doi:10.1037/a0022657.

## Quantified Choice of RMSEAs for Evaluation and Power Analysis of Small Differences between Structural Equation Models

Libo Li and

UCLA Integrated Substance Abuse Programs, University of California, Los Angeles

Peter M. Bentler

Departments of Psychology and Statistics, University of California, Los Angeles

### Abstract

MacCallum, Browne, and Cai (2006) proposed a new framework for evaluation and power analysis of small differences between nested structural equation models (SEMs). In their framework, the null and alternative hypotheses for testing a small difference in fit and its related power analyses were defined by some chosen root-mean-square error of approximation (RMSEA) pairs. In this article, we develop a new method that quantifies those chosen RMSEA pairs and allows a quantitative comparison of them. Our method proposes the use of single RMSEA values to replace the choice of RMSEA pairs for model comparison and power analysis, thus avoiding the differential meaning of the chosen RMSEA pairs inherent in the approach of MacCallum et al. (2006). With this choice, the conventional cutoff values in model overall evaluation can directly be transferred and applied to the evaluation and power analysis of model differences.

### Keywords

structural equation model; model comparison; RMSEA; equi-discrepancy line; power analysis

---

Comparison of nested structural equation models (SEMs) was traditionally based on the null hypothesis that competing models have the same fit in population. By far, the classical normal theory-based likelihood ratio (NTLR) test (e.g., Jöreskog, 1971; Steiger, Shapiro, & Browne, 1985) is the most common method used for testing this null hypothesis. Unquestionably, the power of the NTLR test to detect any difference in fit between competing models should be an important issue for this null hypothesis testing.

Despite the popularity of this approach, MacCallum, Browne, and Cai (2006) suggested that the traditional null hypothesis that competing models have the same fit in population is always false and will be rejected by the NTLR test in practice when the sample size ( $N = n + 1$ ) is large enough. Instead, they advocated that a null hypothesis of testing some small difference in fit between nested models should replace the traditional one for comparison of SEMs. Under this new null hypothesis, the restricted model can be selected if the NTLR test doesn't detect a *large enough* difference in fit between two nested models. Similarly, for this type of null

---

Correspondence concerning this article should be addressed to Libo Li, UCLA Integrated Substance Abuse Programs, 1640 S. Sepulveda Boulevard, Suite 200, Los Angeles, CA 90025. lilibo@ucla.edu.

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/pubs/journals/met](http://www.apa.org/pubs/journals/met)

hypothesis testing, the power of the NTLR test to detect a *large enough* difference in fit between two nested models should be an important issue, too.

MacCallum et al. (2006) proposed a unified approach to such model comparison and power analysis. In their approach, the null and alternative hypotheses needed for model comparison and power analysis are defined by pairs of root-mean-square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger & Lind 1980) values chosen for competing models. Under these newly defined hypotheses, the power of the NTLR test and the comparison of models by the NTLR test can be investigated.

In this article, we first review this new approach to model comparison and power analysis. Then we propose a new method to quantify the RMSEA pairs chosen for the model comparison and the power analyses in MacCallum et al. (2006). By this new method, the choice of single RMSEA values is suggested to replace the choice of RMSEA pairs to define the null and alternative hypotheses needed for model comparison and power analysis. Our new method is then compared with the choices of RMSEA pairs by MacCallum et al. (2006) and applied to some empirical examples in the literature, followed by a discussion at the end of the article.

## An Approach to Model Comparison and Power Analysis

Let us assume for the moment that the model is estimated by the normal theory-based maximum likelihood (ML) discrepancy function, which is defined as

$$F_{ML} = \log|\hat{\Sigma}| - \log|S| + \text{tr}(S\hat{\Sigma}^{-1}) - q,$$

where  $S$  is the sample covariance matrix,  $\hat{\Sigma}$  is the covariance matrix implied by the model, and  $q$  is the number of observed variables. Let two models  $M_1$  and  $M_2$  have degrees of freedom  $\nu_1$  and  $\nu_2$  respectively and  $M_2$  be nested in  $M_1$ . Let  $\hat{F}_1$  and  $\hat{F}_2$ , respectively, be the minimized  $F_{ML}$  values of  $M_1$  and  $M_2$  in sample. Let  $F_1$  and  $F_2$  be the corresponding population values of  $\hat{F}_1$  and  $\hat{F}_2$  respectively. Suppose for the moment that the data are normal, and two nested models are not badly specified and satisfy the so-called population drift assumptions (see Steiger et al., 1985, for detail). The well-known NTLR statistic  $T_{ML} = n\hat{F}_2 - n\hat{F}_1$  will asymptotically follow the noncentral  $\chi^2_{\nu_2 - \nu_1, nF_2 - nF_1}$  distribution with degrees of freedom  $\nu_2 - \nu_1$  and noncentrality parameter  $nF_2 - nF_1$ .

In SEM, analysis of statistical power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 = 0$ , the null hypothesis of no difference in fit between nested models, is a traditional issue. For this power analysis, because the value of  $F_2 - F_1$  is generally unknown, we need an alternative hypothesis  $H_1 : F_2 - F_1 = \delta$ , where  $\delta$  is a positive chosen value. Then  $T_{ML}$  will have an approximate central  $\chi^2_{\nu_2 - \nu_1}$  distribution with degrees of freedom  $\nu_2 - \nu_1$  under  $H_0$  and an approximate noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta}$  distribution with degrees of freedom  $\nu_2 - \nu_1$  and noncentrality parameter  $n\delta$  under  $H_1$ . Let us use  $\alpha = .05$ . The critical value for  $T_{ML}$  to reject  $H_0$  will be the 95 percent quantile of the central  $\chi^2_{\nu_2 - \nu_1}$  distribution. Let us denote this critical value by  $\chi^2_{\nu_2 - \nu_1, .95}$ . The approximate power of  $T_{ML}$  to reject  $H_0$  under  $H_1$  is the probability of  $T_{ML}$  greater than  $\chi^2_{\nu_2 - \nu_1, .95}$  under  $H_1$ . For this power analysis, given  $\nu_2 - \nu_1$ ,  $n$ , and  $\alpha$ , the distribution of  $T_{ML}$  and the critical value  $\chi^2_{\nu_2 - \nu_1, .95}$  are fixed under  $H_0$ . However, a value must be chosen for  $\delta$  to define the noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta}$  distributions of  $T_{ML}$  under  $H_1$ .

MacCallum et al. (2006) defined the value of  $\delta$  as follows. Let  $r_1$  and  $r_2$  denote the population RMSEAs of  $M_1$  and  $M_2$  respectively and be defined as  $r_1 = \sqrt{F_1/\nu_1}$  and  $r_2 = \sqrt{F_2/\nu_2}$ . Then  $F_2 - F_1$  can be expressed as

$$F_2 - F_1 = \nu_2 r_2^2 - \nu_1 r_1^2. \quad (1)$$

Mimicking the expression of  $F_2 - F_1$  in (1), MacCallum et al. (2006) defined

$$\delta = \nu_2 r_{2\delta}^2 - \nu_1 r_{1\delta}^2, \quad (2)$$

where  $r_{1\delta}$  and  $r_{2\delta}$  are some RMSEA values chosen for  $M_1$  and  $M_2$ , respectively, to define  $\delta$  in  $H_1$  for the power analysis.

