model. With regard to a threshold, Byrne and Van de Vijver (2010) consider an item to be non-invariant if its deletion increases the CFI by 0.01 relative to the baseline model. Justification for the value of 0.01 is provided by Cheung and Rensvold (2002), who consider it to be the critical value for overall measurement invariance. To summarize, the estimated set of non-invariant items S_{BV} of this procedure is given by

$$S_{\mathrm{BV}} := \left\{ i \;\middle|\; \mathrm{CFI}\left(\mathcal{M}_{(-i)}^{\mathrm{strong}}\right) \geq \mathrm{CFI}\left(\mathcal{M}^{\mathrm{strong}}\right) + 0.01 \right\} \tag{46}$$

4.2 A Novel Approach to Non-invariant Item Detection

The novel approach proposed and tested in this thesis deviates from the existing approaches in several ways. Most fundamentally, instead of being based on model comparison or relying on the model parameters *per se*, it makes direct use of the implications of the relationship between latent variables and items in CFA models. More specifically, it is based on the residuals of these linear relationships. The use of residuals for diagnostic purposes in a CFA framework is by no means an innovative idea. For example, Costner and Schoenberg (1973) use correlations of all items' residuals to identify relationships that are missing from the model structure of the specified model.[13] Regardless, to the best of my knowledge, residual analysis has not been used for detecting non-invariance. The original contribution of this thesis is therefore to devise such a method for detecting non-invariant items under partial MI by studying patterns in the residuals of the linear relationships of a given CFA model. I begin with an introduction of the fundamental logic of this method and prove that residuals can theoretically be used to identify non-invariance by laying out the residual behavior under MI or violations thereof. I then provide visual examples to further strengthen the intuition behind the idea. Finally, I introduce the actual implementation of the method and how it may be further improved with a step-wise extension.

4.2.1 Residual Behavior

For the introduction, assume a single latent factor model with p items that load on the latent variable η. We further consider a single item for which we want to determine its MI status. For the moment, we can thus omit the subscript i and denote the item Y. Regressing Y on η yields a regression intercept γ and slope β which allow us to write the residuals r in the regression as

$$r := Y - (\gamma + \beta\eta). \tag{47}$$

Note, that by construction the residuals of this regression (or any simple linear regression,

[13] However, they also caution that "this approach can be very misleading" by providing some examples where a modified model is not in line with the true data generating model (Costner & Schoenberg, 1973, p.172)

for that matter) satisfy the following two properties:

$$\mathbb{E}[r] = 0 \;\&\tag{48}$$

$$\mathrm{Cov}(r, \eta) = 0.\tag{49}$$

However, these two properties do not necessarily hold for groups within the data. To see this, the true data-generating model of the item must be allowed to vary across groups. In terms of MI, recall that there are four different group-specific DGPs: A simultaneous violation of both metric and scalar MI, a violation of metric MI, a violation of scalar MI, and the case of MI. Let Y^l denote the group-specific item and η^l the group-specific latent variable for group $l = 1, ..., g$, where η^l is a random variable with expectation μ^l and variance ϕ^l. Explicitly writing out the four DGPs in this notation gives

- **Simultaneous violation of metric and scalar MI:**

$$Y^l = \tau^l + \lambda^l \eta^l + \varepsilon\tag{50}$$

  where $\tau^l \neq \tau^{l'}$ and $\lambda^k \neq \lambda^{k'}$ for some $l, l', k, k' \in \{1, ..., g\}$.
- **Violation of metric MI:**

$$Y^l = \tau + \lambda^l \eta^l + \varepsilon\tag{51}$$

  where $\lambda^l \neq \lambda^{l'}$ for some $l, l' \in \{1, ..., g\}$
- **Violation of scalar MI:**

$$Y^l = \tau^l + \lambda \eta^l + \varepsilon\tag{52}$$

  where $\tau^l \neq \tau^{l'}$ for some $l, l' \in \{1, ..., g\}$
- **Perfect MI:**

$$Y^l = \tau + \lambda \eta^l + \varepsilon.\tag{53}$$

Given these group-specific DGPs, the residuals in group l can be written as

$$r^l \coloneqq Y^l - \left(\gamma + \beta \eta^l\right)\tag{54}$$

where it is important to note that γ and β are still the slope of a linear regression of Y on η in the pooled data and not just in group l. In other words, the regression parameters are obtained while ignoring any (potential) grouping in the data.

