

# Seeing Perfectly Fitting Factor Models That Are Causally Misspecified: Understanding That Close-Fitting Models Can Be Worse

Educational and Psychological  
Measurement

2014, Vol. 74(6) 905–926

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164414527449

epm.sagepub.com



Leslie Hayduk<sup>1</sup>

## Abstract

Researchers using factor analysis tend to dismiss the significant ill fit of factor models by presuming that if their factor model is close-to-fitting, it is probably close to being properly causally specified. Close fit may indeed result from a model being close to properly causally specified, but close-fitting factor models can also be seriously causally misspecified. This article illustrates a variety of nonfactor causal worlds that are perfectly, but inappropriately, fit by factor models. Seeing nonfactor worlds that are perfectly yet erroneously fit via factor models should help researchers understand that close-to-fitting factor models may seriously misrepresent the world's causal structure. Statistical cautions regarding the factor model's proclivity to fit when it ought not to fit have been insufficiently publicized and are rarely heeded. A research commitment to understanding the world's causal structure, combined with clear examples of factor mismodeling should spur diagnostic assessment of significant factor model failures—including reassessment of published failing factor models.

## Keywords

factor analysis, factor model, testing, close fit, structural equation modeling

Factor analysis and path analysis existed for decades as separate statistical approaches before it became commonly recognized that the measurement and causal structures of

---

<sup>1</sup>University of Alberta, Edmonton, Alberta, Canada

## Corresponding Author:

Leslie Hayduk, Department of Sociology, University of Alberta, Tory 5-21, Edmonton, Alberta, Canada T6G 2H4.

Email: LHayduk@ualberta.ca

factor and path models could be combined under the rubric of structural equation modeling (Joreskog, 1970; Spearman, 1927; Wright, 1921). The cohabitation of path+factor within structural equation modeling (SEM) proceeded statistically smoothly, but in recent decades fundamental conflicts have arisen over differences in operating procedures, rules of thumb, and intellectual commitments. For example, Anderson and Gerbing's (1988, 1992) recommendation to routinely use factor analysis prior to employing full structural equation models was challenged by Fornell and Yi (1992a, 1992b), and additional criticisms of the factor-before-path idea were provided by Hayduk (1996) and Hayduk and Glaser (2000a, 2000b). More recently, Hayduk and Littvay (2012) considered the appropriate number of indicators required for measuring latent variables in structural equation models and concluded that it is preferable to use the few best indicators (one, or two, rarely three)—which is fewer than what some factor analytic researchers find comfortable.

One fundamental difference between the factor and path traditions concerns model testing. Barrett's (2007) call for careful attention to all significant  $\chi^2$  model ill fit garnered very different responses from those with factor analytic backgrounds (Millsap, 2007; Mulaik, 2007; Steiger, 2007) than from those who were more SEM oriented (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). There was general agreement that all models should be tested and that significant ill fit between a model's implications and the data must be reported as evidence speaking against the model. But, considerable divergence remained regarding the seriousness and dedication with which researchers should attend to, and investigate, the reasons behind model ill fit. Those with factor histories tended to defend indexing of model ill fit, whereas the SEM oriented called for careful diagnostic investigation of all beyond-chance evidence of model ill fit because this potentially signals important model misspecifications (Hayduk et al., 2007; Millsap, 2007; Mulaik, 2007; Steiger, 2007).

There is a natural hesitancy to report one's model as failing, and there is a general perception that failing models are less likely to be published—despite calls for publishing informative failing models (Hayduk et al., 2007). These concerns apply to all models and hence cannot explain factor analysts' exceptional hesitancy to attend to model ill fit. A more likely explanation is that disregard for testing originated in factor analytic procedures and rules of thumb that became entrenched before the  $\chi^2$  model test appeared. The comparatively recent arrival of model testing pitted the relative newcomer—the  $\chi^2$  model test—against entrenched factor assessment traditions. Basic factor analytic training, factor interpretations, and even the intent to “simplify” rather than mirror the world, all contribute to disregarding significant factor model ill fit.

Factor models are rarely presented or interpreted as causal models. Statisticians were hesitant to call factor models causal models, in contrast to the clear causal commitments of those doing path-oriented structural equation modeling (Bollen, 1989; Duncan, 1975; Hayduk, 1987; Hayduk & Pazderka-Robinson, 2007; Heise, 1975; Pearl, 2000; Wright, 1921). Some factor analysts have recently acknowledged that factor models are indeed causal models (e.g., Mulaik, 2010; Wright & Villalba,

2012), but the impact of this is not yet widely appreciated. Once factor models are viewed as constituting claims about the world's causal structure, consistency with worldly based data becomes mandatory. It is no longer adequate to identify latent factors merely by pointing to the meanings of high-loading indicators because that dubiously conflates "similarity in meaning" with "dependence on a common cause." Similar indicator meanings do not certify dependence of those indicators on a common cause.

Seeking simplicity by having the data "suggest" only a few underlying latent factors also obscures the causal nature of factors. Rules-of-thumb requiring retention of a factor for each eigenvalue exceeding 1.0, or corresponding to the break-point in a scree plot, or "accounting" for a substantial proportion of "common variance," do not encourage, let alone force, the researcher to attend to the causal nature of retained latent factors. Exploratorily seeking a few latent factors directly acknowledges that the researcher has no commitment to a specific number of underlying latent causal factors, and consequently it becomes nearly impossible for the researcher to convincingly "retrospectively theorize" about causal structures linking the retained latent factors to the indicators or to other latents (Hayduk & Glaser, 2000a, 2000b). Disturbingly many factor analysts refer to "loadings," "common variance," and "average variance explained" in ways that avoid acknowledging how factors as common causes produce the variance in indicators and the correlations between indicators.

Even the intent of minimizing the number of latent factors conspires against causal understanding. The world has some degree of causal complexity, but routinely proceeding as if the world's structure is likely to arise from a minimum number of common causes seems presumptive in the extreme. We as researchers cannot simplify the world—the best we can do is match the world's causal complexity, however simple or complex the world happens to be.

This article urges researchers to respect and conscientiously diagnostically investigate all significant factor model ill fit because even small amounts of real ill fit may signal important causal misspecifications. We begin with multiple illustrations of how the factor model can fit and satisfy standard factor analytic criteria, despite the factor model being importantly causally misspecified. Once it is realized that seriously wrong factor models can fit, it becomes obvious that slight but statistically significant factor model ill fit may be the first detectable evidence of important factor model misspecification.

The following sections of this article present hypothetical worldly models that would provide covariance data satisfying the criteria factor analysts commonly use to confirm or support a factor model. Specifically, we retain all the factors having an eigenvalue greater than 1.0, which is often called the Kaiser–Guttman rule (Kaiser, 1960; Mulaik, 2010), and all the factors whose scree plot shows a noticeable disjunction from the scree tail (Cattell, 1966; Ruscio & Roche, 2012). We attend to the variance-explained (actually variance and covariance explained) by retaining as many factors as are required to render the residual variances and covariances

insignificant—so the model fits according to the  $\chi^2$  test. Nearly all the following models satisfy these criteria with only a single factor, but only some of the true worldly models are factor models, and none of the true models are one-factor models. These examples illustrate worldly causal structures that therefore risk being mistakenly described as one-factor models. We expect that readers will be disturbed by the traditional factor criteria and even the stringent  $\chi^2$  test criteria, pointing toward “accepting” wrong factor models. We conclude by considering the wider disciplinary implications of misspecified factor models and what can be done to mitigate the problems.

