Factorial Invariance:

Historical Trends and New Developments.

Roger E. Millsap
Arizona State University

William Meredith
University of California, Berkeley

Paper presented at the "Factor Analysis at 100" Conference, May 13-15, 2004, L.L. Thurstone Psychometric Laboratory, University of North Carolina.

Overview

(1) The Invariance Problem: What is factorial invariance?

(2) History of the Problem

a) Selection theory and invariance

b) Rotational approaches

c) Confirmatory factor analytic approaches

(3) New Developments and Future Directions

a) The Meaningfulness Problem: When does a violation of invariance matter? What should we do about it?

b) The Specification Search Problem: Which parameters are responsible for the violation of invariance?

(4) Final Points

The Factorial Invariance Problem

X = px1 vector-valued observed variable

Multiple populations indexed by k=1,...,K

Common factor representation in the kth population:

$$X_k = \tau_k + \Lambda_k \xi_k + \delta_k$$

$$\mu_{xk} = \tau_k + \Lambda_k \kappa_k \qquad \Sigma_k = \Lambda_k \phi_k \Lambda_k' + \Theta_k$$

Λ_k is pxr pattern matrix, τ_k is px1 latent intercept

κ_k, ϕ_k are mean vector, covariance matrix for ξ_k

Θ_k is covariance matrix for δ_k

Question:

Are the parameters $[\tau_k, \Lambda_k, \Theta_k]$ invariant over k?

History of the Invariance Problem

<u>Selection theory and Invariance</u>:

1) Aitken (1934), building on Pearson (1902), described how the alteration of a portion of a covariance matrix through direct selection would affect the rest of the matrix indirectly, assuming multivariate normality.

2) Lawley (1943) showed that Aitken's results hold under weaker conditions: linearity and homoscedasticity of regressions.

3) Thomson & Ledermann (1939) used the Aitken results to show the implications of selection for the factor structure in the selected groups. The key here is that direct selection operates on a subset of the **measured variables.**

Thomson (1939) concluded:

"All these considerations make it very doubtful indeed whether any factors, and any loadings of factors, have absolute meaning. They appear to be entirely dependent upon the population in which they are measured..."

Thurstone and Simple Structure

1) Thurstone (1947) emphasized "simple structure" as an important consideration in identifying factors with primary or important processes.

Simple Structure ==> each measured variable has at least one zero loading

2) Thurstone (1947) made two general claims regarding the impact of selection on factor structure:

A) Under univariate selection, simple structure in loadings is preserved, although factor correlations may vary as a function of selection.

B) Under multivariate selection, simple structure is also preserved for the <u>primary factors</u>, but selection can introduce additional incidental factors that depend on selection. Incidental factors will not hold up across different selections.

"The analysis of these various cases of selection is very encouraging, in that a simple structure has been shown to be invariant under widely different selective conditions." (Thurstone, 1947)

Rotational Approaches to Invariance

From general selection principles, attention shifted to ways of finding an invariant factor structure, if it exists.

1) **Ahmavaraa (1954)** derived an expression for the factor pattern matrix after selection. He confirmed Thurstone's claim regarding simple structure invariance under certain conditions.

2) **Meredith (1964a,b)** showed that under Lawley's selection theorem, we have invariance in the factor pattern matrix provided that

A) measures are expressed in common units across populations,

B) the factor solution is not required to be orthogonal in all populations,

C) **the regressions of the common factors on the selection variables are linear and homoscedastic.**

Here the "selection variables" are external and need not be measured or even known.

How can we find the invariant pattern from separate, within-population, exploratory factor analyses?

(1) Meredith (1964b) provided two rotational methods for taking separate factor solutions from each population and finding an invariant pattern matrix that is "best-fitting."

(2) Cattell (1944) had earlier developed the rotational principle of "parallel proportional profiles" which argued for rotating factor solutions so that the pattern matrices are columnwise proportional. See McArdle & Cattell (1994).

--Applied to solutions from two populations, the principle leads to a method for finding proportional solutions that are also orthogonal in each population.

--The principle is limited to two populations, and the observed measures must not be standardized within-population.

(3) Procrustes rotational methods can be applied to the multiple-population case to seek an invariant pattern. Especially useful are methods that require only partially-specified target patterns ("hyperplane fitting"). Contributors: **Mosier, Green, Horst, Browne,Tucker,**

Schönemann, Lawley, Maxwell, Jöreskog, Meredith
Confirmatory Factor Analytic Approaches

(1) Jöreskog (1971) first presented a confirmatory factor analytic approach to studying factorial invariance in multiple populations. The method permitted a direct test of fit for an invariant factor pattern, and an estimation algorithm for finding this best-fitting pattern.

--other aspects of the factor structure could be evaluated for invariance (e.g., unique factor variances).

(2) Sörbom (1974) extended Jöreskog's approach to include mean structures, adding latent intercept parameters as an essential part of the invariance question.

(3) Muthén & Christoffersson (1981) extended CFA to include dichotomous measures in multiple populations, with tests of invariance on model parameters. Further extensions to the polytomous case have been considered by both Jöreskog and Muthén.