MacCallum et al. (2006) stated that  $H_0 : F_2 - F_1 = 0$  is too restrictive and always would be false in practice. As an alternative, they advocated the *good enough principle* (Serlin & Lapsley, 1985) for model comparison, and proposed a new null hypothesis of the form

$$H_0 : F_2 - F_1 \leq \delta_*, \quad (3)$$

where  $\delta_*$  is some chosen small positive value. They argued that this null hypothesis of small difference in fit between nested models is more realistic than the null hypothesis of no difference in fit between nested models for SEM model comparison in practice. Under this null hypothesis,  $M_2$  can be selected as long as the difference in fit between  $M_1$  and  $M_2$  is *small enough*. Under this null hypothesis, the critical value for  $T_{ML}$  to reject  $H_0$  now becomes the 95 percent quantile of the noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta_*}$  distribution and can be denoted by  $\chi^2_{\nu_2 - \nu_1, .95, n\delta_*}$ . If  $T_{ML}$  is greater than  $\chi^2_{\nu_2 - \nu_1, .95, n\delta_*}$ ,  $M_1$  would be preferred. Otherwise,  $M_2$  would be preferred. MacCallum et al. (2006) defined

$$\delta_* = \nu_2 r_{2\delta_*}^2 - \nu_1 r_{1\delta_*}^2, \quad (4)$$

where  $r_{1\delta_*}$  and  $r_{2\delta_*}$  are some RMSEA values chosen for  $M_1$  and  $M_2$ , respectively, to determine the value of  $\delta_*$ .

As with the power of rejecting  $H_0 : F_2 - F_1 = 0$ , the statistical power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 \leq \delta_*$  should be an important issue for testing the null hypothesis  $H_0 : F_2 - F_1 \leq \delta_*$ . Notice that the power of  $T_{ML}$  in this context no longer refers to the probability of detecting any difference in fit between competing models as in traditional power analysis. Instead, it now becomes the probability of detecting a *large enough* difference in fit between models.

For this type of power analysis, an alternative hypothesis  $H_1 : F_2 - F_1 = \delta_{**}$  is needed. Under this alternative hypothesis,  $T_{ML}$  has an approximate noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta_{**}}$  distribution. The approximate power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 \leq \delta_*$  is the probability of  $T_{ML}$  greater than  $\chi^2_{\nu_2 - \nu_1, .95, n\delta_*}$  under the alternative hypothesis. MacCallum et al. (2006) again defined  $\delta_{**}$  in terms of a RMSEA pair as

$$\delta_{**} = \nu_2 r_{2\delta_{**}}^2 - \nu_1 r_{1\delta_{**}}^2, \quad (5)$$

where  $r_{1\delta^{**}}$  and  $r_{2\delta^{**}}$  are some RMSEA values chosen for  $M_1$  and  $M_2$ , respectively, to determine  $\delta^{**}$ . Notice that  $\delta^{**}$  here must be greater than  $\delta^*$  for this type of power analysis.

## A New Method for the Choice of RMSEAs

From the description above, it is clear that the choice of RMSEA pairs, such as  $(r_{1\delta}, r_{2\delta})$  in (2),  $(r_{1\delta^*}, r_{2\delta^*})$  in (4), and  $(r_{1\delta^{**}}, r_{2\delta^{**}})$  in (5), is crucial to the power analyses and model comparison methodology proposed in MacCallum et al. (2006). In this article, we propose a way to quantify those RMSEA pairs chosen for model comparison and power analysis and suggest a new method to replace the choice of RMSEA pairs to define  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  across models and conditions.

We use  $\delta_C$  to represent the chosen value for  $\delta$ ,  $\delta^*$  or  $\delta^{**}$ , and use  $(r_{1C}, r_{2C})$  to represent the chosen values for the defining RMSEA pairs in (2), (4), and (5). Then the equations in (2), (4), and (5) can be expressed in a general equation, that is,

$$\delta_C = \nu_2 r_{2C}^2 - \nu_1 r_{1C}^2. \quad (6)$$

We use Figure 1 to illustrate our idea. Let us assume for moment that  $\nu_1 = 20$  and  $\nu_2 = 22$ . Let the horizontal and vertical axes in Figure 1 represent  $r_{1C}$  and  $r_{2C}$  respectively. Then in Figure 1 the area of any reasonable  $(r_{1C}, r_{2C})$  pair to define  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  should be above Line 0, where Line 0 is the line  $r_{2C} = \sqrt{\nu_1/\nu_2} \cdot r_{1C}$  with  $\delta_C = 0$ . This is because the minimum value of  $F_2 - F_1$  is zero and  $\delta_C$  defined by  $(r_{1C}, r_{2C})$  must be greater than zero. In addition, Line 0 clearly should be lower than the diagonal line since  $M_2$  is assumed to be nested in  $M_1$  and  $\sqrt{\nu_1/\nu_2} < 1$ .

Let  $p_1$  denote the point  $(r_{1C}, r_{2C}) = (.06, .08)$  in Figure 1 and be our choice of RMSEA pair, for example for  $(r_{1\delta^{**}}, r_{2\delta^{**}})$ . There should exist one line, Line A in Figure 1, which consists of a series of points  $(r_{1C'}, r_{2C'})$  in Figure 1 and satisfying the following equality

$$\delta_C = \nu_2 r_{2C}^2 - \nu_1 r_{1C}^2 = \nu_2 r_{2C'}^2 - \nu_1 r_{1C'}^2. \quad (7)$$

Similarly, when the point  $p_2 = (.07, .10)$  is chosen for  $(r_{1\delta^{**}}, r_{2\delta^{**}})$ , Line B exists in Figure 1 by (7). In fact, to any RMSEA pair chosen for such as  $(r_{1\delta}, r_{2\delta})$  in (2),  $(r_{1\delta^*}, r_{2\delta^*})$  in (4), and  $(r_{1\delta^{**}}, r_{2\delta^{**}})$  in (5), there will always exist a line in the figure consisting of infinitely many RMSEA pairs satisfying (7). Let us call this line the equi-discrepancy line because all RMSEA pairs on this line have the same value of  $\delta_C$  and represent the same degree of discrepancy between two nested models.

It is not hard to see that Line 0 is also a line satisfying (7), where all RMSEA pairs represent the same zero discrepancy between nested models. Suppose that the points on Line 0 are chosen as  $(r_{1C}, r_{2C})$  for model comparison. The null hypothesis of no difference in fit between nested models can be redefined as  $H_0: F_2 - F_1 = \nu_2 r_{2C}^2 - \nu_1 r_{1C}^2 = \nu_2 r_{2C'}^2 - \nu_1 r_{1C'}^2 = 0$  and can be interpreted as a null hypothesis of whether the true RMSEA pair  $(r_1, r_2)$  is on Line 0 or not. In addition, combining the identity  $F_2 - F_1 = \nu_2 r_2^2 - \nu_1 r_1^2$  in (1) with (3) and (4), we obtain a null hypothesis equivalent to (3), that is,

$$H_0: \nu_2 r_2^2 - \nu_1 r_1^2 \leq \delta_* = \nu_2 r_{2\delta_*}^2 - \nu_1 r_{1\delta_*}^2. \quad (8)$$

By (8), testing the null hypothesis of small difference in fit between nested models can be interpreted as testing if the true RMSEA pair  $(r_1, r_2)$ , which we don't know exactly, is inside the area between the equi-discrepancy line defined by the chosen RMSEA pair  $(r_{1\delta^*}, r_{2\delta^*})$  and Line 0 or not. When the chosen RMSEA pair  $(r_{1\delta^*}, r_{2\delta^*})$  represents a higher equi-discrepancy line, the area of testing is larger,  $\delta^*$  is larger, and a greater discrepancy is allowed between nested models. In contrast, when  $(r_{1\delta^*}, r_{2\delta^*})$  represents a lower line, the area of testing is smaller,  $\delta^*$  is smaller and less discrepancy is allowed.