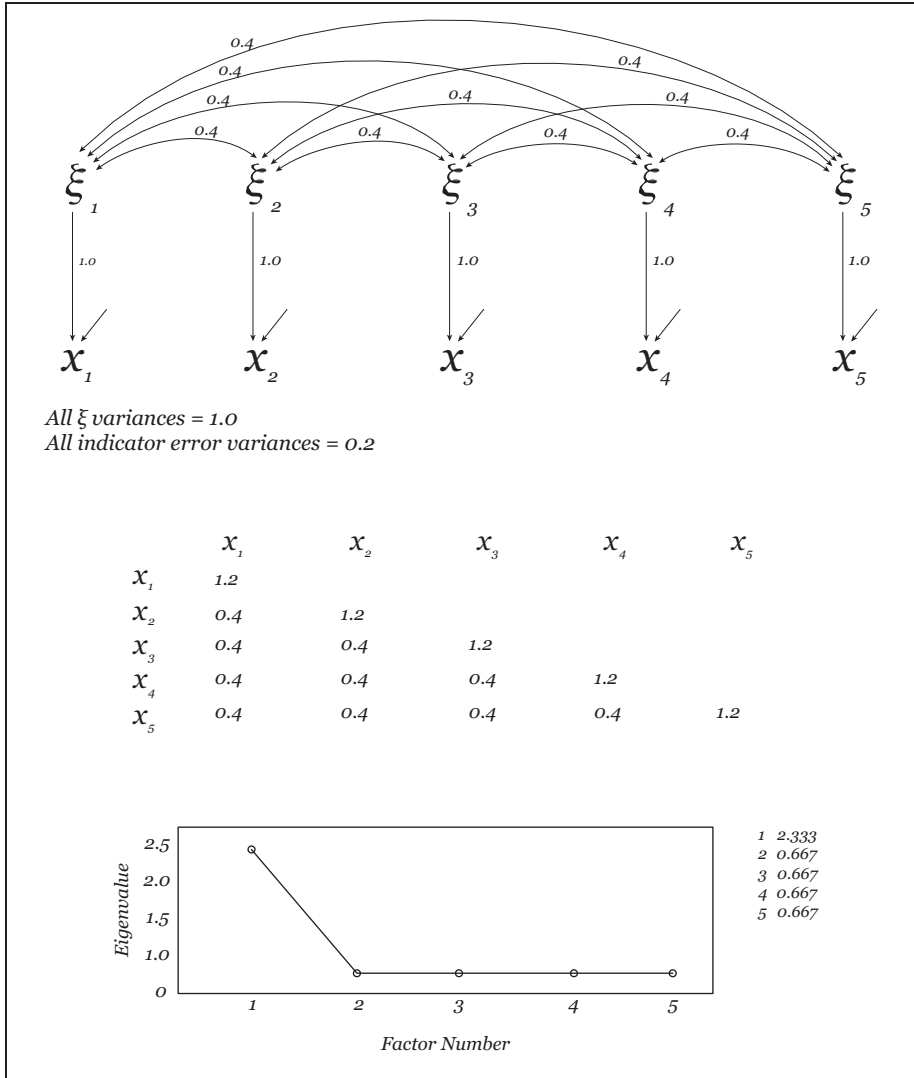
## Illustrative Models

We present each “worldly” model as a figure containing parameter values entered as fixed coefficients in LISREL 8.8 (Joreskog & Sorbom, 1996) to compute the corresponding model-implied covariance matrix. This matrix constitutes the real-world or population covariance matrix that would appear for the indicators if the depicted model constituted the worldly causal forces. Each figure also provides the eigenvalues and scree plot for the corresponding correlation matrix (from PASW Statistics 18; the name applied to SPSS during an ownership transition). The eigenvalues and scree from the correlation matrix are reported because the “most common objective criteria that has been used to decide on the number of factors is Kaiser’s ‘eigenvalues greater than 1.0’ rule” (Chaplin, 2005, p. 632), and “many commercial factor analysis programs have had the effect of popularizing this rule by offering it as a default” (Mulaik, 2010, p. 186). While it is possible to use eigenvalues from communality-adjusted matrices (Mulaik, 2010), or other nontraditional assessment criteria, we know of no alternative criterion that would resolve the factor model difficulties reported below. For example, “parallel analysis” (Horn, 1965) would not help because our use of population covariance matrices eliminates sampling variability.

The text accompanying each model/figure reports the fit of a one-factor model using the covariance matrix and maximum likelihood estimation from LISREL 8.8. This provides protection against the types of errors potentially arising from applying covariance models to correlation matrices (Cudeck, 1989).  $N$  is specified as 200 so the  $\chi^2$  test has neither excessive nor insufficient power. The reported  $\chi^2$  values should not be given the usual interpretation but instead should be treated as  $\chi^2$  noncentrality parameters—where zero  $\chi^2$  reports that the model would be rejected at only the selected Type 1 error rate and where the program’s reported  $p$  provides a sense of how easily the factor model could be confused with the true model given even perfect data.

### Model 1

The model in Figure 1 illustrates that *a one-factor model can fit any covariance matrix containing uniform positive indicator correlations, no matter how those*



**Figure 1.** Uniform indicator correlations.

*uniform correlations were in fact produced by the world.* All the indicator correlations for this model equal 0.333 (namely  $0.333 = 0.4/\sqrt{(1.2 \times 1.2)}$ ).

There is a sharp break in the scree plot after one factor, and only a single eigenvalue exceeds 1.0, so these common rules of thumb are satisfied by a one-factor model—even though here the one-factor model is known to be seriously causally misspecified. A one-factor model with uniform positive loadings implies uniform

positive correlations among the indicators, and the one-factor model estimated for the Figure 1 covariance matrix results in perfect fit ( $\chi^2 = 0.00$ ,  $df = 5$ ,  $p = 1.0$ ) with all the standardized loadings 0.577.

This illustrates that a nonfactor causal world that implies uniform positive indicator correlations can be inappropriately fit by a *one-factor* model. Factor model fit via the eigenvalue, scree, and  $\chi^2$  test criteria does *not* assuredly report that one latent factor constitutes the worldly causal source of the indicators.

The latent level of the Figure 1 model was intentionally left causally-vague to emphasize that the inappropriate-fitting of the one-factor model is not dependent on a specific kind of latent causal structure fortuitously matching the factor model's causal structure. In this instance there were five latents, one for each indicator. The next model emphasizes that *any* worldly model providing uniform positive indicator covariances can be fit by a *one-factor* model.

### Model 2

Figure 2 illustrates a uniform indicator covariance matrix resulting from the causal actions of three independent latent factors. Each factor has a different but consistent causal effect leading to all the indicators. The resultant uniform indicator covariance matrix is again perfectly fit by a one-factor model ( $\chi^2 = 0.00$ ,  $df = 35$ ,  $p = 1.0$ ), this time with all standardized loadings 0.913, and again the eigenvalues and scree plot clearly satisfy the usual rules of thumb. Thus even a world consisting of multiple independent factors may be inappropriately fit by a one-factor model!

The scree test's erroneous pointing to a single factor is not an artifact of using the eigenvalues of the indicator correlation matrix, rather than the eigenvalues of a reduced matrix having communalities on the diagonal (Mulaik, 2010). The eigenvalues for the reduced matrix for the Figure 2 model are 8.33, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00 (using the known true variances from Figure 2)—which clearly but equally erroneously points to a one-factor model. For the Figure 1 model the reduced matrix eigenvalues are 1.67 followed by a tail of zeroes, and for the Figure 3 model the reduced matrix eigenvalues are 8.99 followed by a tail of zeroes, so the problem is deeper than merely which eigenvalues to use in a scree-like assessment.