To further clarify what that entails, it is instructive to show how the pooled Y and η can be written in terms of their group-specific constituents with the help of a multinomial random variable representing group-membership. Let

$$G \overset{iid}{\sim} \mathrm{Multinomial}(1, \boldsymbol{\pi})\tag{55}$$

with

$$\boldsymbol{\pi} = \left(\pi^1, ..., \pi^g\right)^\mathsf{T} = \frac{1}{N}\left(n^1, ..., n^g\right)^\mathsf{T} \tag{56}$$

being membership probabilities equal to the relative group sizes such that $\sum_{l=1}^{g} \pi^k = 1$ and $\pi^k \in (0, 1)$. The pooled random variable Y can then be defined as

$$Y \coloneqq \sum_{l=1}^{g} G^l Y^l \tag{57}$$

and the pooled η as

$$\eta \coloneqq \sum_{l=1}^{g} G^l \eta^l. \tag{58}$$

We have the following result for the regression of these two variables.

Lemma 4.1. *Under perfect MI, regressing Y on η while ignoring the group structure yields intercept γ and slope β*

$$\gamma = \tau \tag{59}$$

$$\beta = \lambda \tag{60}$$

Proof. The regression coefficient β from regressing Y on η is given by

$$\begin{aligned} \beta &= \frac{\mathrm{Cov}\left(Y, \eta\right)}{\mathrm{Var}\left(\eta\right)} \overset{(53)}{=} \frac{\mathrm{Cov}\left(\tau + \lambda\eta + \varepsilon, \eta\right)}{\mathrm{Var}\left(\eta\right)} \\ &= \frac{\mathrm{Cov}\left(\lambda\eta, \eta\right)}{\mathrm{Var}\left(\eta\right)} \\ &= \lambda \end{aligned} \tag{61}$$

and the intercept γ by

$$\begin{aligned} \gamma &= \mathbb{E}\left[Y\right] - \beta\mathbb{E}\left[\eta\right] \overset{(53)}{=} \mathbb{E}\left[\tau + \lambda\eta + \varepsilon\right] - \beta\mathbb{E}\left[\eta\right] \\ &= \tau + \lambda\mathbb{E}\left[\eta\right] - \beta\mathbb{E}\left[\eta\right] \overset{(61)}{=} \tau + \lambda\mathbb{E}\left[\eta\right] - \lambda\mathbb{E}\left[\eta\right] \\ &= \tau. \end{aligned} \tag{62}$$

□

The fundamental idea of the new method is that the result of Lemma 4.1 cannot apply under any violation of metric or scalar MI for all group-specific CFA loadings and intercepts. The intuition can be reduced very simply to this: We conduct only a single regression of Y on η, yielding a single intercept and slope. At the same time, the nature of a violation

of metric or scalar MI is that at least two groups have a different intercept or loading. As a result, under an MI violation, there are at least three parameters relevant for the true DGP at the pooled level. For example, suppose a two-group setting with groups A and B and a pure scalar MI violation. The DGP thus contains two group-specific intercept parameters, τ^A and τ^B, and the loading λ which is shared by both groups. Even if in the regression, $\beta = \lambda$ (which is not true in general), it is still impossible for $\gamma = \tau^A = \tau^B$ because $\tau^A \neq \tau^B$. In other words, for some or all groups, regression intercept and slope will not be equal to their group-specific intercept and loading in the true DGP. The exact difference between the group-specific parameters in the DGP and the regression parameters can be computed and I derive them in the appendix for further illustration. However, they are not necessary for the remainder of this section. Instead, it is important to realize that this discrepancy results in systematic patterns with regard to the expectation of the residuals and their covariance with the latent variable for some or all of the groups. To formalize this, the behavior of the residuals under different violations of MI can be condensed in the following theorem.