This is also the case for power analysis. When the chosen RMSEA pair  $(r_{1C}, r_{2C})$  for defining  $\delta$  or  $\delta^{**}$  in the alternative hypothesis represents a higher equi-discrepancy line,  $T_{ML}$  will have more power of rejection given  $n$ ,  $\nu_2 - \nu_1$ ,  $\alpha$  and a fixed null hypothesis. Otherwise, it will have less power of rejection.

Furthermore, no matter whether a RMSEA pair is chosen for model comparison or power analysis, its corresponding equi-discrepancy line in Figure 1 by (7) will cross the vertical axis and have a point  $(r_{1c'_0}, r_{2c'_0})$  for which  $r_{1c'_0} = 0$ . In Figure 1, for example,  $(r_{1c'_0}, r_{2c'_0}) = (0, .056)$  for Line A, and  $(r_{1c'_0}, r_{2c'_0}) = (0, .074)$  for Line B, illustrate this phenomenon. Combining  $(r_{1c'_0}, r_{2c'_0})$  with (6) and (7), we obtain

$$\delta_c = \nu_2 r_{2c}^2 - \nu_1 r_{1c}^2 = \nu_2 r_{2c'_0}^2. \quad (9)$$

By (9), any discrepancy defined by a chosen RMSEA pair  $(r_{1C}, r_{2C})$  for model comparison or power analysis is equivalent to a discrepancy between the saturated model and a close-fitting model with degrees of freedom  $\nu_2$  and true  $RMSEA = r_{2c'_0}$ . For example,  $p_1$  on Line A can be considered to represent an overall discrepancy of a model with degrees of freedom 22 and true  $RMSEA = .056$  against the saturated model, while for  $p_2$  on Line B, this close-fitting model has degrees of freedom 22 and a true  $RMSEA = .074$ . By (9), we translate the choice of  $(r_{1C}, r_{2C})$  for model comparison and power analysis into an equivalent choice of  $r_{2c'_0}$  for a close-fitting reference model with degrees of freedom  $\nu_2$ .

Given this equivalence, we can be free from a two-dimensional choice of  $(r_{1C}, r_{2C})$  and instead define  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  for model comparison and power analysis based on the choice of  $r_{2c'_0}$ . By definition, RMSEA indicates discrepancy per degree of freedom. Choosing  $r_{2c'_0}$  to define  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  would retain the same meaning for  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  in spite of the change of  $\nu_1$  or  $\nu_2$ . Specifically, even though the value of  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  and the shape of the corresponding equi-discrepancy line may change with  $\nu_1$  or  $\nu_2$  by (7) and (9), the  $\delta$ ,  $\delta^*$  or  $\delta^{**}$  defined by a specific  $r_{2c'_0}$  means the same average discrepancy over degrees of freedom across studies.

In contrast, this is not the case when a two-dimensional choice of  $(r_{1C}, r_{2C})$  is used for defining  $\delta$ ,  $\delta^*$  or  $\delta^{**}$ . For example, MacCallum et al. (2006) selected  $r_{1\delta^*} = .05$  and  $r_{2\delta^*} = .06$  for  $\delta^*$  in (3) to test small difference between models for two published empirical studies: Joireman, Anderson, and Strathman (2003) and Dunkley, Zuroff, and Blankstein (2003). In Joireman et al. (2003),  $\nu_1 = 10$  and  $\nu_2 = 12$ . By (9), with  $r_{1\delta^*} = .05$  and  $r_{2\delta^*} = .06$ , the corresponding  $r_{2c'_0} = \sqrt{(12 \times .06^2 - 10 \times .05^2)/12} = .039$ . In Dunkley et al. (2003),  $\nu_1 = 337$  and  $\nu_2 = 338$ . With  $r_{1\delta^*} = .05$  and  $r_{2\delta^*} = .06$ ,  $r_{2c'_0} = \sqrt{(338 \times .06^2 - 337 \times .05^2)/338} = .033$  correspondingly. By

our method, we find that the values of  $\delta_*$  defined by  $r_{1\delta_*} = .05$  and  $r_{2\delta_*} = .06$  as in MacCallum et al. (2006) vary in terms of  $F_{2c'_0}$  and hence have different meaning across these two studies. The same phenomenon could happen to  $\delta$  or  $\delta_{**}$  when a constant RMSEA pair ( $r_{1C}$ ,  $r_{2C}$ ) is selected for the power analyses across studies.

In addition to eliminating this differential meaning of RMSEA pairs inherent in the approach of MacCallum et al. (2006), the choice of  $F_{2c'_0}$  also allows us to utilize the RMSEA cutoff criteria, established in SEM for a close-fitted reference model, for our model comparison and power analyses. It is well-known that RMSEAs equal to .056 or .074 may represent discrepancies between some mildly misspecified models and the corresponding saturated models. Thus,  $p_1$  and  $p_2$  by their corresponding  $F_{2c'_0}$  equal to .056 and .074 respectively can be considered to allow too much discrepancy to define  $\delta_*$  in (3) for small difference evaluation. In SEM, .05 is the widely used RMSEA cutoff value for model close fit (e.g., Browne & Cudeck, 1993). It can be used as  $F_{2c'_0}$  to define  $\delta_*$  for model comparison.

MacCallum, Browne, and Sugawara (1996) used  $F_{2c'_0} = .05$  to define the alternative hypothesis and estimated the power, or the minimum  $N$ , of the likelihood ratio test to reject the exact fit of SEM models with degrees of freedom  $\nu_2$ . In this article, by (9) and the equi-discrepancy lines above, we demonstrated the connection between choices of  $F_{2c'_0}$  for model overall evaluation and for model comparison. This connection also bridges our power analysis with MacCallum et al. (1996). By the same rationale, we could use  $F_{2c'_0} = .05$  to define  $\delta$  in the alternative hypothesis and then estimate the power, or the minimum  $N$ , of  $T_{ML}$  to reject the null hypothesis of no difference in fit between nested models.

Moreover, with  $\delta_*$  defined by such as  $F_{2c'_0} = .05$  in the null hypothesis, the  $F_{2c'_0}$  for defining  $\delta_{**}$  in the alternative hypothesis should be greater than .05. With such choice of  $F_{2c'_0}$ , the corresponding  $\delta_{**}$  may represent some degree of discrepancy between a mildly misspecified model and the saturated model. MacCallum et al. (1996) used  $F_{2c'_0} = .08$  for the alternative hypothesis and estimated the power of the likelihood ratio test to reject model close fit.

Similarly, we could use  $F_{2c'_0} = .08$  to define  $\delta_{**}$  and then estimate the power of  $T_{ML}$  to reject the null hypothesis of a small difference in fit between nested models.

Of course, we need to emphasize here that the cutoff values above borrowed from the model overall evaluation are merely guidelines. We have no intention to insist any constant RMSEA value for all situations. For example, some other values less than .05 such as .033 or .039 mentioned above can be selected as  $F_{2c'_0}$  for model comparison too. MacCallum et al. (2006) provided a good discussion on how these cutoff criteria for model overall evaluation may not consistently yield valid conclusions in practice. In total, the choices of  $F_{2c'_0}$  for model difference evaluation and power analysis still depend upon the scrutiny of investigators for their particular application. However, no matter what values they choose, our method is still valid and there would be no contradiction between their choices and our method.

Another interesting point we should mention here is that the interval between Line 0 and any other equi-discrepancy line always shrinks as  $r_{1C}$  increases as in Figure 1. Substantively, this means that under the same tolerable discrepancy for two nested models, the model comparison as in (8) allows more restriction or parsimony (larger  $r_2 - r_1$ ) when the general model contains less misspecification (smaller  $r_1$ ), or equivalently, less restriction (smaller  $r_2 - r_1$ ) when the general model becomes less trustable (larger  $r_1$ ). Although we do not know the values of  $r_1$



and  $r_2$ , the model comparison as in (8) automatically sets a corresponding standard for  $r_1$  and  $r_2$  by the equi-discrepancy lines.

