

A COMMENT ON A PAPER BY H. WU AND M. W. BROWNE (2014)

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA

I congratulate Hao Wu and Michael W. Browne (henceforth, WB) on their thought-provoking approach to specification error in moment structure analysis. To my reading, the novel and challenging issue of their paper is to interpret specification error as a (stochastic) second-level variation of the sample covariance matrix, variation that is said to be induced by (physical, real?) "adventitious error" that selects the actual population from where sample data are extracted. Instead of the adage "all models are wrong" ["...but some are useful"], WB imply that each model is wrong in the current population but is true in an hypothetical super-population. The RMSEA emerges as the natural descriptor of the "distance" between the actual and a hypothetical population. My comments elaborate this view and aim to widen the perspective of WB's paper by relating their approach to alternatives in the literature.

1. WB Versus Chen (1979)

At first reading, one sees a striking overlap between WB's approach and Chen (1979) (henceforth, Ch). Let  $s = \text{vec}(S)$  where  $S = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$ , and  $x_1, \dots, x_n$  are iid observations from a  $p$ -variate random  $x$ . Here,  $\text{vec}(\cdot)$  denotes the usual vectorization operator.

Both WB and Ch assume the following two-level variation set-up for  $s$ :

1. Level-one variation,  $s$  is a realization of a distribution with mean  $\sigma$  and covariance matrix  $\Gamma_\sigma$ , say

$$\text{L1: } s \sim (\sigma, \Gamma_\sigma);$$

2. Level-two variation,  $\sigma$  is a realization of a distribution with mean  $\omega$  and covariance matrix  $\Gamma_\omega$ , say

$$\text{L2: } \sigma \sim (\omega, \Gamma_\omega).$$

WB and Chen coincide with the distributions assigned to L1 and L2: a Wishart distribution with degrees of freedom ( $df$ )  $n$  to L1; an Inverted Wishart distribution with (unknown) scalar parameter  $m$  to L2;<sup>1</sup> these distributions imply that  $\Gamma_\sigma$  and  $\Gamma_\omega$  are functions of  $\sigma$  and  $n$ , and  $\omega$  and  $m$ , respectively. Both WB and Ch specify a moment structure on  $\omega$ , namely

$$M_\omega: \omega = \omega(\tau), \quad \tau \in \Upsilon \subset R^q,$$

Work supported by grant EC02011-28875 from the Spanish Ministry of Science and Innovation.  
Correspondence should be made to Albert Satorra, Universitat Pompeu Fabra, Barcelona, Spain.  
Email: albert.satorra@upf.edu

<sup>1</sup> Ch denotes as  $\nu$  the parameter that WB denote as  $m$ , and vice-versa. Moreover, WB use  $\nu$  to denote  $1/m$ .

where  $\omega(\tau)$  is a (continuously differentiable) function of  $\tau$ , a  $q$ -dimensional vector of parameters varying in  $\mathcal{Y}$  an open and compact subset of  $R^q$ ; finally, both WB and Ch estimate the parameter vector  $\tau$  and the scalar parameter  $m$  using the maximum likelihood (ML) method. WB and Ch, however, diverge in the computational approach used: Ch uses the EM algorithm, while WB use the Newton–Raphson method. They also diverge in that Ch uses the Bayesian approach, and WB use (frequentist) asymptotic methods.

In our view, however, the fundamental difference between WB and Ch is conceptual; it is the role assigned to the moment structure  $M_\omega$  and the parameter vector  $\tau$ . WB view  $\tau$  as the vector of structural parameters (loadings, regression coefficients, etc.) and  $M_\omega$  as a true model for the super-population mean vector  $\omega$ . In contrast, Ch views  $\tau$  as a vector of incidental parameter with no substantive interpretation and  $M_\omega$  as a prior information that serves to improve estimation of  $\theta$ , the vector of structural parameters (loadings, regression coefficients, etc.), a function (explicit or implicit) of the moment vector  $\sigma$ ,  $\theta = \theta(\sigma)$ . In an example where  $M_\omega$  is a factor model and a component of  $\theta$  is a regression coefficient ( $\beta = \sigma_{12}/\sigma_{22}$ ), Ch illustrates how the introduction of the variation level L2 and  $M_\omega$  improve the MSE of the classical estimator of  $\beta$ , especially when sample size  $n$  is small. For large samples, Ch's estimates converge to the classical ones that assume only L1 variation. In Ch's approach,  $M_\omega$  is a parsimonious parameterization of the super-population parameter  $\omega$ ; if  $M_\omega$  is saturated, no reduction in MSE is achieved. Ch's variation L2 is a statistical artifact to produce a regularization estimator of  $\theta$  for small samples.