Theorem 4.2. *Iff metric and scalar MI are satisfied,*

$$\mathbb{E}\left[r^l\right] = 0, \textit{ and} \tag{63}$$

$$\mathrm{Cov}\left(r^l, \eta^l\right) = 0 \tag{64}$$

for all $l = 1, ..., g$.

Proof. We begin with the case of a simultaneous violation of metric and scalar MI.

$$\begin{aligned} \mathrm{Cov}\left(r^l, \eta^l\right) &\overset{(50)}{=} \mathrm{Cov}\left(\tau^l + \lambda^l\eta^l + \varepsilon - \gamma - \beta\eta^l, \eta^l\right) \\ &= \mathrm{Cov}\left(\lambda^l\eta^l - \beta\eta^l, \eta^l\right) \\ &= \left(\lambda^l - \beta\right)\mathrm{Var}\left(\eta^l\right). \end{aligned} \tag{65}$$

Because $\mathrm{Var}\left(\eta^l\right) > 0$,

$$\mathrm{Cov}\left(r^l, \eta^l\right) = 0 \iff \lambda^l = \beta. \tag{66}$$

For this to hold in all groups, we require that $\beta = \lambda^1 = \lambda^2 = ... = \lambda^g$ which contradicts the assumed metric violation.

The same argument can be made in the presence of a sole violation of metric MI. Formally,

$$\begin{aligned} \mathrm{Cov}\left(r^l, \eta^l\right) &\overset{(51)}{=} \mathrm{Cov}\left(\tau + \lambda^l\eta^l + \varepsilon - \gamma - \beta\eta^l, \eta^l\right) \\ &= \left(\lambda^l - \beta\right)\mathrm{Var}\left(\eta^l\right). \end{aligned} \tag{67}$$

So the same logic that β cannot be equal to λ^l for all l applies.

Next, consider the case of a violation of scalar MI. We have

$$\begin{aligned}\mathbb{E}\left[r^l\right] &\overset{(52)}{=} \mathbb{E}\left[\tau^l + \lambda\eta^l + \varepsilon - \gamma - \beta\right] \\ &= \tau^l - \gamma + (\lambda - \beta)\mu^l.\end{aligned} \tag{68}$$

If $\mu^l = 0$ for all l, we can argue analogously as for the loadings above that

$$\mathbb{E}\left[r^l\right] = \tau^l - \gamma = 0 \iff \tau^l = \gamma \tag{69}$$

which contradicts the violation of scalar MI.

If $\mu^l \neq 0$ for some or all l, $\mathbb{E}\left[r^l\right] = 0$ iff

$$\tau^l - \gamma = (\beta - \lambda)\mu^l. \tag{70}$$

In other words, the difference between the regression intercept and the group-specific intercept in the DGP may be cancelled out and the expected residual may be equal to zero. Suppose there was some constellation of all group-specific intercepts and all latent variable expectations that ensured that $\mathbb{E}\left[r^l\right] = 0$ for all l then this implies that

$$(\beta - \lambda) \neq 0 \iff \beta \neq \lambda. \tag{71}$$

In turn, this yields the following result:

$$\begin{aligned}\text{Cov}\left(r^l, \eta^l\right) &\overset{(52)}{=} \text{Cov}\left(\tau^l + \lambda\eta^l + \varepsilon - \gamma - \beta\eta^l, \eta^l\right) \\ &= \text{Cov}\left((\lambda - \beta)\eta^l, \eta^l\right) \\ &= (\lambda - \beta)\text{Var}\left(\eta^l\right) \neq 0\end{aligned} \tag{72}$$

because $\text{Var}\left(\eta^l\right) > 0$.