## Relevant Developments

The equi-discrepancy line above is defined by a series of infinite equivalent alternatives to an arbitrary pair of RMSEA values. In fact, the equivalence of infinitely many RMSEA pairs is not a completely new idea. MacCallum, Lee, and Browne (2010) used a series of infinite alternatives to an arbitrary RMSEA pair to define their isopower contour. For any RMSEA pair on the isopower contour, the power of the likelihood ratio statistic to reject model overall fit (either exact or close fit) remains the same. The isopower contours in their Figures 9 and 10 look similar to our equi-discrepancy lines in Figure 1.

However, we need to point out that the equivalence of their infinite alternatives is defined by a power calculation which requires (among other things) a choice of discrepancy function, a test statistic, the distribution of the test statistic under the null hypothesis, the distribution of the test statistic under the alternative hypothesis, sample size  $N$ , and a choice of  $\alpha$ . Each point on the isopower contour represents the power of a statistic for model overall evaluation. On the other hand, the equivalence of our infinite alternatives requires none of these and is defined by (7) only. The points on our equi-discrepancy lines do not yield a plot for the power of a test statistic without further information.<sup>1</sup> Although coming from different directions, the idea of the equivalence of infinitely many RMSEA pairs is shared by the two studies and may be applicable to other SEM contexts beyond these studies in the future.

## Comparison and Application to Issues in SEM

Our new method quantifies the RMSEA pairs chosen for the model comparison and the power analyses proposed by MacCallum et al. (2006) and allows discrepancies defined by different RMSEA pairs to be compared with each other quantitatively in a common metric. By our method, the choice of a single  $r_{2c'_0}$  is suggested to replace the choice of a RMSEA pair ( $r_{1C}$ ,  $r_{2C}$ ) for the model comparison and the power analyses proposed by MacCallum et al. (2006). In this section, we reexamine the RMSEA pair choices for the examples in MacCallum et al. (2006) using our method and apply our choice of  $r_{2c'_0}$  to issues in SEM.

### Quantification of the RMSEA pair choices

MacCallum et al. (2006) suggested to use a large set of pairs of  $r_{1C}$  and  $r_{2C}$  in a reasonable range to define  $\delta$  or  $\delta^{**}$  for power analysis. For example, in their Table 1 and Table 2, given  $\nu_1 = 20$  and  $\nu_2 = 22$ , MacCallum et al. (2006) selected  $r_{1\delta}$  from .03 to .09 and  $r_{2\delta}$  from .04 to .10 to define  $\delta$  in the alternative hypothesis and estimated the power, or minimum  $N$ , of  $T_{ML}$  to reject the null hypothesis of no difference in fit between nested models. We apply our method to compare those choices and use (9) to calculate the  $r_{2c'_0}$  values for the pairs of  $r_{1\delta}$  and  $r_{2\delta}$  selected by MacCallum et al. (2006) in their tables. The R code used for the calculation is given in the Appendix and the calculated values are presented in our Table 1.

In our Table 1, eight pairs of  $r_{1\delta}$  and  $r_{2\delta}$  have  $r_{2c'_0}$  values less than .05, while all other twenty pairs have  $r_{2c'_0}$  values greater than .05. By our method, we would say that for this power analysis eight pairs of  $r_{1\delta}$  and  $r_{2\delta}$  suggested by MacCallum et al. (2006) may represent a discrepancy between a close-fitting model and the saturated model while all other twenty pairs may represent a larger discrepancy. Furthermore, in our Table 1, with the same  $\nu_1$  and  $\nu_2$ , the

<sup>1</sup>To further distinguish the difference between the isopower contour and our equi-discrepancy line, please see the Appendix.

RMSEA pairs at the higher end of the RMSEA scale such as (.08,.10) have larger  $r_{2c'_0}$  values and represent larger discrepancies than the pairs at the lower end of the RMSEA scale with the same difference of  $r_{1\delta}$  and  $r_{2\delta}$  such as (.03,.05) or (.04,.06). This phenomenon in turn accounts for the increasing statistical power in detecting model difference or the decreasing minimum  $N$  required for the same power along the diagonal bands of Table 1 and Table 2 in MacCallum et al. (2006).

### The choice of $r_{2c'_0}$ for issues in SEM

The choice of  $r_{2c'_0}$  was recommended in this article to replace the choice of RMSEA pairs to define  $\delta$ ,  $\delta_*$  and  $\delta_{**}$  for model comparison and power analysis. MacCallum et al. (2006) selected  $r_{1\delta} = 0.4$  and  $r_{2\delta} = .06$  to define  $\delta$  in the alternative hypothesis  $H_1 : F_2 - F_1 = \delta$  and conduct power analysis for two published empirical studies (Manne & Glassman, 2000; Sadler & Woody, 2003). In Sadler and Woody (2003),  $\nu_1 = 24$  and  $\nu_2 = 27$ . In Manne and Glassman (2000),  $\nu_1 = 19$  and  $\nu_2 = 26$ . Suppose that we choose a single value  $r_{2c'_0} = .05$  to define  $\delta$  for power analysis. By (9),  $\delta = \nu_2 r_{2c'_0}^2 = 27 \times .05^2 = .0675$  for Sadler and Woody (2003) and  $\delta = \nu_2 r_{2c'_0}^2 = 26 \times .05^2 = .065$  for Manne and Glassman (2000). Given these  $\delta$ s, we can calculate the power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 = 0$  in each study. As mentioned before,  $T_{ML}$  would have an approximate central  $\chi_{\nu_2 - \nu_1}^2$  distribution under  $H_0 : F_2 - F_1 = 0$  and have an approximate noncentral  $\chi_{\nu_2 - \nu_1, n\delta}^2$  distribution under  $H_1 : F_2 - F_1 = \delta$ . With  $\alpha = .05$ , this power is the probability of  $T_{ML}$  to exceed the critical value  $\chi_{\nu_2 - \nu_1, .95}^2$  under  $H_1 : F_2 - F_1 = \delta$ . In Sadler and Woody (2003),  $\nu_2 - \nu_1 = 3$  and  $N = 112$ . Given  $\alpha = .05$ , the power is .62. In Manne and Glassman (2000),  $\nu_2 - \nu_1 = 7$  and  $N = 191$ . Given  $\alpha = .05$ , the power is .73. The steps of power calculation for the two studies are coded in R and given in the Appendix.

Both Sadler and Woody (2003) and Manne and Glassman (2000) failed to detect a significant difference between their nested models by  $T_{ML}$ . As a result, both studies selected the restricted models after comparison. Given the power analyses above, the nonsignificance of  $T_{ML}$  in both studies, as mentioned by MacCallum et al. (2006), may partially be due to the moderate power of rejection in both studies. Given this possibility, we decided to calculate the minimum  $N$  to achieve a .80 power of rejection. With  $\alpha = .05$ , this minimum  $N$  is the minimum sample size that the probability of  $T_{ML}$  to exceed  $\chi_{\nu_2 - \nu_1, .95}^2$  under  $H_1 : F_2 - F_1 = \delta$  is equal to or greater than .80. In Sadler and Woody (2003), given  $\nu_2 - \nu_1 = 3$  and  $\delta$  defined by  $r_{2c'_0} = .05$  above, we find that this minimum  $N$  is 163. In Manne and Glassman (2000), given  $\nu_2 - \nu_1 = 7$  and  $\delta$  defined by  $r_{2c'_0} = .05$  above, we find that this minimum  $N$  is 222. The steps for the calculation of minimum  $N$ s are coded in R and given in the Appendix. As in MacCallum et al. (2006), our minimum  $N$ s calculated here suggest that more additional subjects are needed for  $T_{ML}$  in both studies to reach a rejection power of .80.