In contrast, in WB's approach, L2 confers to  $M_\omega$  the status of a true model, not for the "operative population" (i.e., for  $\sigma$ ), but for the mean vector  $\omega$  of the postulated super-population. The vector  $\sigma$  is just a realization of L2; so,  $M_\omega$  is always misspecified for  $\sigma$ . WB assign even a (physical?) motor for the variation L2, what they call "adventitious error," and they even get specific on the likely sources for this error.

In our view, however, data do not provide empirical evidence of the existence of L2 (nothing in that vein is pointed in WB), so we will just contemplate L2 as a mathematical device for modeling. The assumption that the distribution of L2 is an Inverted Wishart distribution seems to be critical for the whole WB's set-up. This distribution assumption and the presumption that  $M_\omega$  is not saturated permit identification of a key scalar parameter:  $\nu = 1/m$  (if  $M_\omega$  is saturated then the WB's approach collapses). The parameter  $\nu$  takes the role of a scalar measure of the intensity of the variation L2, a kind of "distance" between  $\sigma$  and  $\omega$  (the mean vectors of the "operative" and hypothetical populations). The assumption of an Inverted Wishart distribution provides even foot for deriving an estimate of  $\nu$ , which turns out to be the square of the well-known RMSEA (Browne & Cudeck, 1992). Asymptotic derivations based on  $n \rightarrow \infty$  and  $\nu \rightarrow 0$  provide even a standard error for this estimate. Interestingly, the two-level variation set-up of WB offer a (theoretical) re-visiting of the RMSEA.

WB point to insufficiencies of the traditional approach to moment structure analysis.

## 2. The Traditional Approach, TA

WB claim that their approach overcomes the problems of what they call the "traditional approach" (TA). TA is simply L1, L2, and  $M_\omega$  with  $\Gamma_\omega = 0$ ; i.e., the case where L2 vanishes and thus  $\sigma \equiv \omega$  and  $M_\omega \equiv M_\sigma$  is a model for  $\sigma$ .<sup>2</sup> As in Ch, TA addresses estimation of the vector  $\theta$  of structural parameter, a function of  $\sigma$  (implicit in  $M_\sigma$ ). For large samples, TA coincides with Ch. TA and Ch diverge from WB in small and large samples, except when  $M_\omega$  is saturated, in which case the three approaches are the same.

<sup>2</sup> Note that in current TA, no distribution is specified in L1, so  $\Gamma_\sigma$  is not necessarily a function of  $\sigma$ , a non-parametric estimate of  $\Gamma_\sigma$  is in place.

WB questions the validity of the Pitman's drift used by TA when  $M_\omega$  is misspecified (recall that in TA  $\omega \equiv \sigma$ ); we read "[TA] though fairly successful, is still imperfect" (end of Section 1.2) and argues that "[the Pitman's drift device] is implausible in practice" (end of Section 2.3).

### 2.1. On the Pitman's Drift Device

Let  $\delta \equiv \min_\tau \|(\sigma(\tau) - \sigma)\|$  be a measure of the deviation of the model  $M_\sigma$  from  $\sigma$  (here  $\|\cdot\|$  denotes a norm). Let  $\delta_n$  be  $\delta$  when sample size is  $n$ . In case of misspecification of  $M_\sigma$ , i.e., when  $\delta \neq 0$ , TA derives asymptotic distributions for estimates, test statistics, and model diagnostics, using the mathematical device that  $\sqrt{n}\delta_n \rightarrow \mu$ , where  $\mu$  is finite (e.g., Satorra, 1989). It is known that this device produces asymptotic approximations to the actual distributions that will be accurate when  $n$  is large and  $\delta$  is small. In the case of estimates, this device ensures the estimates are consistent for the parameters of a limit model where  $\delta = 0$ . For test statistics, this device produces non-central distributions for approximating the actual distributions. It is not needed that "physically" the populations move with  $n$ , as it seems implicit in WB's comment "[this Pitman drift] is implausible in practice because the population should not be affected by sample size" (end of Section 2.3), what is needed is that the model is not too deviant from a model where  $\delta = 0$ , i.e., that the posited model approximates the true model. The Pitman's drift should be viewed just as a mathematical device to obtain an asymptotic approximation of the distribution of statistics of interest (the accuracy of the approximation being judged usually by Monte Carlo evaluation).

