Proposal

Detecting Noninvariant Indicators in Confirmatory Factor Analysis under Partial Invariance

Pit Rieger
prieger@ethz.ch
19-951-102

August 17, 2021

1 TL;DR

Motivation: Omnibus tests of measurement invariance are conservative under partial invariance. Changing the model can help increase the quality of latent variable recovery, but requires knowledge about noninvariant items. Existing methods for detecting noninvariant items are complicated/difficult to implement/require additional assumptions.

Idea: Use regression form of CFA model to study residuals across groups to detect noninvariant manifest variables under partial invariance. Regression form allows to study properties of residuals for each indicator across groups. Measurement invariance should imply mean zero and uncorrelated residuals in each subset.

What's next: Formalize the idea and think about the question when it breaks down. Will then try to determine the settings of noninvariance under which this method would work to determine its usefulness.

2 Please give feedback/comments on

- Method/Literature: I was surprised that this idea isn't being used already. It appears almost too simple. Do applied researchers use this informally or am I overlooking something?
- Data: Other potential applications than Castanho Silva et al. (2020), perhaps outside political science and more in psychometric literature? Ideally single latent trait with many indicators.

3 Introduction & Motivation

Oftentimes, social scientists are interested in latent constructs rather than directly observable variables. To accommodate this, scholars frequently turn to methods such as **confirmatory factor analysis (CFA)**. However, for CFA to yield estimates of the latent construct that are comparable across groups (e.g. in cross-national research), we require **measurement invariance (MI)**. Different types of MI exist and are satisfied in the CFA model

$$\boldsymbol{Y} = \boldsymbol{\tau} + \Lambda \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

when for all groups g the same model structure is appropriate (configural invariance), the loadings are constant across groups such that $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$ (metric invariance), or both loadings and intercepts are constant across groups such that $\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}$ and $\boldsymbol{\tau}_g = \boldsymbol{\tau}$ (scalar invariance) (Davidov et al., 2014, for an overview, see). In practice, scholars implicitly assume (some of) these properties when interpreting CFA models that were fitted on the full data. It is therefore good practice to test whether they are satisfied, which can be done in several ways. Typically, scholars reason about MI on the basis of the χ^2-distributed likelihood ratio tests of a so-called configural model, i.e. a model where parameters are estimated freely for each group, and a model where parameters have equality constraints across groups (Jöreskog, 1971, c.f. multi-group CFA (MGCFA)). However, under partial invariance (Byrne et al., 1989, c.f.), this approach as well as its alternatives may be unnecessarily conservative. For example, Pokropek et al. (2019) show with a simulation study that freeing noninvariant items from their equality constraint across groups (or removing these items altogether) can yield very good results under a partial invariance setting. Clearly, this hinges on researchers' ability to detect noninvariant manifest variables. Thus, detection of noninvariant items can be considered a crucial first step in analyzing MI and an important aspect for increasing the quality of the recovery of latent variables. However, the detection of noninvariant items in the CFA model remains rare and existing methods are complicated, which may be discouraging applied researchers from going beyond omnibus tests of overall measurement invariance.

4 Existing Approaches

De Roover et al. (2014) list some prominent methods for detecting noninvariant items within the MGCFA framework before presenting a novel approach of their own. First, the sequential model modification procedure (MacCallum, 1986; MacCallum et al., 1986) relaxes equality constraints on the basis of modification indices to alleviate configural or metric invariance. Second, a fairly simple procedure consists of removing individual items whenever a model with all but that item is significantly closer to being measurement invariant (Byrne & Van de Vijver, 2010; Cheung & Rensvold, 1999). However, De Roover et al. (2014, p.1) point out that all existing detection methods "require researchers to run a multitude of analyses and [...] imply assumptions that are often questionable". As a remedy, they propose a method based on clustering which doesn't require such high numbers of invariance tests. Unfortunately, how and why their method works is by no means obvious.

Moreover, outside the realm of frequentist methods, Bayesian alternatives exist (Barendse et al., 2014, e.g.). These should also be explored later, but were skipped for now. However, since applied researchers tend to use the standard frequentist implementations of CFA (e.g. the `lavaan` package in `R`), the switch to Bayesian methods for the detection step of their analysis likely poses an additional complication.