We demonstrated before that when the  $\delta_*$  in  $H_0 : F_2 - F_1 = \delta_*$  is defined by  $r_{1\delta_*} = .05$  and  $r_{2\delta_*} = .06$  as in MacCallum et al. (2006), the corresponding  $r_{2c'_0}$  would be equal to .039 for Joireman et al. (2003) and be equal to .033 for Dunkley et al. (2003). Now we instead select a single value, for example  $r_{2c'_0} = .033$ , to define  $\delta_*$ . By (9),  $\delta_* = \nu_2 r_{2c'_0}^2 = 12 \times .033^2 = .013$  for Joireman et al. (2003). With  $\nu_2 - \nu_1 = 2$ ,  $\alpha = .05$  and  $N = 154$ , the critical value for rejecting  $H_0 : F_2 - F_1 = \delta_*$  is  $\chi_{\nu_2 - \nu_1, .95, n\delta_*}^2 = 10.84$ . The observed value of  $T_{ML} = 19.13$  yields a significant



level of  $p < .05$ , indicating rejection of the null hypothesis of small difference in fit for this study. For Dunkley et al. (2003),  $\delta^* = 338 \times .033^2 = .368$  by (9). With  $\nu_2 - \nu_1 = 1$ ,  $\alpha = .05$  and  $N = 163$ , the critical value is  $\chi^2_{\nu_2 - \nu_1, .95, n\delta^*} = 87.74$ . The observed value of  $T_{ML} = 3.84$  yields a nonsignificant level of  $p > .99$ , suggesting that the null hypothesis of small difference in fit is acceptable. The steps for the calculation of critical and  $p$ -values are coded in R and given in the Appendix.

Both Joireman et al. (2003) and Dunkley et al. (2003) observed the significance of  $T_{ML}$  when they compared a general model with a more parsimonious and interpretable model under the traditional approach. As a result, Joireman et al. (2003) selected the general model while Dunkley et al. (2003) selected the parsimonious one despite the significance of  $T_{ML}$ . Like the traditional one, our model comparison under the null hypothesis of small difference in fit supports the choice of the general model for Joireman et al. (2003). However, for Dunkley et al. (2003), it is different from the traditional approach and supports the parsimonious model, the choice of the authors in their article.

For our model comparisons above, a related issue is the power of  $T_{ML}$  to reject the null hypothesis of small difference in fit. In their example with  $\nu_1 = 20$  and  $\nu_2 = 22$ , MacCallum et al. (2006) selected  $r_1 \delta^{**} = .05$  and  $r_2 \delta^{**} = .08$  to define  $\delta^{**}$  in the alternative hypothesis  $H_1 : F_2 - F_1 = \delta^{**}$ . From our Table 1, we find the corresponding  $r_{2c'_0} = .064$ . As a result, we set  $r_{2c'_0} = .064$  to define  $\delta^{**}$  for our power analysis. By (9),  $\delta^{**} = \nu_2 r_{2c'_0}^2 = 12 \times .064^2 = .0492$  for Joireman et al. (2003) and  $\delta^{**} = \nu_2 r_{2c'_0}^2 = 338 \times .064^2 = 1.3844$  for Dunkley et al. (2003). Given these  $\delta^{**}$ s and the  $\delta^*$ s defined by  $r_{2c'_0} = .033$  above, we can calculate the power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 = \delta^*$  in the two studies. In this case,  $T_{ML}$  would have an approximate noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta^*}$  distribution under  $H_0 : F_2 - F_1 = \delta^*$  and have an approximate noncentral  $\chi^2_{\nu_2 - \nu_1, n\delta^{**}}$  distribution under  $H_1 : F_2 - F_1 = \delta^{**}$ . With  $\alpha = .05$ , this power is the probability of  $T_{ML}$  to exceed the critical value  $\chi^2_{\nu_2 - \nu_1, .95, n\delta^*}$  under  $H_1 : F_2 - F_1 = \delta^{**}$ . Our steps of power calculation for the two studies are coded in R and given in the Appendix. Our calculation shows that the power of rejection is .35 for Joireman et al. (2003) and is 1 for Dunkley et al. (2003). Notice that the degrees of freedom of competing models in Dunkley et al. (2003) are high. This may explain the higher power in their study with the same  $r_{2c'_0}$ . The  $r_{2c'_0}$  value may be adjusted downward given the high degrees of freedom of their models (see p. 32 of MacCallum et al., 2006).

Our power analysis results above strengthen the selection of the general model for Joireman et al. (2003) and the restricted model for Dunkley et al. (2003). In Joireman et al. (2003), the observed  $T_{ML}$  is significant even though the power for this study is just .35. In contrast, the observed  $T_{ML}$  in Dunkley et al. (2003) is nonsignificant even though the power is 1. Given the low power in Joireman et al. (2003), we calculate the minimum  $N$  for  $T_{ML}$  to reach a .80 power of rejection in this study. The steps for this calculation are coded in R and given in the Appendix. Our calculation shows that this minimum  $N$  is 555. Some additional subjects are needed for this study to gain a desirable power.

## Discussion and Recommendations

In this article, we followed the footsteps of MacCallum, Browne, and Cai (2006) and developed a new method for conducting model comparison and power analysis consistent with the approach of MacCallum et al. (2006). In this development, we defined an equi-discrepancy line by equation (7). Our equi-discrepancy line reminds investigators that there always exists

a series of infinite alternatives to their chosen RMSEA pair to equivalently define the null or alternative hypothesis for model comparison and power analysis as in MacCallum et al. (2006). With these infinite alternatives, testing no or a small difference between nested models can be graphically translated or understood as testing whether the true RMSEA pair of nested models falls on Line 0 or the region below the equi-discrepancy line in Figure 1.

Among these infinite alternatives, we also identified a unique RMSEA pair that is at the cross of the equi-discrepancy line and the vertical axis in Figure 1. With this unique RMSEA pair, any chosen RMSEA pair can be quantified into the RMSEA value of a reference SEM model. Consequently, for investigators, their choices of RMSEA pairs for model comparison and power analysis can be compared quantitatively. This comparison as in our Table 1 also provides a reminder to investigators that the RMSEA pairs at the different parts of the RMSEA scale, for example, (.03,.05) vs. (.08,.10) in our Table 1, would represent different discrepancies for model comparison and have different implications for power analysis even though they have the same difference of RMSEA values.

More importantly, given this quantification, the choice of RMSEA pairs in MacCallum et al. (2006) now is recommended to be replaced by the choice of a single RMSEA value instead. For model comparison and power analysis following MacCallum et al. (2006), this choice of single RMSEA value can eliminate the differential meaning of the RMSEA pairs inherent in the original approach of MacCallum et al. (2006). This choice also bridges the two branches of SEM model evaluation, model overall evaluation and model difference evaluation, and makes possible for the criteria established for model overall evaluation to be transferred and applied to the issues of SEM model difference evaluation and power analyses.

As a result, in our new procedure illustrated in the examples above, we propose to first choose some reasonable RMSEA values to define the null and alternative hypothesis for model comparison and power analysis as per MacCallum et al. (2006). These choices could be made based on some established criteria for model overall evaluation (e.g., Browne & Cudeck, 1993; MacCallum et al., 1996) or some other values upon the scrutiny of investigators. Once these defining RMSEA values are chosen, model comparison and power analysis as in MacCallum et al. (2006) can be conducted under the corresponding null and alternative hypotheses as illustrated in our examples above.