The Pitman's drift (also called sequence of local alternatives) is classical in statistics (e.g., Wald, 1943) and has been used extensively. Satorra and Saris (1985) used this device to justify a procedure to approximate the power of the chi-square goodness-of-fit test (see, e.g., Satorra & Saris, 1983; Satorra, Saris, & de Pijper, 1991, for Monte Carlo evidence on the accuracy of this procedure). More recently, Chun and Shapiro (2009) showed that, for small deviations from the null, the Pitman's drift device provides more accurate approximations to the actual distribution of the chi-square goodness-of-fit test than fix alternatives.

We note that the asymptotic results of WB are obtained under the double limit  $n \rightarrow \infty$  and  $\nu \rightarrow 0$ , we read "This assumption [ $\nu \rightarrow 0$ ] is more plausible than the traditional Pitman drift assumption in that adventitious error is only assumed small, but not assumed to get smaller when the sample size increases" (mid of Section 10). Clearly,  $\nu$  is tied to the size of misspecification and is thus the counterpart of  $\delta$ , so  $\nu \rightarrow 0$  is tantamount to  $\delta_n \rightarrow 0$ , and, to our view, the WB's asymptotics stay on identical foot as the Pitman's drift with regards to being "plausible in practice." Whether the limits of  $n$  and  $\delta_n$  operate in one dimension, the Pitman's drift device of  $\sqrt{n}\delta_n \rightarrow \mu$ , or in two dimensions,  $n \rightarrow \infty$  and  $\nu \rightarrow 0$  independently, do not change their plausibility; except that, for undertaking a Monte Carlo evaluation of the accuracy of these asymptotic approximations, the Pitman's drift offers a much more simplified frame than the double limit involved in WB.

## 3. TA and WB's Two-Level Data

Assume that data arise from the two-level set-up of WB, under an exact model  $M_\omega$  for  $\omega$ . What would be the consequences of applying standard TA analysis to data  $x_1, \dots, x_n$  sampled from  $\sigma$ ? This section investigates this issue using a Monte Carlo illustration.

We consider the following data generating process. For a given choice of  $\omega$  and  $\nu = 1/m$ , a population vector  $\sigma$  is sampled from L2 and, given  $\sigma$ , a sample  $x_1, \dots, x_n$  of size  $n$  is extracted from L1, with the Wishart and Inverted Wishart distributions used in L1 and L2, respectively. The sample covariance matrix  $S$  is computed from  $x_1, \dots, x_n$  and fitted—using TA—to the

following single-factor model  $M_\sigma : \sigma = \sigma(\theta)$ , where  $\sigma = \text{vec } \Sigma$ ,

$$\Sigma(\theta) = \begin{pmatrix} \lambda_1^2 + \psi_1 & & & & \\ \lambda_1\lambda_2 & \lambda_2^2 + \psi_2 & & & \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2 + \psi_3 & & \\ \lambda_1\lambda_4 & \lambda_2\lambda_4 & \lambda_3\lambda_4 & \lambda_4^2 + \psi_4 & \\ \lambda_1\lambda_5 & \lambda_2\lambda_5 & \lambda_3\lambda_5 & \lambda_4\lambda_5 & \lambda_5^2 + \psi_5 \end{pmatrix}$$

and  $\theta = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \psi_1, \psi_2, \psi_3, \psi_4, \psi_5)$ . For the simulations, we use  $\omega = \text{vec } \Omega$  with

$$\Omega = \begin{pmatrix} 1 & .64 & .64 & .64 & .64 \\ .64 & 1 & .64 & .64 & .64 \\ .64 & .64 & 1 & .64 & .64 \\ .64 & .64 & .64 & 1 & .64 \\ .64 & .64 & .64 & .64 & 1 \end{pmatrix},$$

the covariance matrix implied by a single-factor model with loading parameters equal to .8, unique variances equal to .36, and variance of the common factor equal to 1. Thus  $M_\sigma$  is a true model for  $\omega$ , but it is a misspecified model for  $\sigma$ . The degrees of freedom of the chi-square goodness-of-fit model test is  $df=5$ .