### Model 3

Figure 3 provides a uniform indicator covariance matrix implied by a world consisting of three correlated latent factors. Again, a one-factor model, this time with all standardized loadings 0.948, fits perfectly ( $\chi^2 = 0.00$ ,  $df = 35$ ,  $p = 1.0$ ), and the scree plot and eigenvalues seem convincing. This indicates that the perfect fit of Model 2 was not an artifact of the independence of the latent factors. Together, Models 2 and 3 illustrate that it is sometimes possible for a one-factor model to perfectly, but inappropriately, match covariance data originating in a world composed of several latent

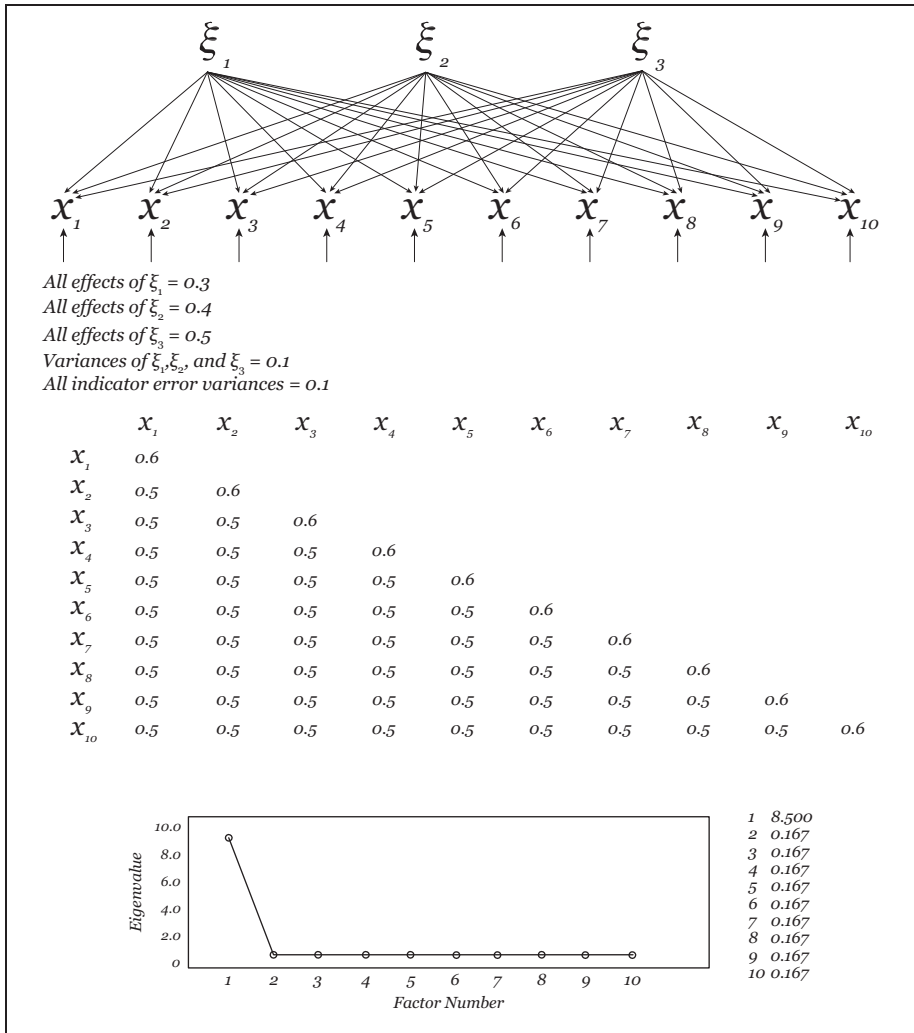


Figure 2. Three factors fit by a one-factor model.

factors, whether those factors are independent or not. That is, Models 2 and 3 illustrate that it is possible for a one-factor model to misrepresent, or mismodel, even other factor models. There are multifactor worlds (with different loading specifications) that could not be fit by a one-factor model, so these examples do not illustrate that the one-factor model will always be wrong—they merely illustrate that satisfying the usual rules of thumb for a one-factor model should not be treated as compelling evidence that a single underlying latent factor actually constitutes the underlying causal world.

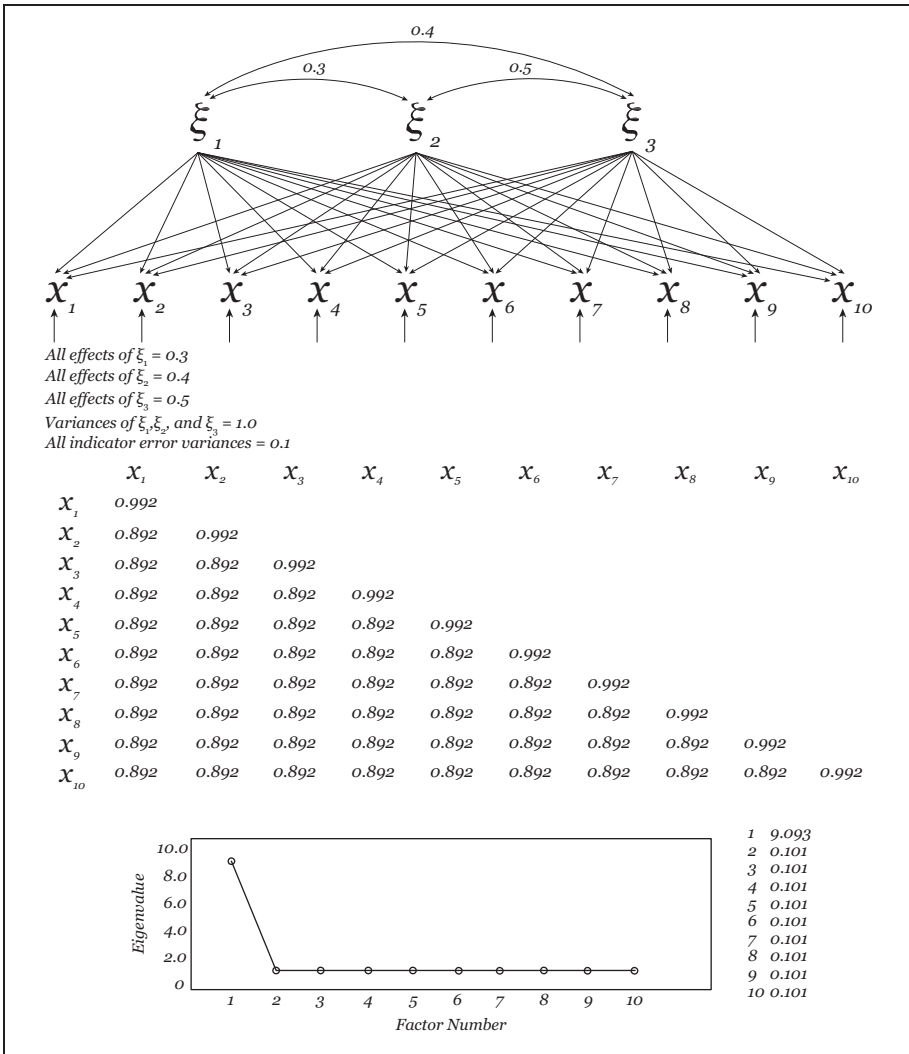


Figure 3. Three correlated factors fit by a one-factor model.