Of course, as a followup to MacCallum et al. (2006), our proposed method for model comparison and power analysis would inevitably inherit many issues encountered by MacCallum et al. (2006). These issues include assumptions such as  $T_{ML}$  to be central or non-central chi-square distributed, estimation methods other than the ML discrepancy function, limitations of RMSEA cutoff criteria for practical use, different approaches to model comparison (e.g., Satorra & Saris, 1985), etc. For all these issues, please refer to MacCallum et al. (2006) for a thorough discussion in the context of the general framework. In addition, in multisample SEM studies, due to constraints across groups, Steiger (1998) proposed a modification of the definition of RMSEA. That is,  $r_1 = \sqrt{G} \sqrt{F_1/\nu_1}$  and  $r_2 = \sqrt{G} \sqrt{F_2/\nu_2}$  for  $M_1$  and  $M_2$  respectively, where  $G$  is the number of groups. Under this definition, by dividing the right side of (2), (4), (5) or (6) by  $G$ , our proposed method could be still valid to define the value of  $\delta_C$  for model comparison and power analysis in MacCallum et al. (2006). Of course, RMSEA in this case doesn't have the interpretation as discrepancy per degree of freedom as pointed out by MacCallum et al. (2006).

MacCallum et al. (2006) pointed out that their framework can be generalized into some fit indices other than RMSEA. Correspondingly, our developments are also applicable to those fit indices. For example, let  $\gamma_{1C}$  and  $\gamma_{2C}$  denote the values of Steiger's (1989)  $\gamma$  chosen for  $M_1$  and  $M_2$  respectively. Then  $\delta_C$  can be defined as  $\delta_C = (p/2) \cdot (1/\gamma_{2C} - 1/\gamma_{1C})$  (see Kim,

2005). Similarly, let  $Mc_{1C}$  and  $Mc_{2C}$  denote the values of McDonald's (1989) fit index chosen for  $M_1$  and  $M_2$  respectively. Then  $\delta_C$  can be defined as  $\delta_C = -2\log(Mc_{2C}) + 2\log(Mc_{1C})$  (see Kim, 2005). Following the same logic for RMSEAs in (7) and (9), we obtain

$$\delta_c = \frac{p}{2} \left( \frac{1}{\gamma_{2c}} - \frac{1}{\gamma_{1c}} \right) = \frac{p}{2} \left( \frac{1}{\gamma_{2c'}} - \frac{1}{\gamma_{1c'}} \right) = \frac{p}{2} \left( \frac{1}{\gamma_{2c'_0}} - 1 \right)$$

and

$$\delta_c = -2\log(Mc_{2c}) + 2\log(Mc_{1c}) = -2\log(Mc_{2c'}) + 2\log(Mc_{1c'}) = -2\log(Mc_{2c'_0})$$

where  $\gamma_{1C}$  and  $\gamma_{2C'}$  constitute infinite alternatives to  $\gamma_{1C}$  and  $\gamma_{2C}$ ,  $Mc_{1C'}$  and  $Mc_{2C'}$  constitute infinite alternatives to  $Mc_{1C}$  and  $Mc_{2C}$ , and  $\gamma_{2c'_0}$  and  $Mc_{2c'_0}$  are the corresponding values of Steiger's  $\gamma$  and McDonald's fit index for the reference model. Again, by these infinite alternatives, different equi-discrepancy lines can be drawn. By these equations, the choice of  $\gamma_{2c'_0}$  and  $Mc_{2c'_0}$  can replace the choice of pairs of Steiger's  $\gamma$  and McDonald's fit index to define  $\delta$ ,  $\delta^*$  and  $\delta^{**}$  in the null and alternative hypotheses for model comparison and power analysis. Similarly, some conventional cutoff values of these fit indices (e.g., Hu & Bentler, 1999) can be used for the choice of  $\gamma_{2c'_0}$  or  $Mc_{2c'_0}$  for model comparison and power analysis.

## Acknowledgments

Research supported in part by grants DA00017, DA01070 and P30 DA016383 from the National Institute on Drug Abuse. Part of this article has been presented at the 2008 annual international meeting of Psychometric Society, Durham, NH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse.

## Appendix

1. The R code to calculate the  $r_{2c'_0}$  values for the pairs of  $r_{1\delta}$  and  $r_{2\delta}$  selected by MacCallum, Browne, and Cai (2006) in their Table 1 and Table 2:

```
dfM1<-20 # degrees of freedom of M1
dfM2<-22 # degrees of freedom of M2
r1<-seq(.03,.09,.01) # RMSEA choice for M1
r2<-r1+.01 # RMSEA choice for M2
Fn<-matrix(0,length(r2),length(r1))
for (i in 1:length(r2)) {
  for (j in 1:i) {
    Fn[i,j]<-sqrt((r2[i]^2*dfM2-r1[j]^2*dfM1)/dfM2)
  }
}
print(Fn,digits=2)
```

2. The R code to calculate the power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 = \delta$  for Sadler and Woody (2003):

```

n<-111 # (Sample size-1)
dfM1<-24 # degrees of freedom of M1
dfM2<-27 # degrees of freedom of M2
delta<-dfM2*.05^2 # define delta in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1) # Critical value under the null hypothesis
power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta)
print(power,digits=2)

```

For Manne and Glassman (2000):

```

n<-190 # (Sample size-1)
dfM1<-19 # degrees of freedom of M1
dfM2<-26 # degrees of freedom of M2
delta<-dfM2*.05^2 # define delta in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1) # Critical value under the null hypothesis
power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta)
print(power,digits=2)

```

3. The R code to calculate the minimum  $N$  for  $T_{ML}$  to achieve the .80 power of rejecting  $H_0 : F_2 - F_1 = 0$  in Sadler and Woody (2003):

```

dfM1<-24 # degrees of freedom of M1
dfM2<-27 # degrees of freedom of M2
delta<-dfM2*.05^2 # define delta in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1) # Critical value under the null hypothesis
for (n in 1:100000) {
  power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta)
  if (power>=.80) break # reach the desirable power
}
N<-n+1
print(N)

```

For Manne and Glassman (2000):

```

dfM1<-19 # degrees of freedom of M1
dfM2<-26 # degrees of freedom of M2
delta<-dfM2*.05^2 # define delta in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1) # Critical value under the null hypothesis
for (n in 1:100000) {
  power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta)
  if (power>=.80) break # reach the desirable power
}
N<-n+1
print(N)

```

4. The R code to calculate the  $p$ -value for  $T_{ML}$  to reject  $H_0 : F_2 - F_1 \geq \delta_*$  in Joireman, Anderson, and Strathman (2003):

```
Tml<-19.13 # the observed value of the likelihood ratio test
n<-153 # (Sample size-1)
dfM1<-10 # degrees of freedom of M1
dfM2<-12 # degrees of freedom of M2
delta<-dfM2*.033^2 # define delta* in the null hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1,n*delta) # Critical value under H0
pvalue<-1-pchisq(Tml,df=dfM2-dfM1,ncp=n*delta)
print(cv,digits=4)
print(pvalue,digits=2)
```

For Dunkley, Zuroff, and Blankstein (2003):

```
Tml<-3.84 # the observed value of the likelihood ratio test
n<-162 # (Sample size-1)
dfM1<-337 # degrees of freedom of M1
dfM2<-338 # degrees of freedom of M2
delta<-dfM2*.033^2 # define delta* in the null hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1,n*delta) # Critical value under H0
pvalue<-1-pchisq(Tml,df=dfM2-dfM1,ncp=n*delta)
print(cv,digits=4)
print(pvalue,digits=2)
```