Given the above  $\Omega$ , for each combination of  $m$  and  $n$  listed in the first two columns of Table 1, 1000 replicates of  $S$  and the corresponding TA analyses are undertaken. TA is carried out using `lavaan` (Rosseel, 2012) with specification ML, and all the computations are performed in R (R Development Core Team, 2008). The Monte Carlo distribution of the TA parameter estimates and standard errors (se's) as well as the distribution of a scaled version of the regular chi-square goodness-of-fit test is reported in Table 1. For simplicity, we only consider the estimate of  $\lambda_1$ , the other estimates perform in a similar manner.

We can reason that the unconditional expectation of  $E_u(S)$  (i.e., the expectation under variations L1 and L2) is

$$E_u(S) = E_{L2}(E_{L1}(S | L1)) = E_{L2}(\Sigma) = \Omega,$$

where  $E_L$  denotes expectation conditional to level  $L$ ; that is, the unconditional expected matrix value of  $S$  is the matrix  $\Omega$  for which  $M_\sigma$  holds exactly. Consistency of  $S$  as an estimate of  $\Omega$  could also be achieved but at the cost, however, of asymptotic limits with  $n \rightarrow \infty$  and  $\nu \rightarrow 0$ . Thus, we should expect the TA estimate of  $\lambda_1$  to be centered around its true value 0.8 despite the fact that the model  $M_\sigma$  analyzed by TA is misspecified for  $\sigma$ . The standard errors (se's) produced by TA, however, are necessarily conditional to  $\sigma$ , the population from where  $x_1, \dots, x_n$  has been extracted. That is, the se's produced by TA do not take into account the variation L2, the sampling variation of  $\sigma$  within L2. Unconditional se's for TA estimates can be deduced from the unconditional distribution of  $S$ . Clearly, given the nature of the distribution assumed in L2, the unconditional asymptotic variance of  $\text{vec } S$  is obtained by multiplying the conditional one (the one that assumes only variation L1) by the following scaling factor

$$c = (1 + n/m),$$

a number that necessarily is  $c > 1$ . For this result, we require  $n \rightarrow \infty$  and  $\nu \rightarrow 0$ . That is, multiplication by  $\sqrt{c}$  transforms the (conditional to L1) se's produced by TA to the unconditional

TABLE 1.  
Monte Carlo results for TA analysis of WB's two-level data.

$m = 1/\nu$	$n$	$m(\hat{\lambda}_1)^a$	$sd(\hat{\lambda}_1)^b$	$\overline{se}_{\hat{\lambda}_1}^c$	$\sqrt{c} \times \overline{se}_{\hat{\lambda}_1}^d$	$c^e$	5 % $T/c^f$
60	200	0.8340	0.1363	0.0632	0.1315	4.3333	5.50 <sup>g</sup>
600	200	0.7993	0.0700	0.0607	0.0701	1.3333	4.70
6000	200	0.8010	0.0630	0.0606	0.0616	1.0333	5.90
60	2000	0.8437	0.1273	0.0201	0.1175	34.3333	5.70

<sup>a</sup> mean of the 1000 replications of estimates of  $\lambda_1$ .

<sup>b</sup> sd of the 1000 replications of estimates of  $\lambda_1$ .

<sup>c</sup> mean of the 1000 replications of ML standard error of  $\lambda_1$ .

<sup>d</sup> mean of the 1000 replications of the scaled ML standard error of  $\lambda_1$ .

<sup>e</sup> the scaling factor  $c = 1 + n/m$ .

<sup>f</sup> rejection rate for the scaled ML chi-square 5 %-level test.

<sup>g</sup> the 95 % margin of error bound is  $\pm 1.4$ .

ones that incorporate the variation of L2. Moreover, scaling by  $c$  the regular TA chi-square goodness-of-fit test (as in Satorra & Bentler, 1994) produces an asymptotic chi-square statistics that takes into account both L1 and L2 variations.<sup>3</sup> These expectations will now be confronted with the Monte Carlo results shown in Table 1.

We observe that: (i) the Monte Carlo mean of the TA estimate  $\hat{\lambda}_1$  is close to the true population value .8 (see column 3); (ii) the mean of the se's computed by TA (column 5) show a downward bias with respect to the true unconditional se's estimated by Monte Carlo (column 4); (iii) the scaled se's (column 6) are fairly close to the Monte Carlo se's (column 4); and (iv) the Monte Carlo estimate of the rejection rate of the (scaled) chi-square 5 %-level test agrees with its nominal value 5 % (the last column of the table). Thus, Table 1 confirms our expectation. Noteworthy is the large value of  $c$  in the case of large misspecification ( $m = 60$ ) and large sample size  $n = 2000$ .