5 Idea for a New Method

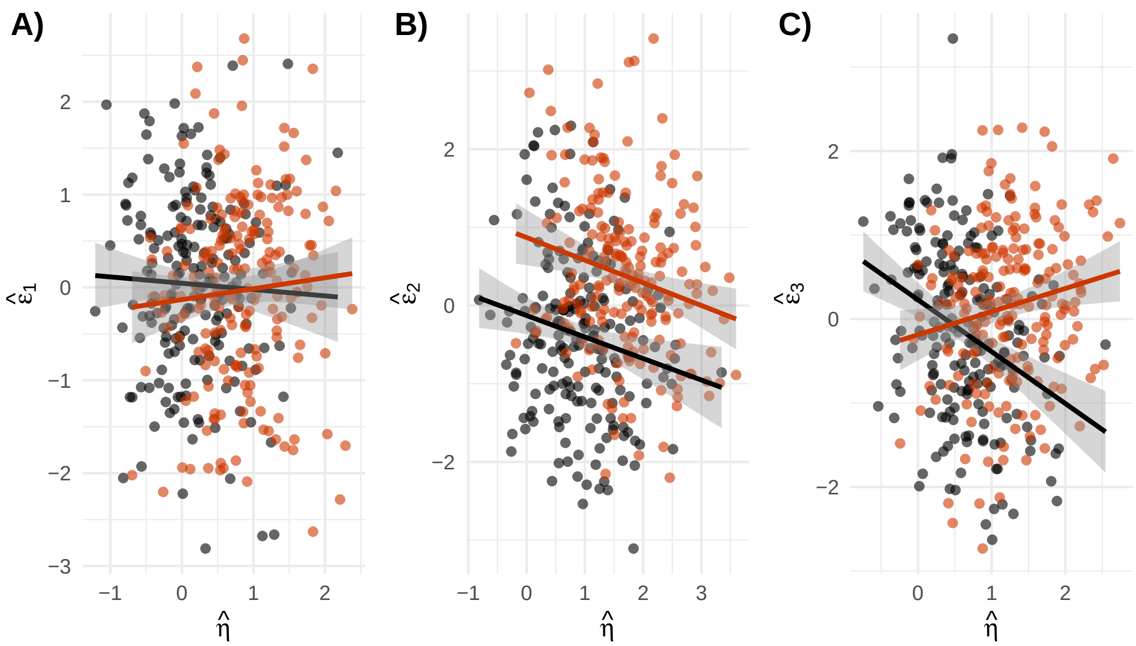

Figure 1: *Illustration of the residuals from three DGPs across two groups (black and red). Panel A shows the setting where invariance is given and the data come from the same DGP and the only difference across groups is a shift in the expectation of the latent variable. In panel B, scalar invariance is violated and the red group's item-specific intercept is shifted up by 0.9 units. In panel C, metric invariance is violated and the red group's loading on η is increased by 0.5 units.*

I will now sketch an idea for detecting noninvariant items that doesn't require fitting many differently configured models and appears to be much more intuitive than existing approaches. The fundamental idea is to leverage the fact that the CFA model takes the form of a linear regression function. This means that we can compute residuals from regressing observable indicators Y_j on the previously estimated latent scores $\hat{\eta}$. Depending on whether item j satisfies invariance or not, we would expect to observe different patterns in the residuals. Figure 1 visualizes the differences between residuals from a data-generating process (DGP) that emulates measurement invariance or violations thereof. Recall, that, by construction, residuals have mean zero and are uncorrelated with their linear predictor when we're considering all data. Yet, these properties do not necessarily hold for subsets of the full data that was used to fit the regression. As a result, when analyzing the subset of residuals for a group, the residual mean needn't be zero and the residuals may be correlated with their linear predictor. In fact, measurement invariance should imply that this holds for all groups and in turn cases where it doesn't indicate noninvariance[1]. It should therefore be possible to create a procedure that conducts tests on these residuals in order to detect noninvariant items.

The preliminary goal of my master thesis will therefore be to develop such a procedure that formalizes how one can validly test these properties under different settings such as number of indicators, number of groups, types of measurement invariance, etc. One anticipated drawback is that this can only work if $\hat{\boldsymbol{\eta}}$ is an adequate estimator. I will therefore begin by studying different settings of noninvariance to determine the limitations of this method. This can be done with a simulation study of detection rates and also enables me to compare it with other approaches. Finally, I aim to apply this method to real-world data. The top contender so far are replication survey data for a comparison of different scales of populist attitudes by Castanho Silva et al. (2020). Their survey was conducted in 9 different countries and includes a great number of questions that are commonly used for constructing different populism scales. This data will allow me to apply the method to the differently specified CFA models from the populism literature and compare their individual invariance properties.

[1] Note that the general idea is very similar to parent detection in invariant causal prediction (Peters et al., 2016)

References

Barendse, M. T., Albers, C. J., Oort, F. J., & Timmerman, M. E. (2014). Measurement bias detection through bayesian factor analysis. *Frontiers in Psychology*, *5*, 1087. https://doi.org/10.3389/fpsyg.2014.01087

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456.

Byrne, B. M., & Van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132.

Castanho Silva, B., Jungkunz, S., Helbling, M., & Littvay, L. (2020). An empirical comparison of seven populist attitudes scales. *Political Research Quarterly*, *73*(2), 409–424. https://doi.org/10.1177/1065912919833176

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*(1), 1–27.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

De Roover, K., Timmerman, M. E., De Leersnyder, J., Mesquita, B., & Ceulemans, E. (2014). What's hampering measurement invariance: Detecting non-invariant items using clusterwise simultaneous component analysis. *Frontiers in Psychology*, *5*, 604. https://doi.org/10.3389/fpsyg.2014.00604

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426.

MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*(1), 107–120.

MacCallum, R., Roznowski, M., & Necowitz, L. B. (1986). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504.

Peters, J., Bühlmann, P., & Meinshausen, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *78*(5), 947–1012. http://www.jstor.org/stable/44682904

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A monte carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 724–744. https://doi.org/10.1080/10705511.2018.1561293