### Model 4

The model in Figure 4 moves away from a factor-structured world toward a regression-structured world. Here multiple uniformly coordinated and uniformly effective latents cause a dependent latent, with each latent having a single indicator. This regression-like latent causal world results in a nearly uniform covariance matrix, and the one-factor model again fits perfectly ( $\chi^2 = 0.00, df = 9, p = 1.0$ ) with



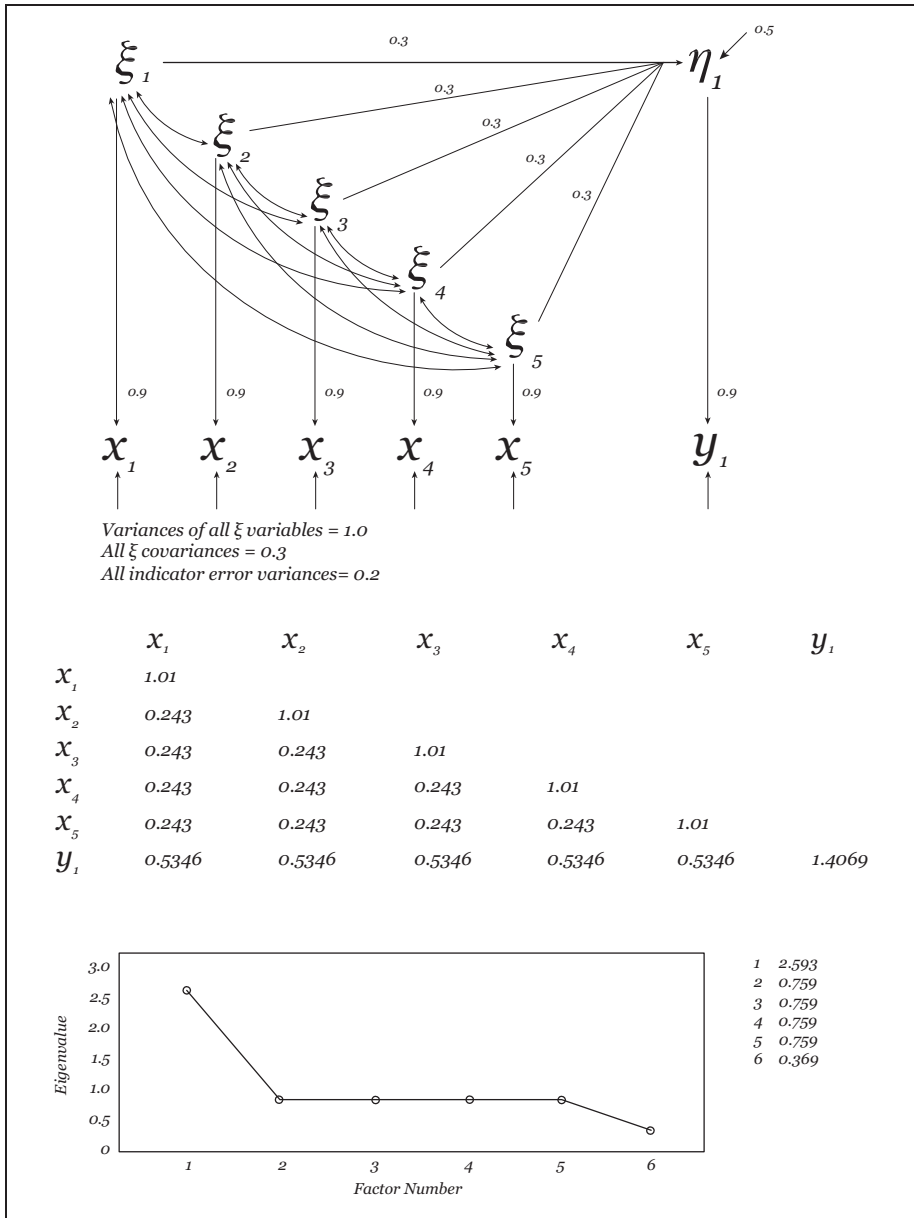


Figure 4. A regression model as almost one-factor.

a 0.491 standardized loading for each  $x$  and a 0.914 loading for  $y_1$ . The scree plot and eigenvalues again satisfy the usual rules of thumb. This illustrates a specific non-factor latent causal world that would be inappropriately fit by a one-factor model.

### Model 5

The model in Figure 5 is a bit of a mongrel. It contains a first-order latent “factor” ( $\eta_6$ ), a latent almost like a second-order factor ( $\eta_5$ ) and other model segments that are reminiscent of full structural equation modeling (creating direct, indirect, and reciprocal effects among the latents). With the indicated parameter values this worldly model provides a reasonably, but not perfectly, uniform covariance matrix, which has one large eigenvalue, a scree plot “suggesting” a single-factor, and which can be closely though not perfectly fit by a one-factor model ( $\chi^2 = 6.853$ ,  $df = 5$ ,  $p = .232$ ) with standardized loadings 0.362, 0.913, 0.873, 0.913, and 0.362. Here, a seriously misspecified one-factor model provides fit that would often be close-enough even for researchers attending to significant  $\chi^2$  ill fit.

Model 5 illustrates a feature that some factor analysts might prefer to forget—namely that the world may contain more latent variables than indicators. “*There is no guarantee . . . that any proposed battery of  $n$  measurements will involve less than  $n$  common factors*” (Thurstone, 1947, emphasis in the original). There is no hope of an estimable factor model matching a world having more latents than indicators, and indeed factor models in general are unable to estimate even as many factors as indicators (Bekker & ten Berge, 1997; Lederman, 1937). Factor practitioners typically prefer far fewer latents than indicators, so the possibility of *more* latents than indicators may be a factor analyst’s nightmare. Science is festooned with instances of the world turning out to be more complicated than initially imagined, and we see nothing that precludes personality or other styles of indicators from having more than a preferred “small” number of underlying latent causes. This possibility deserves serious consideration.

### Model 6

With the indicated parameter values, the Figure 6 worldly causal model provides a relatively uniform covariance matrix, with eigenvalues and scree plot that satisfy the traditional rules-of-thumb for one factor. A one-factor model having all standardized loadings 0.971 fits this world’s covariance matrix marginally ( $\chi^2 = 22.094$ ,  $df = 14$ ,  $p = .077$ ). Hence, although significant ill fit is not assured, additional sampling variations would be more likely to lead to significant  $\chi^2$  values—and tempt some researchers to abandon  $\chi^2$  in favor of close-fit indices for “seeming-agreement” with the eigenvalues and scree.

The model in Figure 6 is rather unusual—it has no exogenous latents, and it contains multiple causal loops. To some factor analysts this model might even seem unimaginable. This model instructs us that competent structural equation modeling may require imagining, and seeking, worldly causal structures radically different than anything previously considered. Again, the fact that a one-factor can satisfy the traditional criteria should not lull a researcher into thinking that the world’s causal structure must therefore be like, or almost like, a factor model.