In summary, the steps above conclude half of the iff statement in the theorem and prove that a violation of either or both types of MI results in the group-specific residuals having non-zero expectation or non-vanishing correlation with the latent variable. To conclude the proof we further need to show that under perfect MI, the two components hold. We have that

$$\begin{aligned}\mathbb{E}\left[r^l\right] &\overset{(53)}{=} \mathbb{E}\left[\tau + \lambda\eta^l + \varepsilon - \gamma - \beta\eta^l\right] \\ &= \tau + \lambda\mu^l - \gamma - \beta\mu^l\end{aligned} \tag{73}$$

In accordance with Lemma 4.1, we can replace $\gamma = \tau$ and $\beta = \lambda$ and as a result it is easy to see that

$$\mathbb{E}\left[r^l\right] = \tau - \tau + \lambda\eta^l - \lambda\eta^l = 0. \tag{74}$$

Similarly, we have

$$\begin{aligned} \text{Cov}\left(r^l, \eta^l\right) &\overset{(53)}{=} \text{Cov}\left(\tau + \lambda\eta^l + \varepsilon - \gamma - \beta\eta^l, \eta^l\right) \\ &= \text{Cov}\left(\varepsilon, \eta^l\right) \\ &= 0. \end{aligned} \tag{75}$$

which completes the second half of the iff statement and concludes the proof. □

4.2.2 Illustrating Residual Behavior

To further illustrate the implications of Theorem 4.2, Figure 3 visualizes an example of the patterns in the residuals of a single item for two groups for the four DGPs. Across all panels, the groups differ in their true latent mean with the red group scoring higher than the black group. The residuals were obtained from a single linear regression that ignores the group structure of the data. In a next step, these residuals are regressed on the latent variable using separate regressions for each group. The fitted lines of these secondary regressions are shown as solid lines in the figure. Panel A corresponds to perfect MI where the two groups originate from identical DGPs. The fitted lines show that both groups have a mean close to zero and that there is no correlation with the latent variable in either group. In panel B, scalar invariance is violated and the red group's intercept is shifted up by 0.9 units compared to the black group. As a result, there is a slight difference in the residual mean between the groups and a slight correlation with the latent variable in both groups. In panel C, metric invariance is violated and the red group's loading on η is increased by 0.5 units compared to the black group. Finally, panel D combines both violations using the same values. In both panels, the correlation with the latent variable in the black group is very pronounced. Generally, across the panels, it is easy to see that the two criteria of zero mean and vanishing correlation in the groups only hold in the MI setting shown in panel A. For any violation of MI, panels B-D show that the residuals exhibit deviations from at least one of these criteria in at least one of the two groups.

Furthermore, the example shows that it's possible to visualize the MI status of the items in CFA models, which may be helpful for applied researchers. The methods required for creating such visualizations are the bread and butter of empirical researchers and could therefore contribute to spreading the use of item-level detection methods for violations of MI. The existing methods, on the other hand, don't have this advantage. Instead, their reliance on ML theory and the use of many submodels may pose a serious hurdle for newcomers to CFA and MI.

