Unifying and extending methods for measurement invariance using Bayesian regularization

March 17, 2017, SEM Working Group

Sara van Erp\textsuperscript{1} Joris Mulder\textsuperscript{1} Daniel L. Oberski\textsuperscript{2}

\textsuperscript{1}Tilburg University
\textsuperscript{2}Utrecht University
The problem: Measurement invariance
Measurement invariance

MI is an important concept when comparing multiple groups. Model restrictions are needed to ensure that the latent construct is measured similarly across groups.
MI is an important concept when comparing multiple groups. Model restrictions are needed to ensure that the latent construct is measured similarly across groups.

**The problem**
So essentially, MI poses an *identification* problem.
Multiple group factor model

\[ y_{ijg} = \nu_{jg} + \lambda_{jg} \eta_{ig} + \epsilon_{ijg} \]

with \( \eta_{ig} \sim N(\alpha_g, \omega_g^2) \)
and \( \epsilon_{ijg} \sim N(0, \sigma_{jg}^2) \)

For \( i = 1, \ldots, n_g \) individuals, \( j = 1, \ldots, J \) items, and \( g = 1, \ldots, G \) groups.
Popular approaches

Some common methods to assess MI:

- Configural, metric, & scalar invariance
- Partial invariance
- Multiple group factor analysis alignment [Asparouhov and Muthén, 2014]
- Bayesian approximate measurement invariance [Muthén and Asparouhov, 2013]
Popular approaches

Some common methods to assess MI:

- Configural, metric, & scalar invariance
- Partial invariance
- Multiple group factor analysis alignment  
  [Asparouhov and Muthén, 2014]
- Bayesian approximate measurement invariance  
  [Muthén and Asparouhov, 2013]

Limitations

Each method imposes very specific restrictions, which might be unrealistic in practice.
MI is an identification problem.

Its solution should allow important deviations from MI to be automatically modelled, while small deviations are ignored.

Sounds familiar . . . ?
MI is an identification problem.

Its solution should allow important deviations from MI to be automatically modelled, while small deviations are ignored.

Sounds familiar . . . ?

**Sparse regression!**
Bayesian regularization as a general solution (?)
When $p > N$, penalties are used to identify the model, e.g. $L_1$ or $L_2$. 
When $p > N$, penalties are used to identify the model, e.g. $L_1$ or $L_2$.

Estimates under penalties correspond to posterior mode estimates under specific priors.
When $p > N$, penalties are used to identify the model, e.g. $L_1$ or $L_2$.

Estimates under penalties correspond to posterior mode estimates under specific priors.

**Advantages Bayesian regularization**

- Hyperpriors penalty parameters
- Standard errors
- Non-convex penalties
• RegSEM [Jacobucci et al., 2016]
• Bayesian regularized quantile & semiparametric SEM [Feng et al., 2015, Feng et al., 2017, Guo et al., 2012]
• Lasso for DIF [Magis et al., 2014]

Bayesian regularization has not yet been applied in the MGCFA model.
Multiple group factor model:

\[ y_{ijg} = \nu_{jg} + \lambda_{jg} \eta_{ig} + \epsilon_{ijg} \]

with \( \eta_{ig} \sim N(\alpha_g, \omega_g^2) \)

and \( \epsilon_{ijg} \sim N(0, \sigma_{jg}^2) \)
Parametrization

Multiple group factor model:

\[ y_{ijg} = \nu_{jg} + \lambda_{jg} \eta_{ig} + \epsilon_{ijg} \quad \text{with} \quad \eta_{ig} \sim N(\alpha_g, \omega^2_g) \]
\[ \text{and} \quad \epsilon_{ijg} \sim N(0, \sigma^2_{jg}) \]

We specify shrinkage priors for \( \delta^\nu_{jg} \) and \( \delta^\lambda_{jg} \):

\[ \delta^\nu_{jg} = \nu_{jg} - \mu^\nu_j \]
\[ \delta^\lambda_{jg} = \lambda_{jg} - \mu^\lambda_j \]