5. The R code to calculate the power of  $T_{ML}$  to reject  $H_0 : F_2 - F_1 \geq \delta_*$  for Joireman et al. (2003):

```
n<-153 # (Sample size-1)
dfM1<-10 # degrees of freedom of M1
dfM2<-12 # degrees of freedom of M2
delta1<-dfM2*.033^2 # define delta* in the null hypothesis
delta2<-dfM2*.064^2 # define delta** in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1,n*delta1) # Critical value under H0
power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta2)
print(power,digits=2)
```

For Dunkley et al. (2003):

```
n<-162 # (Sample size-1)
dfM1<-337 # degrees of freedom of M1
dfM2<-338 # degrees of freedom of M2
delta1<-dfM2*.033^2 # define delta* in the null hypothesis
delta2<-dfM2*.064^2 # define delta** in the alternative hypothesis
alpha<-.05
cv<-qchisq(1-alpha,df=dfM2-dfM1,n*delta1) # Critical value under H0
power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta2)
```

```
print(power,digits=2)
```

6. The R code to calculate the minimum  $N$  for  $T_{ML}$  to achieve the .80 power of rejecting  $H_0 : F_2 - F_1 = \delta^*$  in Joireman et al. (2003):

```
dfM1<-10 # degrees of freedom of M1
dfM2<-12 # degrees of freedom of M2
delta1<-dfM2*.033^2 # define delta* in the alternative hypothesis
delta2<-dfM2*.064^2 # define delta** in the alternative hypothesis
alpha<-.05
for (n in 1:100000) {
  cv<-qchisq(1-alpha,df=dfM2-dfM1,ncp=n*delta1) # Critical value under H0
  power<-1-pchisq(cv,df=dfM2-dfM1,ncp=n*delta2)
  if (power>=.80) break # reach the desirable power
}
N<-n+1
print(N)
```

7. The difference between the isopower contour of MacCallum, Lee and Browne (2010) and our equi-discrepancy line:

The purpose of the description below is to help researchers distinguish those two concepts in the two studies and facilitate the application of both of them. As a result, the first difference we want to point out is that the context of the isopower contour of MacCallum et al. (2010) is model overall evaluation (exact or close fit). Instead, our equi-discrepancy line applies to comparison of any pair of nested SEM models. For further illustration, we consider a special situation where the less restricted model  $M_1$  is a saturated model with  $\nu_1 = 0$ ,  $F_1 = 0$  and  $r_1 = 0$ , and the more restricted one  $M_2$  is a model with degrees of freedom, for example,  $\nu_2 = 20$  as assumed in Figure 9 of MacCallum et al. (2010)<sup>2</sup>. Notice that the isopower contour and the equi-discrepancy line are now discussed in the same context of overall evaluation of  $M_2$ .

Let  $(r_{1I}, r_{2I})$  be a pair of RMSEA values to define the null hypothesis  $H_0 : r_2 = r_{1I}$  and its alternative one  $H_1 : r_2 = r_{2I}$  for overall evaluation of  $M_2$  as in MacCallum et al. (2010). With a choice of  $\alpha$  and sample size  $N$ , the central or noncentral  $\chi^2$  distribution of  $T_{ML}$  under  $H_0$  and  $H_1$  can be determined and the power of  $T_{ML}$  to reject  $H_0$  under  $H_1$  can be calculated. MacCallum et al. (2010) argued that  $T_{ML}$  can reach the same rejection power when  $H_0$  and  $H_1$  are defined by another RMSEA pair among an infinite number of isopower alternatives to  $(r_{1I}, r_{2I})$ . Let  $\alpha = .05$ ,  $N = 200$ ,  $r_{1I} = .05$  and  $r_{2I} = .10$  as in Figure 9 of MacCallum et al. (2010). Then we calculated some of those isopower alternatives and replicated the isopower contour in their Figure 9 as Line 1 in our Figure 2<sup>3</sup>. The rejection power of  $T_{ML}$  under any set of  $H_0$  and  $H_1$  defined by the points on Line 1 would be the same as .84.

For model comparison and power analysis by our method,  $r_{1C}$  and  $r_{2C}$  are selected for  $M_1$  and  $M_2$  respectively to define  $\delta$  in  $H_1 : F_2 - F_1 = \delta$ , or  $\delta^*$  in  $H_0 : F_2 - F_1 = \delta^*$ , or  $\delta^{**}$  in  $H_1 : F_2 - F_1 = \delta^{**}$ . According to MacCallum et al. (2006),  $r_{1C}$  and  $r_{2C}$  should be some reasonable RMSEA values representing the misspecification of two nested models. When  $M_1$  is assumed to be a saturated model with  $\nu_1 = 0$ ,  $F_1 = 0$  and  $r_1 = 0$ , the only reasonable choice of RMSEA value for  $r_{1C}$  (and its alternative  $r_{1C'}$ )

<sup>2</sup>We thank James Steiger for this valuable suggestion.

<sup>3</sup>We thank Taehun Lee for providing their SAS code to replicate this isopower contour.



should be zero. Correspondingly,  $r_{2C}$  and its alternative  $r_{2C'}$  should be the same by the definition of the equi-discrepancy line in (7). Consequently, in the context of overall evaluation of  $M_2$  as in MacCallum et al. (2010), our equi-discrepancy line is no longer a line and reduces to a point on the vertical axis in Figure 2. For example, when .076 or .126 is selected for  $r_{2C}$ , our equi-discrepancy line reduces to the point  $p_1$  or  $p_2$  in Figure 2, respectively.

In fact, even when  $r_{1C}$  or  $r_{1C'}$  is not limited to zero and can be any value greater than zero,  $r_{2C}$  and  $r_{2C'}$  should still be the same to satisfy (7) because  $M_1$  is a saturated model and  $\nu_1 = 0$  at this time. For example, when  $\nu_1 = 0$ ,  $\nu_2 = 20$ , and the point  $p_1 = (.00, .076)$  is selected as  $(r_{1C}, r_{2C})$ ,  $r_{1C'}$  should always be .076 by (7) in spite of the value of  $r_{1C'}$ . As a result, the equi-discrepancy line in this case would be Line 2 in Figure 2, which is a straight line parallel to the horizontal axis. Similarly, when the point (.08, .126) is selected as  $(r_{1C}, r_{2C})$ , the equi-discrepancy line would be Line 3 in Figure 2, which crosses the vertical axis at  $p_2$ .