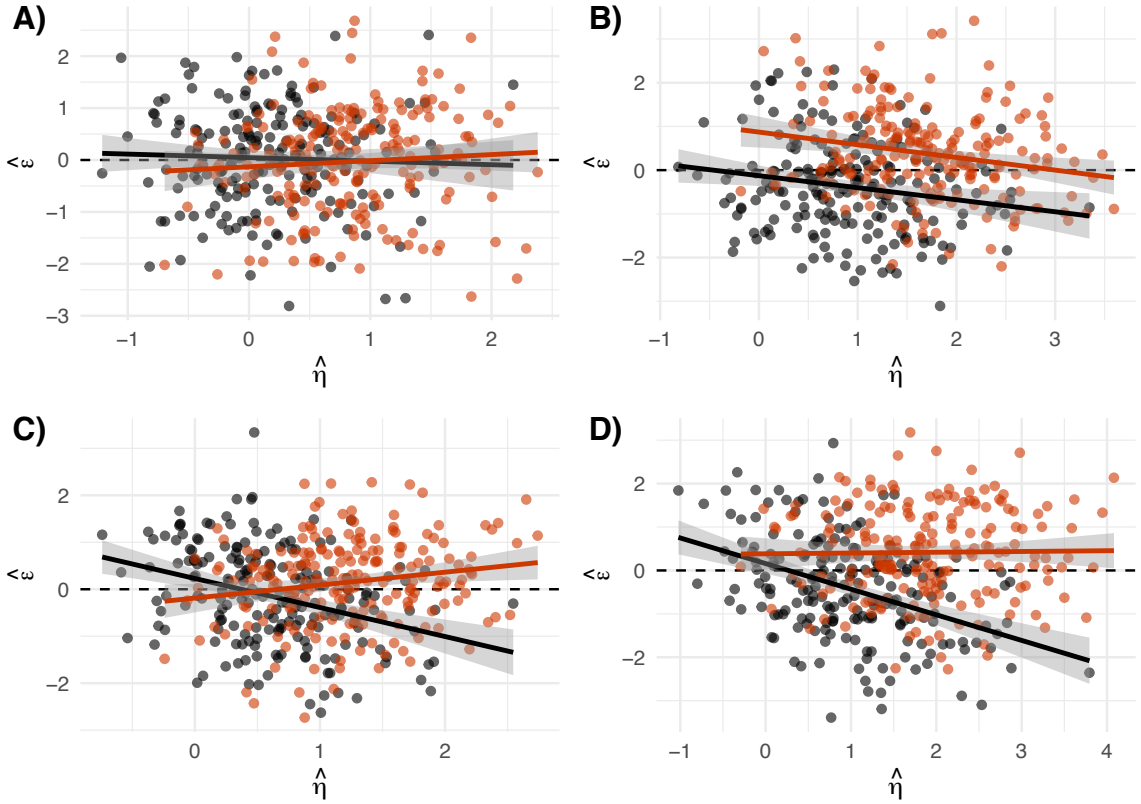

Figure 3: *Illustration of the residuals from four DGPs across two groups (black and red).*

4.2.3 Implementation as a Detection Method (R1)

To turn these results into a detection methods, the two components of a) across-group comparisons of the residual means and b) correlations between the residuals and the latent variable within each group must be formalized for hypothesis testing. In the following, I show that the first component can be formulated as a standard one-way analysis of variance (ANOVA) and the second component can be studied by means of a coefficient test which is equivalent to a correlation test in the absence of control variables. I then describe how to aggregate these two components to yield a test of non-invariance at the item level.

Before going into details, it is important to note the obvious obstacle for devising a method from the results above. The key problem is the very raison d'être of CFA itself: the latent variable η is unknown. Therefore, we cannot really regress Y on η. The obvious - albeit naive - solution is to rely on estimates $\hat{\eta}$ instead which has the obvious shortcoming that these estimates are problematic under non-invariance. Unfortunately, there is no other way around this issue in a world in which latent variables exist. It is the price to pay for the simplicity of the novel approach. The implication of this issue is that we at least require a setting of partial MI such that the estimates at least resemble the "true" latent variables. I return to this issue and the amelioration of its consequences with the implementation of a step-wise version of the detection method below.

In the following, we turn to sample versions and estimates, not just of the latent variable,

but also of the regression parameters, the residuals, etc.. Moreover, we again consider all $i = 1, ..., p$ items for which we want to test for MI. We can thus write the residuals in group l as

$$\hat{r}_i^l = Y_i^l - \hat{\gamma}_i - \hat{\beta}_i \hat{\eta}^l \tag{76}$$

where $\hat{\gamma}_i$ and $\hat{\beta}_i$ are the estimated parameters of the linear regression of Y_i on $\hat{\eta}$ with the pooled data.

Recall that for the first component, the implication of an invariant item is that the residuals have the same expectation of zero across all groups. Let

$$\nu_i^l := \mathbb{E}\left[\hat{r}_i^l\right] \tag{77}$$

denote the expected residual for item i in group l. For each item, the null hypothesis for the ANOVA component of the procedure can be written as the global null hypothesis of a single-mean model, i.e.