So that the model becomes:

\[ y_{ijg} = (\delta^\nu_{jg} + \mu^\nu_j) + (\delta^\lambda_{jg} + \mu^\lambda_j) \eta_{ig} + \epsilon_{ijg} \quad \text{with} \quad \eta_{ig} \sim N(\alpha_g, \omega^2_g) \]
\[ \text{and} \quad \epsilon_{ijg} \sim N(0, \sigma^2_{jg}) \]
Priors

Deviances loadings:

\[ \delta_{jg}^{\lambda} \sim N(0, 0.01) \]

So the loadings are assumed to be approximately invariant.
Priors

Deviances loadings:

\[ \delta_{jg}^\lambda \sim N(0, 0.01) \]

So the loadings are assumed to be approximately invariant.

Nuisance parameters:

\[ \alpha_g \sim C(0, 2.5) \]
\[ \sigma_{jg} \sim C^+(0, 5) \]
\[ \omega_g \sim C^+(0, 5) \]
1. Approximate MI

$$\delta_{jg}^\nu \sim N(0, 0.01)$$
1. Approximate MI

\[ \delta_{jg} \sim N(0, 0.01) \]

2. Ridge

\[ \delta_{jg} \sim N(0, \tau_j^2) \]
\[ \tau_j^2 \sim IG(0.5, 0.5) \]
1. Approximate MI

\[ \delta_{jg}^\nu \sim N(0, 0.01) \]

2. Ridge

\[ \delta_{jg}^\nu \sim N(0, \tau_j^2) \]
\[ \tau_j^2 \sim IG(0.5, 0.5) \]

3. Lasso

[Park and Casella, 2008]

\[ \delta_{jg}^\nu \sim DE(0, \frac{\sigma_{jg}^2}{\tau_j}) \]
\[ \tau_j \sim G(1, 1.78) \]
1. Approximate MI

\[ \delta_{jg}^\nu \sim N(0, 0.01) \]

4. Normal mixture

\[ \delta_{jg}^\nu \sim \gamma_j \times N(0, 1) + (1 - \gamma_j) \times N(0, 0.001) \]
\[ \gamma_j \sim U(0, 1) \]

2. Ridge

\[ \delta_{jg}^\nu \sim N(0, \tau_j^2) \]
\[ \tau_j^2 \sim IG(0.5, 0.5) \]

3. Lasso

[Park and Casella, 2008]

\[ \delta_{jg}^\nu \sim DE(0, \frac{\sigma_{jg}^2}{\tau_j}) \]
\[ \tau_j \sim G(1, 1.78) \]
1. Approximate MI
\[
\delta_{jg}^{\nu} \sim N(0, 0.01)
\]

2. Ridge
\[
\begin{align*}
\delta_{jg}^{\nu} & \sim N(0, \tau_j^2) \\
\tau_j^2 & \sim IG(0.5, 0.5)
\end{align*}
\]

3. Lasso
[Park and Casella, 2008]
\[
\begin{align*}
\delta_{jg}^{\nu} & \sim DE(0, \frac{\sigma_{jg}^2}{\tau_j}) \\
\tau_j & \sim G(1, 1.78)
\end{align*}
\]

4. Normal mixture
\[
\begin{align*}
\delta_{jg}^{\nu} & \sim \gamma_j \times N(0, 1) + (1 - \gamma_j) \times N(0, 0.001) \\
\gamma_j & \sim U(0, 1)
\end{align*}
\]

5. Robust mixture
\[
\begin{align*}
\delta_{jg}^{\nu} & \sim \gamma_j \times C(0, 1) + (1 - \gamma_j) \times N(0, 0.001) \\
\gamma_j & \sim U(0, 1)
\end{align*}
\]
Plot priors