(4) Robust inference in multiple-population CFA has been developed by **Satorra (1993, 2000)** and by **Bentler, Lee, & Weng (1987)**.

New Developments and Future Directions

Much technical progress has been made ....

--we can test any invariance hypothesis using a mixture of global and local fit statistics in the normal case.

--we can examine hypotheses on both mean and covariance structures.

--we can fit and test models for both continuous and discrete observed measures.

--in the non-normal case, we can use robust fit statistics or bootstrap methods to examine hypotheses.

Some technical problems remain (e.g., fit evaluation with modest samples and discrete measures).

We will look at two areas in which new developments are needed, one being concerned with <u>interpretation</u> and the other being <u>technical</u>.

The Meaningfulness Problem

When is a violation of invariance large enough to warrant concern? What should be done about it?

The invariance literature is almost exclusively devoted to procedures for <u>detecting</u> violations of invariance.

Consider the “ideal” level of invariance: **strict factorial invariance.**

$$\mu_{xk} = \tau + \Lambda\kappa_k \qquad\qquad \Sigma_k = \Lambda\phi_k\Lambda^{\prime} + \Theta$$

Here loadings, intercepts, and unique variances are invariant; systematic group differences in observed means and covariance matrices are due to group differences in common factor score distributions.

In practice, strict factorial invariance is seldom found to hold, given enough observed measures and sufficiently large samples.

A more realistic finding is **partial invariance: some, but <u>not all</u>, elements of $[\tau_k, \Lambda_k, \Theta_k]$ are found invariant.**

Example: Suppose that we have p=10 observed measures, with 6 of the 10 measures having invariant loadings. What is the implication of partial invariance for use of the scale formed by the 10 measures? The literature provides little guidance here.

(1) "Go ahead and use the full 10-measure scale because the majority of the measures are invariant."

--this option ignores the <u>magnitudes</u> of the violations of invariance.

(2) "Go ahead and use all measures as long as none of them show loading differences in excess of ____."

--this option uses arbitrary standards for deciding when a difference is "too large".

(3) "Drop any measures that aren't invariant, and use the remaining measures."

--this option results in as many versions of the scale as there are invariance studies.

(4) "Don't use the scale!"

--this option leads to paralysis, or early retirement.

Solution: Consider whether the violations of invariance interfere with the intended use of the scale.

For the case in which the scale will be used explicitly or implicitly for selection, Millsap & Kwok (2004) give an approach for deciding when the violations of invariance lead to inaccurate selection in one or more groups.

Assuming measures fit a single-factor model, and selection is to be based on composite of measures:

Step 1: Arrive at fitted model for all groups with partial invariance.

Step 2: Use fitted model to generate hypothetical bivariate distribution of factor scores and composite of measured variables, pooled across groups.

Step 3: Designate cutpoints on composite score and factor score distributions for selection.

Step 4: Calculate sensitivity, specificity, hit rate for each group, and compare to strict invariance model.

Step 5: Base decision on above accuracy indices.

Example: Single-factor with p=6 measures.

$$\tau_1 = \begin{vmatrix} .4 \\ .1 \\ .5 \\ .2 \\ .1 \\ .4 \end{vmatrix} \qquad \tau_2 = \begin{vmatrix} .4 \\ .1 \\ .5 \\ .4 \\ .3 \\ .8 \end{vmatrix} \qquad \lambda_1 = \begin{vmatrix} .6 \\ .5 \\ .3 \\ .4 \\ .2 \\ .4 \end{vmatrix} \qquad \lambda_2 = \begin{vmatrix} .6 \\ .5 \\ .3 \\ .5 \\ .4 \\ .5 \end{vmatrix}$$

Invariant $\boldsymbol{\Theta}$: diag $\boldsymbol{\Theta} = \begin{vmatrix} .3 & .3 & .4 & .2 & .3 & .2 \end{vmatrix}$

$$\kappa_1 = 0, \quad \phi_1 = 1, \qquad \kappa_2 = .5, \quad \phi_2 = 1$$

Let $x_k = 1' X_k$ be the unweighted sum of the measures.

$$\mu_{x1} = 1.7, \quad \sigma^2_{x1} = 7.46, \qquad \mu_{x2} = 3.9, \quad \sigma^2_{x2} = 9.54$$

Results: **Selection at 90th percentile in pooled group.**

	Partial Inv			Strict Inv		
Group	**Sens.**	**Spec.**	**PPV**	**Sens.**	**Spec.**	**PPV.**
G1	**.40**	**.99**	**.77**	**.68**	**.98**	**.64**
G2	**.80**	**.93**	**.67**	**.72**	**.96**	**.74**

Sens. = Sensitivity, or proportion of true positives that are selected.

Spec. = Specificity, or proportion of true negatives that are rejected.

PPV = positive predictive value, or proportion of those selected that are true positives.

Biggest impact of violation of invariance lies in the reduced sensitivity within Group 1: qualified individuals will be overlooked to a larger degree due

to violation of invariance in this group.

The Specification Search Problem

If strict invariance is violated, how can we accurately locate which parameters are responsible?

