Boris Sokolov
LCSR, Higher School of Economics

LCSR-EBES training methodological workshop
04-05.04.2015

Bayesian Analysis for Comparative Social Research

Overview

- Part I: *Introduction to Bayesian data analysis*
  - General principles of Bayesian analysis
  - Prior distributions
  - MCMC estimation
  - Bayesian model selection

- Part II: *Bayesian structural equation modeling in MPLUS*
  - Overview of Bayesian Features In Mplus
  - Example I: Basic Example of Bayesian CFA
  - Example II: Using of Bayesian approach to detect CFA model misspecifications
  - Example III: Approximate (Bayesian) measurement invariance

Part I: *Introduction to Bayesian data analysis*

Bayesian Analysis for Comparative Social Research

Historical Background

- Bayesian statistics is named after Thomas Revenge Bayes, an English statistician, philosopher and Presbyterian minister of the XVIII century
- Some contemporary authors, however, argue that Pierre-Simon Laplace's contribution in what is now called Bayesian statistics is much larger and therefore justify the use of term "Laplacian" instead "Bayesian"

Why Bayes?

- Bayesian analysis (BA) allows for incorporating of existing information about the model parameters of interest
- BA directly includes uncertainty about parameters in the model, yielding more realistic predictions
- BA provides more narrow (and credible) probability intervals
- BA parameter estimates are unbiased with respect to sample size: particularly, It provides reliable parameters estimates even with small sample size (both on individual and contextual levels)
- BA allows for comparing models with different methods
- BA allows for more complicated models (which usual frequentist and likelihood models unable to estimate)
- BA has a theoretic guarantee of convergence

Some notes on the interpretation of probability

- Physical: e.g. relative frequency of the event in a long run of trials (rolling dice of tossing coin)
- Epistemological: reflect degree of our subjective belief in the probability of event (e.g. readiness to bet on, or against the event)

Bayes' Theorem

- Let's consider probabilities of two events, A and B
- B is an evidence
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- $P(A)$ – the prior, is the initial degree of belief in A
- $P(B|A)$ – likelihood of B given A
- $P(A|B)$ – the posterior, is the degree of belief having accounted for B
- $P(B|A)/P(B)$- the support B provides for A

Bayesian Statistical Analysis

- Let θ be a set of parameters of model M and $data$ is observed sample
- $\Pr(\theta \mid data, M) = \frac{\Pr(\theta|M) \times \Pr(data \mid \theta, M)}{\Pr(data \mid M)}$
- $\Pr(\theta| M)$- prior
- $\Pr(data \mid \theta, M)$– likelihood of the data given parameters$(p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$, where y_i is an individual data point)
- $\Pr(data \mid M)$ - marginal likelihood (evidence)
- $\Pr(\theta| data, M)$ - posterior

- *Posterior is proportional to prior times likelihood*
- $\Pr(\theta| data, M) \propto \Pr(\theta| M) \times \Pr(data \mid \theta, M)$

Frequentist vs. Bayesian

- Frequentist view: parameters are fixed. ML or LS estimates have an asymptotically-normal distribution

- Bayesian view: parameters are variables that have a prior distribution. Estimates have a possibly non-normal posterior distribution. Does not depend on large-sample theory

Priors. What is it?

- $\Pr(\theta \mid M)$ – a prior distribution, which reflects the information about the model parameters before observing data .
- For a given parameter, θ_t, a prior probability is a description of what is known *a priori* about the parameter to be estimated
- For specifying a prior, you should choose a type of distribution (e.g., normal, beta, uniform) and give the values of its parameters (for instance, mean and variance for normal distribution).

Classes of prior distributions:

- Informative vs. Non-Informative
- Proper vs. Improper
- Conjugate priors
- Hyperpriors

Proper vs. Improper Priors

- Let θ is a set of model parameters
- A prior $p(\theta)$ is said to be improper if $\int p(\theta)d\theta = \infty$
- For example, Beta ($\alpha = 0, \beta = 0$), or uniform distribution on an infinite interval (entire real line)
- Improper priors can cause improper posteriors and therefore lead to invalid estimates.
- Nonetheless, improper priors may still be useful sometimes since they usually yield non-informative priors (e.g., uniform distribution)

Informative vs. Non-Informative Priors

- An ***informative*** prior is not dominated by the likelihood and has an impact on posterior distribution
- A prior is ***non-informative*** if it has minimal impact on the posterior distribution of a respective parameter (if the prior is 'flat' relative to the likelihood function).
- ***Weakly informative priors*** (WIP) are used to avoid computational problems
- For instance, a normal distribution with very large variance is the WIP for centered and scaled continuous predictors
- Another example of WIP (so called *vague prior*) is a conjugate prior with large scale parameter
- ***Least informative parameters*** (LIP) are used to 'let data the data speak from themselves'.
- A popular subclass of LIP is *flat priors* (unbounded uniform distribution). FP allow the posterior distribution to be affected solely by the data, with no impact from prior information
- ***Jeffreys' prior*** is a least informative prior that is invariant to transformations (reparameterization of the parameter vector)

Conjugate Priors

- A prior said to be a conjugate prior for a family of distributions if the prior and posterior distributions are from the same family (that is, the form of the posterior had the same distributional form as the prior distribution).
- E.g. normal prior distribution / normal posterior distribution, gamma/gamma, gamma/Poisson, beta/beta.
- Using of conjugate priors allows for avoiding computational complexities: instead solving complex integrals you can just slightly change respective parameters to compute posterior.