$$H_0 : \nu_i^1 = ... = \nu_i^g, \tag{78}$$

which can be conducted with an F-test. More specifically, the test statistic for item i with $N = n^1 + ... + n^g$ samples is given by the ratio of treatment and error mean squares:

$$F_i := \frac{MST_i}{MSE_i} = \frac{\frac{1}{(g-1)} \sum_{l=1}^{g} n^l \left(\hat{\nu}_i^l - \hat{\nu}_i\right)^2}{\frac{1}{(N-g)} \sum_{l=1}^{g} \sum_{j=1}^{n^l} \left(\hat{r}_{ij}^l - \hat{\nu}_i^l\right)^2} \overset{H_0}{\sim} F_{(g-1),(N-g)}, \tag{79}$$

where $\hat{\nu}_i^l$ and $\hat{\nu}_i$ are the sample means of the residuals in group l and in the full sample of size N, respectively. Note, that the latter sample mean is zero by construction. However, it is kept in equation (79) to emphasize that the statistic is the standard ANOVA F-test. For a more detailed account, refer to an introductory ANOVA book, e.g. Oehlert (2000).

With regard to the second component of vanishing correlation between the residuals and the latent variable, the linear relationship can be tested via the corresponding coefficients in separate linear regressions of $\hat{\eta}^l$ on $\hat{r}_i^l$ for each item and each group.[14] It is important to stress that this is a secondary regression after already having obtained the residuals from a first regression. In total, the method entails a total of pg secondary regressions. Let κ_i^l and ω_i^l denote the regression intercepts and slopes of the secondary regressions. The null hypothesis of no correlation can then be written as

$$H_0 : \omega_i^l = 0. \tag{80}$$

[14] Equivalently, this can of course be done in a single regression where $\hat{\eta}$ is interacted with the group.

More specifically, for each item and group the following test statistic applies:

$$\frac{\hat{\omega}_i^l}{\text{se}\left(\hat{\omega}_i^l\right)} \overset{H_0}{\sim} t_{n^l-2} \tag{81}$$

and an estimator for the standard error of the coefficient is given by

$$\hat{\text{se}}\left(\hat{\omega}_i^l\right) = \sqrt{\frac{\sum_{j=1}^{n^l}\left(\hat{r}_{ij}^l - \hat{\kappa}_i^l - \hat{\omega}_i^l\hat{\eta}_j^l\right)^2}{(n^l-2)\sum_{j=1}^{n^l}\left(\hat{\eta}_j^l - \frac{1}{n^l}\sum_{j=1}^{n^l}\hat{\eta}_j^l\right)^2}}. \tag{82}$$

Again, details can be found in any standard introductory statistics textbook introducing linear regression, e.g. Fahrmeir et al. (2013).

What remains to be done is to aggregate the two components to a single test at the item level. To this end, first note that for each item, the procedure is comprised of $g+1$ individual tests: One global ANOVA test for the residual means and g correlation tests. Let $\wp_{iu}$ denote the p-value of the u^{th} test for item i. These p-values can be aggregated by simply considering the minimal p-value for each item and applying a Bonferroni correction such that

$$\widetilde{\wp}_i \coloneqq (g+1)\min(\wp_{i,1}, ..., \wp_{i,g+1}), \tag{83}$$

is a p-value for the test of item i's non-invariance. In a next step, a Holm-Bonferroni correction at the level of items can be applied to yield the set of identified non-invariant items for which the item-level test is rejected:

$$S_{\text{R1}} \coloneqq \left\{ i \;\middle|\; (p-\pi(i)+1)\widetilde{\wp}_i < \alpha \right\}, \tag{84}$$

where $\pi(i)$ denotes the position of the i^{th} item when arranging the p-values in ascending order and α the significance level.