#### 4. Conclusion

WB write “[TA] though fairly successful, is still imperfect” and propose replacing the single-level approximative model L1 of TA by a two-level variation L1 and L2 where the model  $M_\sigma$  is exact, not in the “operative population” but in the (hypothetical) super-population L2. The two-level variation L1 and L2 parallels the formulation of Chen (1979), but with a fundamental difference: while for WB the parameters of interest to the researcher are the parameters of a true super-population model (the model in L2), for Ch, L2 is just a statistical device aimed to improve estimation of  $\theta$ , a vector of structural parameters residing in L1. In Ch, when  $n \rightarrow \infty$ , estimation converges to TA. This contrasts with WB, where the parameter of interest resides in L2, and even when sample size  $n \rightarrow \infty$  (i.e., when first level variation L1 vanishes), the crux of estimation rests on L2 (this is precisely the case of the very large value of  $c$  in Table 1).

The WB's two-level set-up has the virtue of encapsulating the whole model misspecification issue into the single scalar  $\nu$ , or its estimate, the square of the RMSEA. No other model evaluation statistic, however, emerges from the WB's approach. Not even a test for what presumably is a very restrictive assumption, the Inverted Wishart distribution assumed in L2, a distribution assumed to mimic adventitious error variability.

<sup>3</sup> WB show that RMSEA is an estimate of  $\sqrt{\nu}$  (when  $n \rightarrow \infty$  and  $\nu \rightarrow 0$ ), so the scaling factor for the standard errors could be computed using the RMSEA (when  $n$  is large and  $\nu$  is small). This, however, would not work for scaling the chi-square test.

The Monte Carlo illustration has shown that when L2 is present, then TA estimates are valid (consistent, when  $\nu$  is small) estimates for the parameters of the true model  $M_\omega$  that resides in L2. The se's produced by TA, however, are conditional to the "operational population," they account only for L1 variation. A simple scaling factor, however, can be used to expand TA se's to account for the variation added by L2. For given  $m$  and  $n$ , a simple scaling of the regular chi-square goodness-of-fit test produced by TA is an (unconditional distribution) asymptotic chi-square statistic.

WB's paper opens a new interpretation for RMSEA, a measure of the magnitude of "adventitious error." Model evaluation, however, cannot be reduced to a single number. TA offers several statistics to assess misspecification (model test statistic, modification indices, parameter change indicators, etc.). Can these diagnostic statistics be extended to WB's two-level set-up when the covariance structure is postulated in L2? How can we assess empirically the distribution assumed for L2 as well as the whole two-level model set-up? Given the mathematical similarity of Ch and WB, could a connection be made to emerge with more tools for model diagnostic than just the RMSEA? An interesting feature of Ch's approach is the reduction of mean square error of estimates when compared to TA, for small samples.

## References

- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230–258.
- Chen, C.-F. (1979). Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *Journal of the Royal Statistical Society, Series B*, 41, 235–248.
- Chun, S. Y., & Shapiro, A. (2009). Normal versus noncentral chi-square asymptotic of misspecified models. *Multivariate Behavioral Research*, 44, 803–827.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rosseeel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54(1), 131–151.
- Satorra, A., & Bentler, P. M. (1994). Correctio to test statistics and standard errors in covariance structure analysis. In A. van Eye & C. C. Clogg (Eds.), *Latent variable analysis in developmental research* (pp. 285–305). Thousand Oaks, CA: Sage Publications.
- Satorra, A., & Saris, W. E. (1983). The accuracy of a procedure for calculating the power of the likelihood ratio test as used within the LISREL framework. In C. P. Middendorp, B. Niemoller, & W. E. Saris (Eds.), *Sociometric Research 1982* (pp. 127–190). Amsterdam: Sociometric Research Foundation.
- Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50(1), 83–89.
- Satorra, A., Saris, W. E., & de Pijper, W. M. (1991). A comparison of several approximations to the power function of the likelihood ratio test in covariance structure analysis. *Statistica Neerlandica*, 45, 173–185.
- Wald, A. (1943). Test of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426–482.

*Manuscript Received: 13 OCT 2014*