Typically, invariance is studied via a nested sequence of models:

(1) **Configural invariance** (Thurstone, 1947): zero elements of pattern matrices in the same locations for all groups.

(2) **Metric or pattern invariance** (Thurstone, 1947): pattern matrices are fully invariant.

(3) **Strong factorial invariance** (Meredith, 1993): pattern matrices and latent intercepts are fully invariant.

(4) **Strict factorial invariance** (Meredith, 1993): pattern matrices, intercepts, and unique variances are fully invariant.

If any of the models (1)-(4) are rejected, we then seek which parameters led to the rejection. The search requires some exploratory model-fitting, or respecification.

Two problems arise with these searches:

(1) **We know that data-driven respecifications are likely to mislead, especially if many modifications are needed (MacCallum, 1986).**

(2) **Some invariance constraints are needed for identification, but these constraints may disrupt the search if poorly chosen (Cheung & Rensvold, 1999).**

For (1) above, there is remarkably little evidence in the literature regarding the performance of specification searches in multiple-group invariance problems.

One strategy for such a search, following rejection of full invariance for loadings, would use modification indices or LM statistics to sequentially relax invariance constraints on loadings until adequate fit is achieved.

Can this really be done with reasonable accuracy, finding the measures that violate invariance while avoiding making erroneous claims of violations?

The evidence is mixed....

Simulation Example:

---Six measured variables

---Four measures have invariant loadings, and two have loadings that are both smaller in one group.

---Communalities range from .22 to .48

---Multivariate normal data, N=500 per group

---Three loading difference magnitudes: .3, .2, .1

Loading Diff	% Finding Both Violations	% False Positive
Small	11	25
Medium	66	24
Large	98	16

(Source: Myeongsun Yoon, Master's Thesis)

The second problem noted earlier concerns the need to constrain some parameters to invariance for identification purposes.

Example

Single-factor model: Typical practice is to choose one measured variable to serve as a marker variable for scaling, setting its loading to one in all groups, and its latent intercept to zero in all groups.

What if the chosen marker variable violates invariance in loadings and/or intercepts?

---Leads to distortions in estimates for loadings and/or intercepts of other measures, and may disrupt specification search.

This problem can be reduced if we have prior information (e.g., previous studies) regarding which measures might have invariant parameters. If no such information is available, we must rely on alternative strategies.

Proposed Strategies:

1) Rensvold & Cheung (2001; Cheung & Rensvold, 1999): The factor ratio test.

(1) Conduct an exhaustive search procedure shifting the location of the marker across all measures, and pairing the marker with one additional measure that is also constrained to invariance.

(2) Use bonferroni-adjusted critical values for exact chi-square difference tests, testing in each case whether invariance constraint for additional measure leads to significant loss of fit.

(3) The resulting pattern of significant and non-significant results should reveal which subset of measures can be considered to have invariant loadings

This procedure can lead to many tests if the number of measures is large; **N measures = N(N-1)/2 tests**

Relies on chi-square difference test even where initial fit is poor by chi-square.

2) Meredith & Horn (2001): addresses intercept/factor mean identification problem

(1) Conduct invariance study without means. Proceed only if factor loadings are invariant.

--Note: **Do not require perfect simple structure for loadings prior to invariance investigation.**

(2) Test for invariance of intercepts in mean structure model. If that fails, get estimates of factor means by fixing $\mathbf{\Lambda}, \mathbf{\Theta}_k$ to estimated values from (1), with intercepts constrained to invariance. Scale factor means to sum to zero across groups.

(3) Rerun model fixing $\mathbf{\Lambda}, \mathbf{\Theta}_k$, along with $\boldsymbol{\kappa}_k$ from (2), to get estimates of intercepts.

This method is a simplification of a slightly more complex method outlined in Meredith and Horn (2001). Step (2) forces the factor means to carry as much of the group difference in observed means as possible. In Step (3), the intercepts capture what remains of this group difference.

(3) Alternative identification strategy: Full loading and intercept invariance, with factor variance(s) and factor mean in <u>one group</u> set to 1.0 and 0 respectively.

This strategy avoids picking a marker variable explicitly, either for loadings or for intercepts.

(1) Test pattern invariance using the above identification for pattern elements.

(2) If pattern invariance is rejected, begin specification search by sequentially relaxing the invariance constraints on loadings based on modification indices or LM statistics, starting with the constraint that yields the highest index or LM statistic.

(3) Continue relaxing the constraints until either fit is adequate or only one constraint remains.

(4) Repeat these steps with intercepts, confining interest to intercepts for measures that have invariant loadings.

This strategy faces the problems faced by specification searches generally: it may be difficult to accurately determine which measures violate invariance and which do not.

Final Points

(1) Factorial invariance supports the scientific meaningfulness of the factor structure.

(2) Great technical progress has been made in methods for evaluating invariance; progress has been slower on issues of interpretation.

(3) Some likely directions for future research will include:

--technical problems with small samples and ordinal measures in multiple populations;

--latent mixture models and invariance: defining populations as latent classes;

--the impact of violations of invariance on selection, prediction, or decisions;

--explanations for violations of invariance: do violations always stem from additional factors?