4.2.4 Step-wise Version (R2)

As noted previously, the clearest drawback of this approach is that it hinges on $\hat{\eta}$ being reasonably close to the true latent variable. If MI is violated strongly, i.e. for many items, then $\hat{\eta}$ has little resemblance with the true latent variable even if the model structure, i.e. the relationships between latent variables and items is correct. Thus, this, but also the other existing approaches would fail to correctly distinguish non-invariant items. However, the degree of non-invariance at which this and existing methods still work can only be studied with simulations. One potential way of ameliorating this issue is to implement a step-wise version of this approach. Starting with a given CFA model of structure $\mathcal{M}$, the step-wise

approach first selects the item with the lowest p-value in the R1 approach. If its p-value is below the given significance level, it is removed from the model entirely and the next iteration of the process begins. The motivation for doing so is to incrementally improve the estimates of the latent variable by refitting the model with the worst item removed in each iteration until no more items are detected as being non-invariant for a given α-level. Note, that this idea of improving the model by removing items entirely has some resemblance with the approach by Byrne and Van de Vijver (2010). Further note that the fundamental idea of the original method $R1$ still applies. In fact, computationally, its implementation can simply be reused in each iteration of the step-wise approach. Instead of using the set-builder notation, the set of non-invariant items in a CFA model of structure $\mathcal{M}$ can best be described by the following algorithm:

1. Set $t = 0$, $S_{\text{R2}} = \{\}$
2. Apply R1 to the initial model structure $\mathcal{M}_0 = \mathcal{M}$, yielding $\widetilde{\wp}$
3. while $((p - t) \min \widetilde{\wp} < \alpha)$ {
   (a) $S_{\text{R2}} = S_{\text{R2}} \cup \arg\min_i \widetilde{\wp}_i$
   (b) Update model structure $\mathcal{M}_{(t+1)} = \mathcal{M}_t^{(-\arg\min_i \widetilde{\wp}_i)}$, removing item $\arg\min_i \widetilde{\wp}_i$
   (c) Apply R1 to $\mathcal{M}_{t+1}$ and update $\widetilde{\wp}$
   (d) $t = t + 1$
   }

~~4.3 Implementation~~

~~All detection methods were implemented in the `R` programming language (R Core Team, 2020; v4.0.2) and are publicly available in the GitHub repository for this thesis.[15] Every step within the CFA framework, i.e. model fitting, testing, etc., was done using the `lavaan` package (Rosseel, 2012; v0.6-9).~~

4.4 Detection of Violations of Metric MI

~~As discussed above, the detection methods do not distinguish between the type of MI, i.e. metric or scalar, that is violated by any given item. Nonetheless, it is relatively straightforward to implement versions of the detection methods, that only consider violations of metric MI, i.e. the weak requirement for MI. In general, the expectation for these methods is that they classify a subset of the items that are detected by the standard versions.~~

~~In the following, I briefly describe for each method how it needs to be altered in order to focus on the invariance of loadings while disregarding the intercepts. All other steps of the~~

~~[15] https://github.com/pitrieger/masterthesis/tree/main/Rscripts/simulation for models with a single latent variable and https://github.com/pitrieger/masterthesis/tree/main/Rscripts/application for models with multiple latent variables (see section 6 for further details.)~~

methods remain unchanged compared to their original version. For the R approach, I go into slightly greater detail by proving that the residuals enables the drawing of conclusions about the type of violation. Note that I excluded the J approach from the application section because its theoretical weaknesses were corroborated by the results of the simulation study.

4.4.1 Modification Indices (MInd)

Recall that in the standard implementation of the MInd approach, the baseline model constrains the loadings and intercepts across groups to equality and computes a modification index for simultaneously lifting both types of constraints for each item. The metric version works identically, but only constrains the loadings for the baseline model and modification indexes refer to the lifting of single loading constraints. Therefore, the degrees of freedom of the LR test also need to be adjusted accordingly. As a result, the metric version completely disregards the intercepts and lets them vary freely across groups in the baseline model as well as the comparison models.

4.4.2 Cheung & Rensvold (CR)

Similarly, for the metric version of the CR approach, in the construction of $\mathcal{M}_{\overline{ij}}^{\text{base}}$, the equality constraints are only imposed on the loading parameter of item i instead of both the loading and intercept. As with the metric MInd method, the degrees of freedom in the LR test are also adjusted accordingly. Recall that Cheung and Rensvold (1999) are agnostic about the intercepts. Consequently, this may even be their originally intended method.