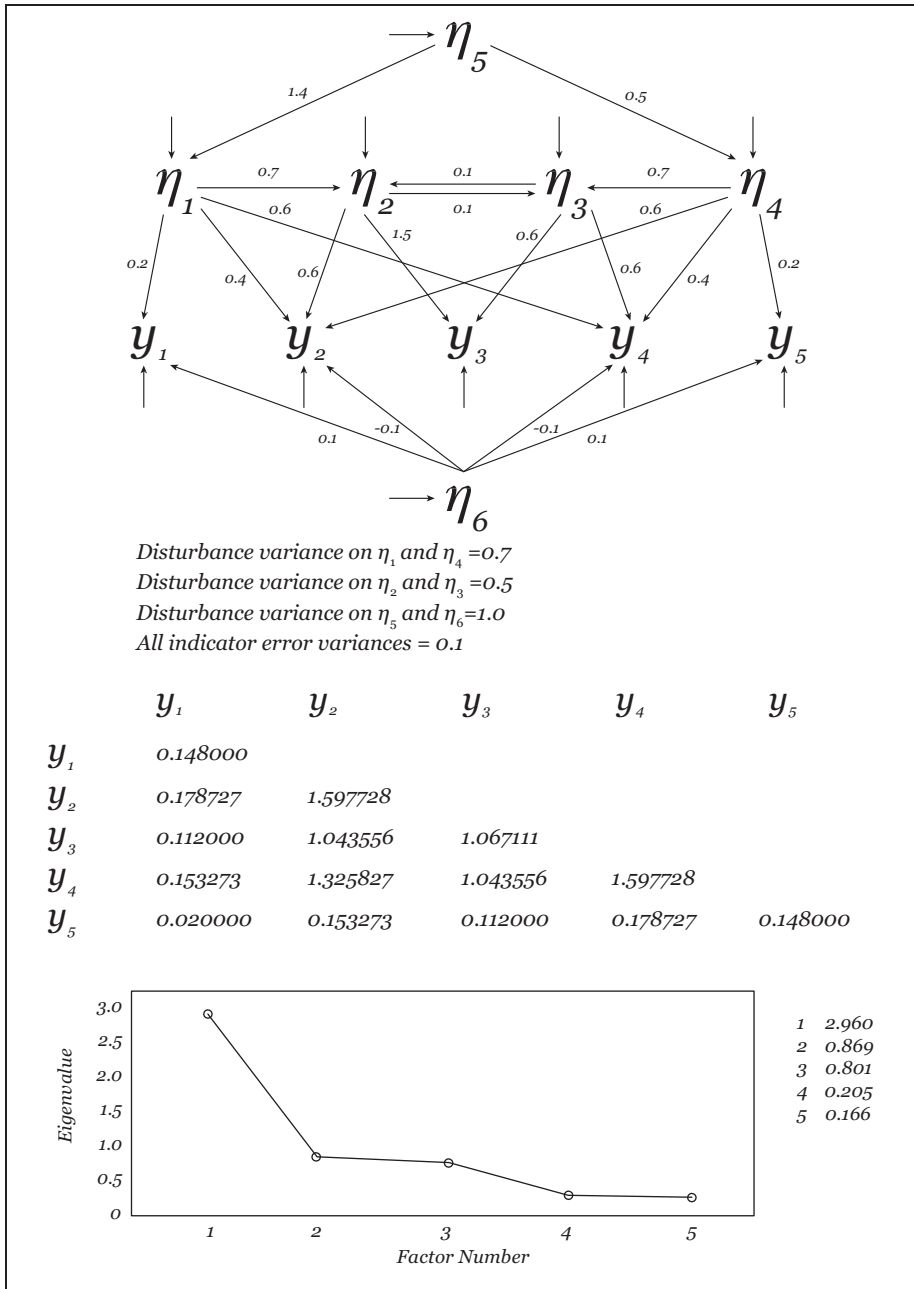


Figure 5. A complex world for one-factor.

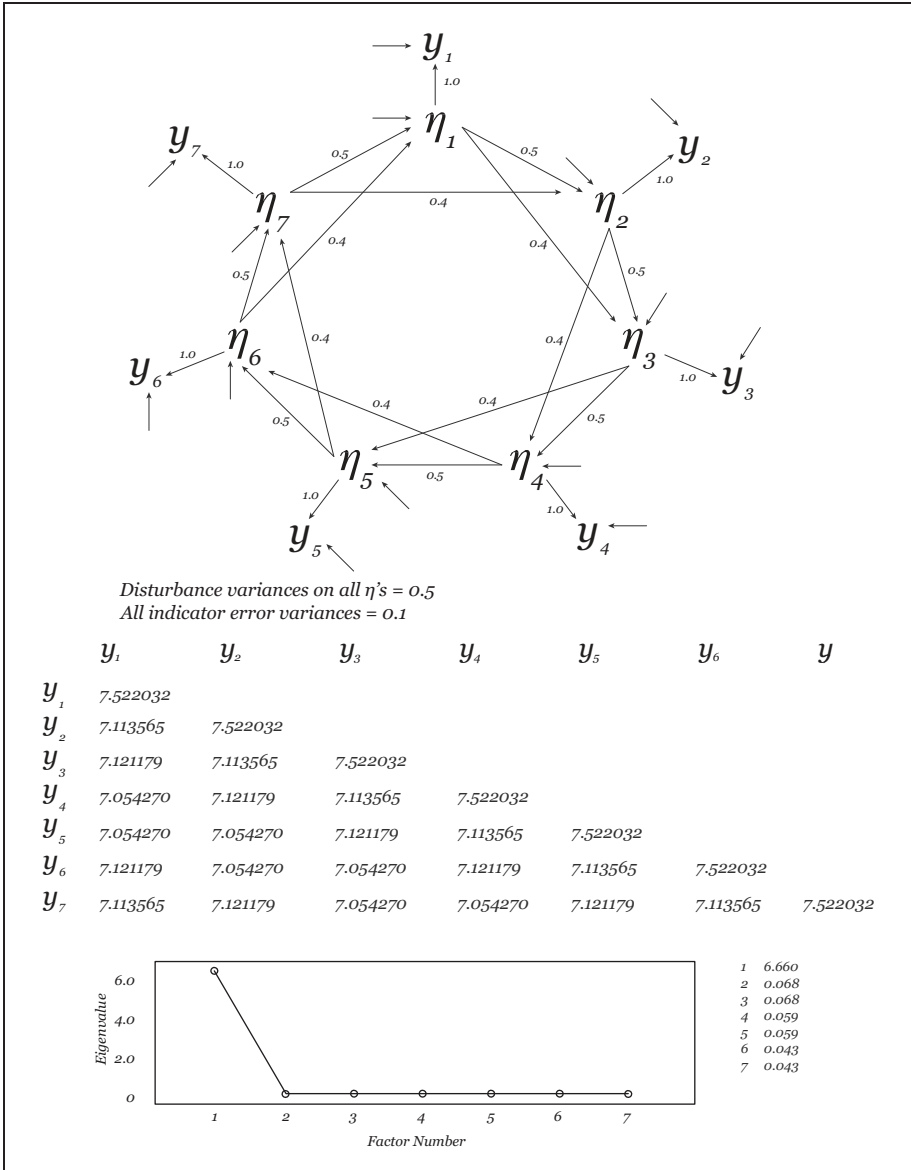


Figure 6. A pretty world as one-factor.