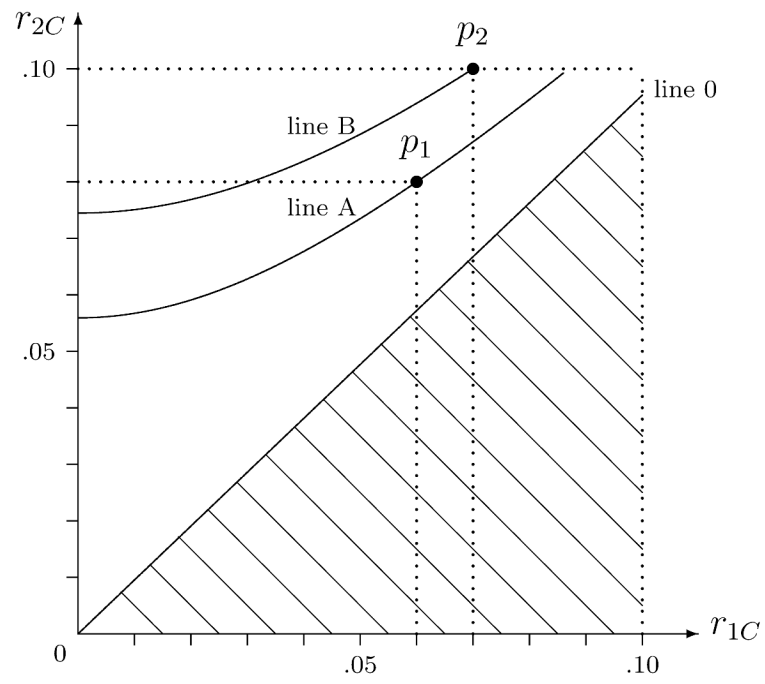
Furthermore, in spite of the value of  $(r_{1C}, r_{2C})$  or  $(r_{1C'}, r_{2C'})$  on the equi-discrepancy line, equation (7) now would reduce to  $\delta_c = \nu_2 r_{2C}^2$  because  $\nu_1 = 0$  and  $(r_{1C}, r_{2C'})$ . Given  $F_1 = 0$  now, substituting this  $\delta_c$  into the null and alternative hypotheses,  $H_1 : F_2 - F_1 = \delta$  could reduce to  $H_1 : r_2 = r_{2C}$ ,  $H_0 : F_2 - F_1 = \delta^*$  could reduce to  $H_0 : r_2 < r_{2C}$ , and  $H_1 : F_2 - F_1 = \delta^{**}$  could reduce to  $H_1 : r_2 > r_{2C}$ . Obviously, in the context of model overall evaluation, the points  $(r_{1C}, r_{2C})$  and  $(r_{1C'}, r_{2C'})$  on an equi-discrepancy line (if this line exists) equivalently define either  $r_{iI}$  of  $H_0 : r_2 < r_{iI}$  or  $r_{2I}$  of  $H_1 : r_2 > r_{2I}$  in MacCallum et al. (2010), instead of both.

Finally, when  $\nu_1$  becomes some positive values rather than zero and  $M_1$  is no longer a saturated model, the isopower contour of MacCallum et al. (2010) doesn't apply while our equi-discrepancy line still exists. We plotted the equi-discrepancy line when  $\nu_1$  is equal to 2, 10 and 19, as Line 4, Line 5 and Line 6 in Figure 2, respectively. The equi-discrepancy lines now become some curves similar to the isopower contour of MacCallum et al. (2010). However, they can not coincide with each other because the isopower contour does not exist in this situation. In addition, notice that the points on an equi-discrepancy line can equivalently define the null or alternative hypothesis for model comparison and its related power analysis. However, unlike the points on the isopower contours, they do not represent a plot of power for any test statistic without further information.

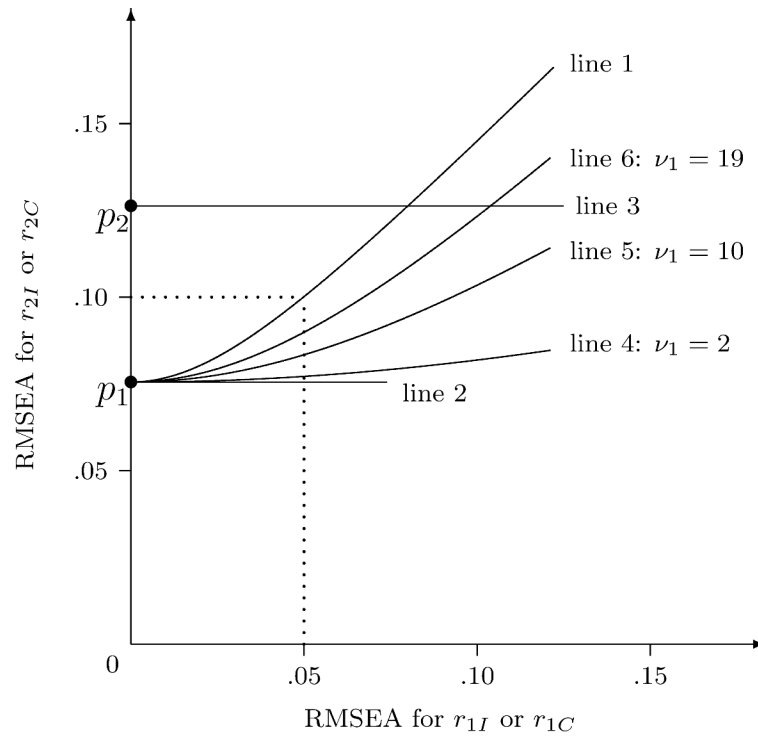
## References

- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-162.
- Dunkley DM, Zuroff DC, Blankstein KR. Self-critical perfectionism and daily affect: Dispositional and situational influences on stress and coping. *Journal of Personality and Social Psychology*. 2003; 84:234–252. [PubMed: 12518982]
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternative. *Structural Equation Modeling*. 1999; 6:1–55.
- Joireman J, Anderson J, Strathman A. The aggression paradox: Understanding links among aggression, sensation seeking, and the consideration of future consequences. *Journal of Personality and Social Psychology*. 2003; 84:1287–1302. [PubMed: 12793590]
- Jöreskog KG. Simultaneous factor analysis in several populations. *Psychometrika*. 1971; 36:409–426.
- Kim KH. The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling*. 2005; 12:368–390.
- MacCallum RC, Browne MW, Cai L. Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*. 2006; 11:19–35. [PubMed: 16594765]

- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1996; 1:130–149.
- MacCallum RC, Lee T, Browne MW. The issue of isopower in power analysis for tests of structural equation models. *Structural Equation Modeling*. 2010; 17:23–41.
- Manne S, Glassman M. Perceived control, coping efficacy, and avoidance coping as mediators between spouses unsupportive behaviors and cancer patients' psychological distress. *Health Psychology*. 2000; 19:155–164. [PubMed: 10762099]
- McDonald RP. An index of goodness-of-fit based on noncentrality. *Journal of Classification*. 1989; 6:97–103.
- Sadler P, Woody E. Is who you are who you're talking to? Interpersonal style and complementarity in mixed-sex interactions. *Journal of Personality and Social Psychology*. 2003; 84:80–96. [PubMed: 12518972]
- Satorra A, Saris WE. Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*. 1985; 50:83–90.
- Serlin RC, Lapsley DK. Rationality in psychological research: The good-enough principle. *American Psychologist*. 1985; 40:73–83.
- Steiger, JH. EzPATH: A supplementary module for SYSTAT and SYGRAPH. Evanston, IL: SYSTAT; 1989.
- Steiger JH. A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*. 1998; 5:411–419.
- Steiger, JH.; Lind, JC. Statistically-based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society; Iowa City, IA. 1980.
- Steiger JH, Shapiro A, Browne MW. On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*. 1985; 50:253–264.



**Figure 1.**  
The equi-discrepancy lines with  $\nu_1 = 20$  and  $\nu_2 = 22$



**Figure 2.**  
The equi-discrepancy line vs. the isopower contour with  $\nu_2 = 20$

Table 1

The  $r_{z'c'}$  for the selected pairs of  $r_{1C}$  and  $r_{2C}$  with  $\nu_1 = 20$  and  $\nu_2 = 22$ .

	$r_{1C}$									
$r_{2C}$	.03	.04	.05	.06	.07	.08	.09	.10	.09	.08
.04	.028									
.05	.041	.032								
.06	.053	.046	.036							
.07	.064	.059	.051	.040						
.08	.075	.070	.064	.056	.044					
.09	.085	.082	.076	.069	.060	.048				
.10	.096	.092	.088	.082	.074	.065	.051			