4.4.3 Byrne & Van de Vijver (BV)

For the BV method, a metric version can be obtained by changing the baseline model from a strongly constrained to a weakly constrained MGCFA model $\mathcal{M}^{\text{weak}}$. Thus, the intercepts are free to vary across groups in the baseline model. The metric BV method still removes items entirely, but the removal of an item will not result in an increase in CFI due to a violation of invariance of the intercepts.

4.4.4 Rieger (R1 & R2)

For the R approaches, the necessary changes aren't quite as obvious. It may be tempting to simply choose the correlation component as relating to loadings, but it is clear from panel B of Figure 3 that non-zero correlation can result under scalar invariance if the true latent means vary across groups. Instead, the following Theorem can provide further insight. To omit some indices for the sake of clarity, we again consider a single item Y as in Theorem 4.2.

Theorem 4.3. *Iff metric MI is satisfied,*

$$\omega^l = \omega^k \tag{85}$$

for all pairs $l, k \in \{1, ..., g\}$.

Proof. First, recall that ω^l, the regression slope in the secondary regressions of item Y's residuals on the latent variable in group l, is given by

$$\omega^l = \frac{\text{Cov}\left(r^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)}. \tag{86}$$

Further note that in the four relevant DGPs, metric MI is satisfied in both the case of a pure scalar MI violation and in the case of perfect MI, shown in (52) and (53), respectively. We thus need to show that the statement of the theorem holds in these cases. For the case of perfect MI, the statement is true by Theorem 4.2 which shows that the covariance in the numerator of equation (86) is equal to zero and thus also

$$\omega^l = 0 \; \forall \; l. \tag{87}$$

For the scalar MI violation, we have

$$\begin{aligned} \frac{\text{Cov}\left(r^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} &\overset{(52)}{=} \frac{\text{Cov}\left(\tau^l + \lambda\eta^l + \varepsilon - \alpha - \beta\eta^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= (\lambda - \beta)\frac{\text{Cov}\left(\eta^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= (\lambda - \beta). \end{aligned} \tag{88}$$

which no longer depends on l.

Next, we need to show that the statement doesn't hold if metric MI is violated, i.e. for the remaining two cases in equations 50 and 51. For the former, we have

$$\begin{aligned} \frac{\text{Cov}\left(r^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} &\overset{(50)}{=} \frac{\text{Cov}\left(\tau^l + \lambda^l\eta^l + \varepsilon - \alpha - \beta\eta^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= (\lambda^l - \beta)\frac{\text{Cov}\left(\eta^l, \eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= \lambda^l - \beta \end{aligned} \tag{89}$$

and for the latter, we have the same result by

$$\begin{aligned}\frac{\text{Cov}\left(r^l,\eta^l\right)}{\text{Var}\left(\eta^l\right)} &\overset{(51)}{=} \frac{\text{Cov}\left(\tau+\lambda^l\eta^l+\varepsilon-\alpha-\beta\eta^l,\eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= (\lambda^l-\beta)\frac{\text{Cov}\left(\eta^l,\eta^l\right)}{\text{Var}\left(\eta^l\right)} \\ &= \lambda^l-\beta.\end{aligned} \tag{90}$$

In a similar argument to the one made in Theorem 4.2,

$$\lambda^l-\beta=\lambda^k-\beta\;\forall\; l,k \iff \lambda^1=...=\lambda^g, \tag{91}$$

which contradicts the presence of a violation of metric invariance. □

In other words, Theorem 4.3 shows that in the secondary regressions of $\hat{r}_i^l$ on η^l, groups have parallel regression lines iff metric MI is satisfied. An example of this can also be seen in panel B of Figure 3.

Thus, for a metric version of the R approach, what is of relevance is that the regression coefficients differ across groups. This can be tested by adding an interaction term to the regression of the residuals on the latent variable estimates. More specifically, for the metric version, the residuals of all groups are regressed on the latent variable estimates, the group membership and the interaction of those two covariates. If the interaction term explains a significant share of the variance in the residuals, this must be due to differences in the group-specific regression coefficients of the latent variable. Formally, this test is implemented as an F-test of the group membership variable. Since the step-wise R2 method internally builds on R1, a metric version can be obtained by simply using the metric R1 version for each step.