### Model 7

The model in Figure 7 is derived from the motif model of Hayduk (1996) and contains a variety of features that are neither particularly common nor alien. The model

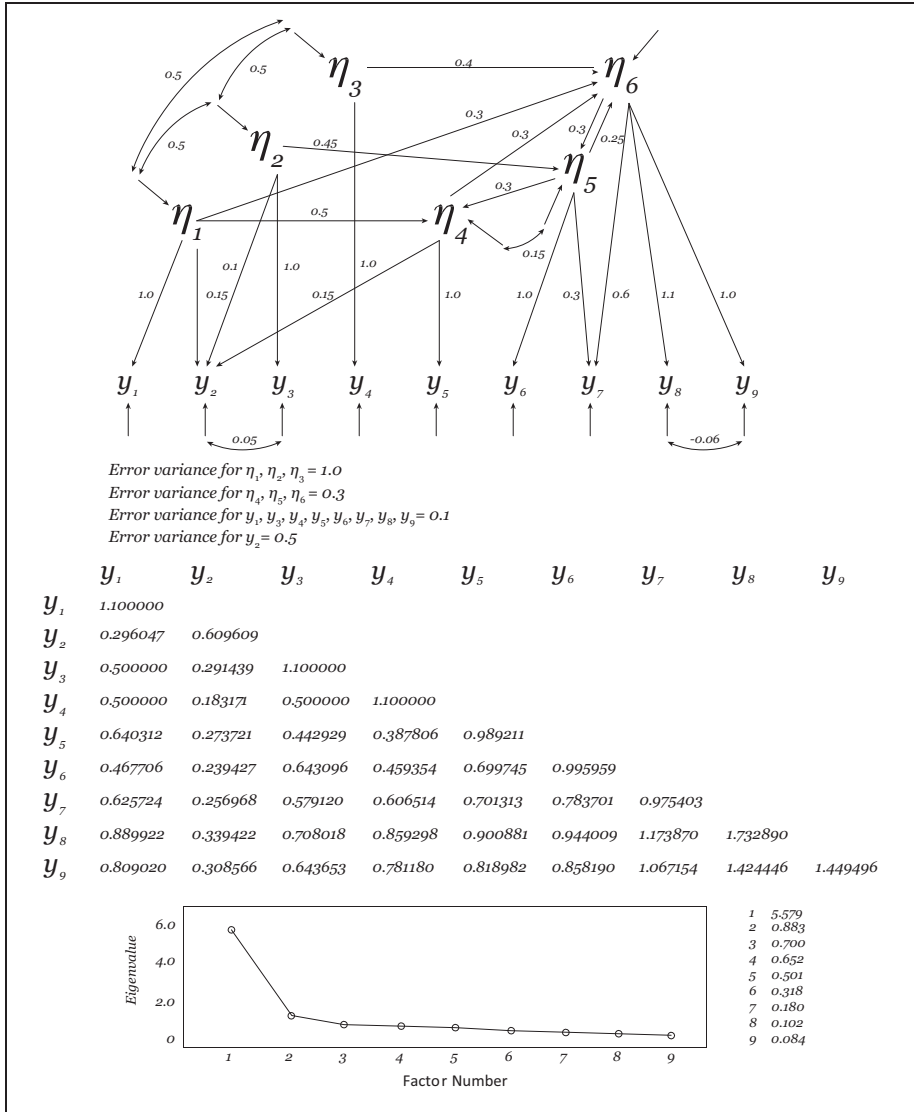


Figure 7. A SEM motif as almost one-factor.

was constructed to be importantly different than a one-factor model but not so different as to be brain-bending. This model results in a somewhat nonuniform covariance matrix but provides a scree plot and eigenvalues seemingly indicative of a single factor. In this instance the worldly covariance matrix is “significantly” ill fit by a

one-factor model ( $\chi^2 = 137.053$ ,  $df = 27$ ,  $p = .000$ ) with standardized loadings 0.668, 0.365, 0.581, 0.632, 0.747, 0.795, 0.956, 0.944, and 0.939.

This model illustrates that significant  $\chi^2$  values will sometimes (often?) do better than the eigenvalues and scree plot at recommending consideration of “ordinary” alternative models and not just wild models like Figure 6 or other factor models like Models 2 and 3. Here the scree plot and eigenvalues were correct in signaling that a one-factor model was the only serious contender from among the family of factor models but what the eigenvalues and scree could not signal was that even the strongest factor-model contender was deficient and seriously causally misspecified. In contrast the  $\chi^2$  test clearly reports the deficiency of the one-factor model. Chi-square from SPSS/PASW also reports that two-factors should fail ( $\chi^2 = 66.076$ ,  $df = 19$ ,  $p = .000$ ) and that three-factors should fail ( $\chi^2 = 29.028$ ,  $df = 12$ ,  $p = .004$ ). (Four-factors would be much more likely to fit  $\chi^2 = 1.278$ ,  $df = 6$ ,  $p = .973$ , but even the true indicator covariance matrix resulted in reported estimation difficulties with four factors.) Again, the eigenvalues and a bent-elbow scree plot do not provide assurance that the proper model is a factor model. The  $\chi^2$  test does a better job of informing the researcher that the proper model is not a factor model, but even the  $\chi^2$  test is not perfect.

## Discussion

Model 1 is not a one-factor model, but it produces uniform indicator covariances and correlations whose eigenvalues, scree pattern, and  $\chi^2$  test would traditionally be taken as convincing support for a one-factor model. The subsequent models illustrate how the traditional criteria can “support” a one-factor model despite the true model being three independent latent factors, three correlated factors, a regression style latent model, a model that is part factor like and part nonfactor like, a model that bends one’s mind with multiple loops and no exogenous latents, and a SEM style model. It is especially disconcerting to learn that it is possible for the eigenvalues and scree to “support” a single factor when the world actually contains three factors (as in Models 2 and 3). Even the  $\chi^2$  model test failed to detect the one-factor model’s misrepresentation of the true underlying causal structures, with the exceptions of Models 6 and 7.

Unfortunately, adjusting for random variations in eigenvalues (Ruscio & Roche, 2012) or using alternative scree-style criteria (Lorenzo-Seva, Timmerman, & Kiers, 2011) do not locate the proper number of latent variables for the models presented above. These factor models were fit to the population covariance matrices, and hence the above observations are not artifacts of sampling fluctuations. Introducing sampling fluctuations into the eigenvalues, when the factor model fits perfectly (e.g., Models 1, 2, 3, and 4), could reject the one-factor model only “by chance” and at the assigned  $\alpha$  or Type I error rate. Similarly, seeking scree-style hull-breakpoints in fit plots (Lorenzo-Seva et al., 2011) would be unable to locate the true number of latents when the population covariance data are improperly yet “perfectly” fit by a

single factor or by a wrong but fitting number of factors. Indeed, when the number of underlying latents exceeds the number of identifiable “factors” (as in Models 1, 4, 5, 6, and 7) there will not even be an identified factor-structured model having as many factors as required latents.

The scree plot and other rules of thumb deserve only part of the blame. The more fundamental difficulty originates in the potentially unsupported presumption that the world actually is causally factor-structured. There is no denying that the multitude of failing factor models might be signaling important model misspecifications, but the examples above demonstrate that even factor models satisfying the traditional criteria warrant skepticism. This raises the possibility of serious and widespread problems in the factor analytic literature.

The coefficients in the “worldly” models above were chosen to provide uniform (or relatively uniform) positive indicator covariance matrices that could be fit with correspondingly uniform loadings, but the problem being illustrated is not confined to uniform matrices. Factor models with nonuniform loadings can match correspondingly nonuniform covariance matrices, and hence the above examples present only a small fraction of the models and parameter values for which factor models fit when they ought not to fit because they are causally misspecified.

The fit provided by the wrong factor models above is not due to “equivalent models” because there are different numbers of coefficients in the illustrated models and the wrong-but-fitting factor models. The wrong but perfectly fitting factor models might be claimed to be “narrowly” (though not “broadly”) covariance fit equivalent (Hayduk, 1996), but none of the factor models are causally equivalent. Statisticians sometimes claim that covariance fitting models can be useful even if wrong, but researchers ought to question the usefulness of models whose causal structures misrepresent the world’s causal structure. Practitioners should be especially wary of significantly failing factor models because knowingly basing advice, decisions, or actions on a model that is significantly inconsistent with the evidence opens one up to being sued for malpractice.