Approximate MI  Lasso  Normal mixture  Ridge  Robust mixture
Plot priors

Approximate MI
Lasso
Ridge
Normal mixture
Robust mixture
Illustrations
Simulated data

Generated data from a one-factor model with 4 items. Two groups with 500 observations per group.

\[
\nu = \begin{bmatrix}
0 & 0 \\
0 & 0 \\
0.5 & 0.5 \\
0.5 & -0.5
\end{bmatrix}
\]

\[
\Lambda = \begin{bmatrix}
0.5 & 0.5 \\
0.5 & 0.5 \\
0.7 & 0.7 \\
0.9 & 0.9
\end{bmatrix}
\]

\[\alpha_1 = 0\]

\[\alpha_2 = 0, 0.3, 0.5, 1, 3, 5, 10\]

\[\omega^2_g = 1\]

Constant error for each dataset: \(\epsilon \sim N(0, 1)\)
Bias factor mean

![Graph showing bias against factor mean for different methods: Scalar MI, Approx. MI, Ridge, Normal mixture, Robust mixture, and Lasso. The graph plots bias on the y-axis and factor mean on the x-axis.]
Bias factor mean for small differences

![Graph showing bias factor mean for small differences. The graph illustrates the relationship between factor mean and bias across different methods: Scalar MI, Approx. MI, Ridge, Normal mixture, Robust mixture, and Lasso.]
Applying the priors to the Holzinger & Swineford (1939) data.

Mental ability scores on 9 variables from children in two schools, modelled with 3 latent factors.
Applying the priors to the Holzinger & Swineford (1939) data.

Mental ability scores on 9 variables from children in two schools, modelled with 3 latent factors.

Some notes:

- The factor covariance matrix was decomposed into scales and a correlation matrix with a $C^+(0, 2.5)$ prior for the scales and an LKJ prior on the correlation matrix with shape $= 2$.
- The first loading in each group was fixed to 1 for identification.
- The same number of observations per school was selected ($n_g = 145$).
Empirical example: Posterior densities $\delta^\nu$

**Approximate MI**

- $\text{diff}_n[2,2,2]$
- Values range from -0.2 to 0.2

**Ridge**

- $\text{diff}_n[2,2,2]$
- Values range from -2 to 2

**Lasso**

- $\text{diff}_n[2,2,2]$
- Values range from -75000 to 0

**Normal mixture**

- $\text{diff}_n[2,2,2]$
- Values range from -0.2 to 0.2

**Robust mixture**

- $\text{diff}_n[2,2,2]$
- Values range from -1.00 to 0.25
Empirical example: Posterior densities $\delta^\nu$

Approximate MI

Ridge

Lasso

Normal mixture

Robust mixture
Empirical example: Factor means group 2

Mean factor 1

95% credible interval

Scalar MI | Approx. MI | Normal mixture | Robust mixture | Lasso | Ridge

Mean factor 2

95% credible interval

Scalar MI | Approx. MI | Normal mixture | Robust mixture | Lasso | Ridge

Mean factor 3

95% credible interval

Scalar MI | Approx. MI | Normal mixture | Robust mixture | Lasso | Ridge
Conclusion
Mixture priors showed the smallest bias for small effects, but can result in multimodal posteriors.
Mixture priors showed the smallest bias for small effects, but can result in multimodal posteriors.

Regularization priors can, in theory, be useful in the context of MI.
Mixture priors showed the smallest bias for small effects, but can result in multimodal posteriors.

Regularization priors can, in theory, be useful in the context of MI.

There is, however, more work to be done before we know whether they are useful in practice.
Work to be done

- Vary number of groups/items/observations
- Different priors (e.g. Horseshoe, Subbotin)
- Regularization priors on deviances loadings and intercepts
- Weights for different sample sizes per group
- Vary prior settings (e.g. variance mixture prior)
- Different parametrizations
- Compare posterior means and modes
- ...
Thank you!

s.vanerp_1@tilburguniversity.edu
Twitter: @SaravanErp