The concern for causal properness applies to all models not just factor models. Nonfactor structural equation models can also fit when they ought not fit (namely when they are not properly causally specified). If we flipped the above demonstrations by specifying the world as factor structured and subsequently estimated the nonfactor models, some of the nonfactor models would inappropriately fit. Most of the illustrated models would not function well for such a demonstration (because they would be underidentified), but the point remains that it would be equally erroneous for true factor models to be mismodeled as nonfactor models. Model covariance fit, factor or otherwise, does not guarantee the model is properly causally specified.

Fortunately, this problem is considerably reduced in the context of general structural equation models because the variety of structural equation models results in these being more latent-theory oriented, less entrenched in disciplinary disregard for evidence of model ill fit, and less subject to model-replicating admonitions (such as if your factor model fails, add one more factor—which produces another

factor-structured model). The practice of attaining fit by routinely adding error covariances and other coefficients suggested by the modification indices seems as prevalent and problematic in SEM as in factor analysis, so SEM will also have its share of misspecified but fitting models.

The  $\chi^2$  test outperformed the eigenvalues and scree plots by clearly signaling misspecification of Model 7, and with large enough  $N$ s it might detect problems with Models 5 and 6 (Satorra & Saris, 1985), but it entirely missed the misspecification of Models 1, 2, 3, and 4. Small parameter adjustments to the worldly models currently being perfectly fit by wrong factor models would provide close-fitting but equally wrong-latent models. This implies that *close fit, or a small amount of covariance ill fit, does not confidently report that the model is close to being properly causally specified*. Maybe the fitting model is close to being causally proper, maybe not. This observation fundamentally challenges the trustworthiness of *all* covariance fit indices in determining the number of factors to retain—though some factor advocates seem not to have noticed (MacCallum, 2009; Mulaik, 2010).

If a one-factor model fails, factor analysts are often advised to try two factors (Lawley & Maxwell, 1971; MacCallum, 2009). A two-factor model will undoubtedly fit better than a failing one-factor model (because the two-factor model contains additional coefficients capable of matching a greater variety of covariance matrices), but the model's causal misspecification may, *or may not*, be resolved via an additional fit-improving factor. None of the above factor models would be rendered correct by inclusion of an additional factor. The ability of factor models with two, three, or more factors to progressively morph toward matching a greater variety of indicator covariance matrices recommends a corresponding escalation in concern for the factor model's causal specification—especially if the expanded factor model was prompted by prior ill fit or remains significantly ill fitting. It is imprudent to point to significantly ill-fitting factor models, with any number of factors, as likely to be close to being properly causally specified. And it is regrettable, but telling, that even senior researchers (e.g., Browne, MacCallum, Kim, Andersen, & Glaser, 2002) have stumbled over this point (as demonstrated by Hayduk, Pazderka-Robinson, Cummings, Levers, & Beres, 2005).

These observations implicitly warn researchers about a progressive reduction in the diagnostic utility of covariance residuals as the number of factors increases. Notice also that the modification indices (and expected parameter change) statistics will not assist in locating the proper model if the true model requires different latents than currently appear in a misspecified factor model. Modification indices can only recommend freeing overidentified coefficients that are currently fixed or constrained and hence can only “suggest” coefficients connected to the current potentially wrong latents (cf. Saris, Satorra, & van der Veld, 2009). Modification indices are severely limited in their ability to report that the latent factors themselves are problematic.

These pessimistic observations seem at odds with many “optimistic” simulations where traditional or improved factor rules of thumb seem to function well. For instance, the above examples illustrate severe *underestimation* of the appropriate



number of underlying latent variables, in contrast to the claim that “it is well known that Kaiser’s rule tends to *overestimate* the number of . . . common factors quite severely” (Lorenzo-Seva et al., 2011, p. 343, emphasis added). This disagreement originates partly in simulations avoiding features that render the eigenvalue and scree unable to properly report on factor-structured worlds such as Models 2 and 3. But the major difference arises because some latent variables are not “factors”—as in most of the above models. Unfortunately, even the latest approaches to locating factors (e.g., Lorenzo-Seva et al., 2011) will fail whenever *factor-fit* arrives before the proper number of *latents* is reached. Simulations are usually based on a variety of identified factor-structured alternative models, with one of those factor models being correct; rather than the factor model being assessed against models that would be incorrectly modeled as factor-structures, as above. Consequently, factor simulations tend to appear “optimistic” because they are set up in a way that avoids the kinds of difficulties illustrated above. Unfortunately, avoiding or disregarding fundamental problems does not eliminate them!

The simulation-presumption of a true underlying factor structure amounts to statisticians off-loading onto the researcher the onus for assuring that the statistician’s factor-enumerating procedure is applied to data originating from a worldly factor causal structure. In research contexts the worldly causal forces are often structured differently than initially imagined, and this renders the match between the causal requirements of the statisticians’ factor-enumerating procedures and the world’s causal structuring, speculative, tentative, and far from reassuring. Researchers often presume that linguistic similarity or the topical coherence of items is sufficient to warrant using a factor model. Regrettably, that is insufficient justification because there is no routine assurance that linguist similarity or topical coherence originates in common-cause structures, rather than in more intricate and complex causal interconnections. Even exact replicate measurements focusing on the same topic and requiring minimalist verbal expression can fail to match factor causal structuring and lead to investigation of alternative causal systems (Hayduk, 1994)—*if* the researcher attends to factor model failure and avoids a knee-jerk reflex to add a factor after encountering a failing factor model.

The rotational and estimation difficulties accompanying factor models have a long statistical history (Lawley & Maxwell, 1971), but notice that none of the one-factor models above displayed estimation problems, and characterizing the difficulty with models having more latents than indicators, as a factor-rotational problem would be obtuse at best. The relevant issues seem more naturally addressed as concern for proper causal specification, where the focus and onus shifts from the statistician’s cipher pad to the researcher’s tentative causal understandings and diagnostic efforts. There is no assured way to obtain a properly specified model, but researchers should attend to the strongest available testing of their models via  $\chi^2$  (Hayduk et al., 2007) and beware the pitfalls accompanying adding a factor or adding modification-index-suggested error covariances. Numerous research areas contain factor models: that do not fit, that were made to fit by adding one more factors, that were fit by adding error

covariances, or that had their failure “suppressed” rather than reported. This leaves the true worldly causal structures in serious doubt and begs for reexamination.

The most convincing assessment of potentially misspecified factor models is to stress or challenge the latent factor by including theorized latent-level causes and/or effects of the factor (Hayduk, 1996; Hayduk & Littvay, 2012). Introducing variables like age or gender as causes, or adding some other latents as effects, probes or challenges the factor causal structuring by adding rows/columns to the indicator covariance matrix. Those additional covariances interrogate or challenge the factor’s causal structure because the covariances between the new indicators and the factor’s indicators must be proportional to the magnitude of the factor’s causal impacts on its indicators. If the factor causal structuring is correct, whatever effects arrive at or depart from the latent factor will have their covariance implications appropriately distributed to the factor’s indicators by the factor’s causal-loadings. The inability of factor “loadings” to appropriately coordinate a factor’s indicators with the indicators of new latent causes/effects constitutes evidence against factor structuring (Hayduk & Littvay, 2012). Even a couple of latent causes or effects of a factor contribute substantial  $\chi^2$  power for detecting misspecified factor structures. It is not the “meaning” of the factor, or commonalities among the meanings of the indicators, that are involved in permitting a factor to survive this style of checkout. Introducing causes/effects of the latent factor interrogates the *causal* capacity of the latent factor to produce its indicators (Borsboom, Mellenbergh, & van Heerden, 2004; Hayduk & Glaser, 2000a, 2000b; Hayduk & Littvay, 2012; Hayduk & Pazderka-Robinson, 2007).

Introducing causes and/or effects of latent factors implicitly quells a long-standing disagreement regarding whether factors should be thought of as representing real-world features or whether factors are mere mathematical fictions that ought not be reified. Factor analysts historically viewed factor analysis as a scientifically helpful way of compacting and organizing worldly observations (Spearman, 1904; Thurstone, 1935; Cattell, 1973, 1978), even though the changing factor loadings accompanying factor rotations and oblique factors, along with factor score indeterminacy, made it impossible to “prove” that the chosen factor structure “is in any sense the correct one” (Thurstone, 1947, p. 124). For decades the simplicity and parsimony of factors stood as the scientific-justification for factor models, even as factor models were acknowledged as nonreifiable and potentially not world-matching. When the  $\chi^2$  test arrived, it could have been greeted as a way of testing the factor model’s causal match to the world, but instead factor models continued to be touted as parsimonious mathematical fictions whose simplicity and nonreified nature excused significant inconsistency with the covariance evidence (Browne & Cudeck, 1992; Browne et al., 2002).

SEM, in contrast, historically began as path analyses containing observed variables and hence reified variables were not an issue (Wright, 1921). When latents were introduced into SEM, these were easily incorporated as true-scores underlying the observed variables, and hence postulated-reification of latent variables comes

naturally to researchers having SEM backgrounds. Assessing factor structuring by introducing causes and/or effects of latent factors implicitly recommends reification of factors as potential real-world entities (Hayduk & Glaser, 2000a). Even if a factor causal structure fails, the researcher is prodded toward reification because justifying alternatives such as formative indicators (Bollen & Lennox, 1991), reactive indicators (Hayduk et al., 2007), or alternative causal connections between modeled latents inevitably appeals to real-world causal structures.

Evidence of factor model deficiency is likely to challenge junior and senior factor analysts in different ways. Senior researchers are likely to become nervous at the thought of reassessments potentially requiring retraction of claims they made in prior factor-based publications. Junior researchers are likely to be uncomfortable challenging the entrenched but problematic literature—especially if they are working under an afflicted senior researcher. The only antidote to both discomforts is a clear understanding of the statistical weakness of the eigenvalue and scree traditions, coupled with honest scientific respect for evidence pointing toward model causal misspecification. Unfortunately, interpersonal conflict and instances of deception, incompetency, and even dishonesty, can be expected in addition to more laudable responses.

There may be instances in which the world is indeed factor-structured, so the above demonstrations do not guarantee that the factor model will always be wrong. These observations should nonetheless spur reconsideration and reassessment of published factor models (whether covariance failing or fitting). Factor analysis does not let the data “speak for itself,” and the eigenvalue and scree-style rules can abet the telling of factor-fibs.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **References**

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411-423.
- Anderson, J. C., & Gerbing, D. W. (1992). Assumptions and comparative strengths of the two-step approach: Comment on Fornell and Yi. *Sociological Methods and Research, 20*, 321-333.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*, 815-824.
- Bekker, P. A., & ten Berge, J. M. F. (1997). Generic global identification in factor analysis. *Linear Algebra and Its Applications, 264*, 255-263.

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305-314.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061-1071.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230-258.
- Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*, 403-421.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.
- Cattell, R. B. (1973). *Factor analysis: An introduction and manual for the psychologist and social scientist*. Westport, CT: Greenwood Press.
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum Press.
- Chaplin, W. F. (2005). Factor analysis of personality measures. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science (Vol. 2, pp. 628-636)*. Chichester, England: Wiley.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317-327.
- Duncan, O. D. (1975). *Introduction to structural equation models*. New York, NY: Academic Press.
- Fornell, C., & Yi, Y. (1992a). Assumptions of the two-step approach to latent variable modeling. *Sociological Methods and Research*, *20*, 291-320.
- Fornell, C., & Yi, Y. (1992b). Assumptions of the two-step approach to latent modeling. *Sociological Methods and Research*, *20*, 334-339.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: Johns Hopkins University Press.
- Hayduk, L. A. (1994). Personal space: Understanding the simplex model. *Journal of Nonverbal Behavior*, *18*, 245-260.
- Hayduk, L. A. (1996). *LISREL issues, debates and strategies*. Baltimore, MD: Johns Hopkins University Press.
- Hayduk, L. A., Cummings, G. G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three—Testing the theory in structural equation models! *Personality and Individual Differences*, *42*, 841-850.
- Hayduk, L. A., & Glaser, D. N. (2000a). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, *7*, 1-35.
- Hayduk, L. A., & Glaser, D. N. (2000b). Doing the four-step, right-2-3, wrong-2-3: A brief reply to Mulaik and Millsap; Bollen; Bentler; and Herting and Costner. *Structural Equation Modeling*, *7*, 111-123.
- Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*, 159. doi:10.1186/1471-2288-12-159
- Hayduk, L. A., & Pazderka-Robinson, H. (2007). Fighting to understand the world causally: Three battles connected to the causal implications of structural equation models. In W. Outhwaite & S. Turner (Eds.), *Sage handbook of social science methodology* (pp. 147-171). London, England: Sage.

- Hayduk, L. A., Pazderka-Robinson, H., Cummings, G. G., Boadu, K., Verbeek, E. L., & Perks, T. A. (2007). The weird world, and equally weird measurement models: Reactive indicators and the validity revolution. *Structural Equation Modeling, 14*, 280-310.
- Hayduk, L. A., Pazderka-Robinson, H., Cummings, G. G., Levers, M.-J., & Beres, M. (2005). Structural equation model testing and the quality of natural killer cell activity measurements. *BMC Medical Research Methodology, 5*, 1.
- Heise, D. (1975). *Causal analysis*. New York, NY: Wiley.
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Joreskog, K. (1970). A general method for analysis of covariance structures. *Biometrika, 57*, 239-251.
- Joreskog, K., & Sorbom, D. (1996). *LISREL 8: Users reference guide*. Chicago, IL: Scientific Software.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed.). London, England: Butterworth.
- Lederman, W. (1937). On the rank of the reduced correlation matrix in multiple-factor analysis. *Psychometrika, 2*, 85-93.
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*, 340-364.
- MacCallum, R. (2009). Factor analysis. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 23-47). London, England: Sage.
- Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences, 42*, 875-881.
- Mulaik, S. A. (2007). There is a place for approximate fit in structural equation modelling. *Personality and Individual Differences, 42*, 883-891.
- Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.) Boca Raton, FL: CRC Press.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*, 282-292.
- Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecification? *Structural Equation Modeling, 16*, 561-582.
- Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83-89.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15*, 201-293.
- Spearman, C. (1927). *The abilities of man*. New York, NY: Macmillan.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences, 42*, 893-898.
- Thurstone, L. L. (1935). *The Vectors of Mind: Multiple-factor Analysis for the Isolation of Primary Traits*. Chicago, IL: University Press.
- Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of mind*. Chicago, IL: University of Chicago Press.

- Wright, D. B., & Villalba, D. K. (2012). Exploratory factor analysis. In G. M. Breakwell, J. A. Smith, & D. B. Wright (Eds.), *Research methods in psychology* (4th ed., pp. 279-318). London, England: Sage.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.