Warnings!

- Subjectivity of the choice for prior distributions
- You should perform sensitivity analysis (run several models with different priors) to check how various prior distributions affect the results.
- Non-informative priors leads to parameter estimates which are almost identical to maximum likelihood estimations.

Application

- Order-constrained hypotheses
- Computational problems
- Multiple imputation
- Approximate measurement Invariance
- Model misspecifications

To specify prior distribution correctly:

- Use findings from previous studies
- Rely on theoretical considerations
- Conduct exploratory data analysis

MCMC

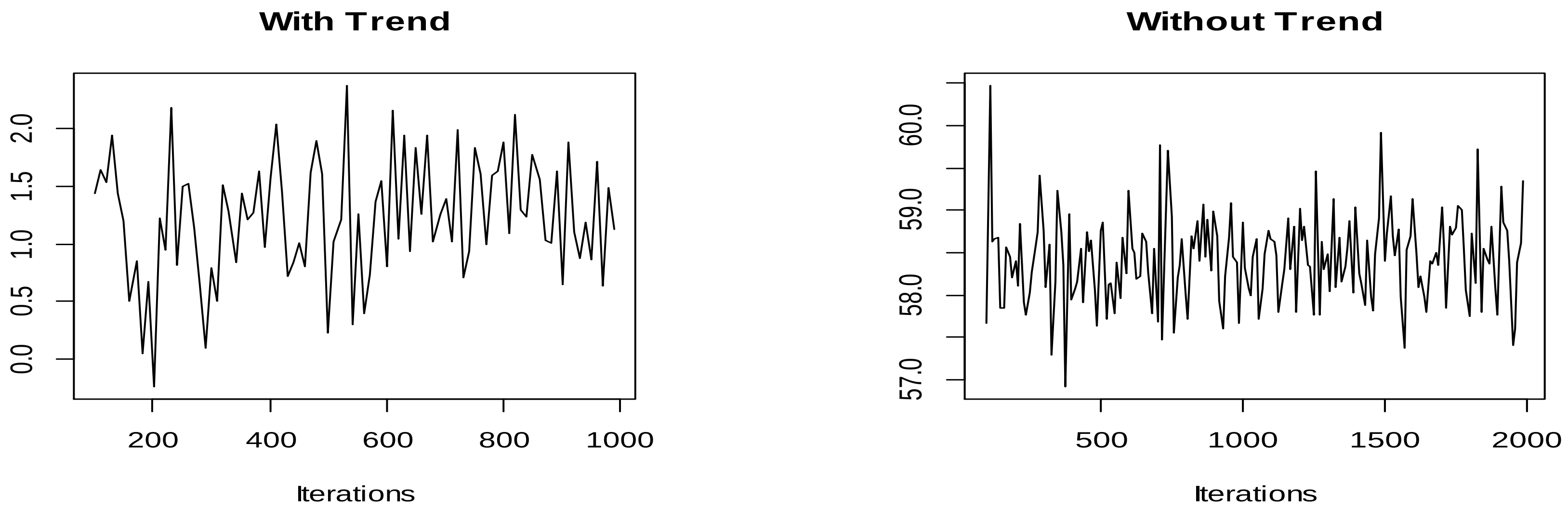
- MCMC is for Monte-Carlo Markov Chains.
- MCMC is used to sample posterior distributions of model parameters
- Mean or mode, and variance of these posterior distributions are Bayesian equivalents of parameter estimates and standard errors in classic statistics
- The quality of the posterior distribution improves as a function of the number of steps in MCMC
- MCMC is well-suited but not the only tool for numerical approximation
- Popular MCMC algorithms: Gibbs, Metropolis-Hastings, Hamiltonian MC, etc.

MCMC Iteration Issues

- Autocorrelation: correlation between consecutive iterations for a parameter. Low correlation desired
- Mixing: the MCMC chain should visit the full range of parameter values, i.e. sample from all areas of the posterior
- Convergence: the MCMC chain should converge to stationary process (should not show an upward or downward trend).
- Burn-in phase: early iterations should be removed due to high autocorrelation.

Convergence Diagnostics

- Visual inspection of **trace plot**: there should be no trend in the chain



- Autocorrelation: **less than 0.1** is considered satisfactory

Potential Scale Reduction Factor (PSR, or Gelman-Rubin diagnostics)

- When several MCMCC iterations carried out in parallel independent chains, PSR consider n iterations in m chains, where $\theta_{ij}$is the value of parameter θ in iteration I of chain j
- $\bar{\theta}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij}$
- $\bar{\theta}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_{ij}$
- $B = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_{.j} - \bar{\theta}_{..})^2$
- $W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_{.j})^2$
- $PSR = \sqrt{\frac{W+B}{W}}$
- PSR **should be slightly higher than 1.** PSR > 1.2 is usually considered to be indicating convergence failure.
- It is also possible to perform Kolmogorov-Smirnoff test of the equality of posterior distributions from different chains.

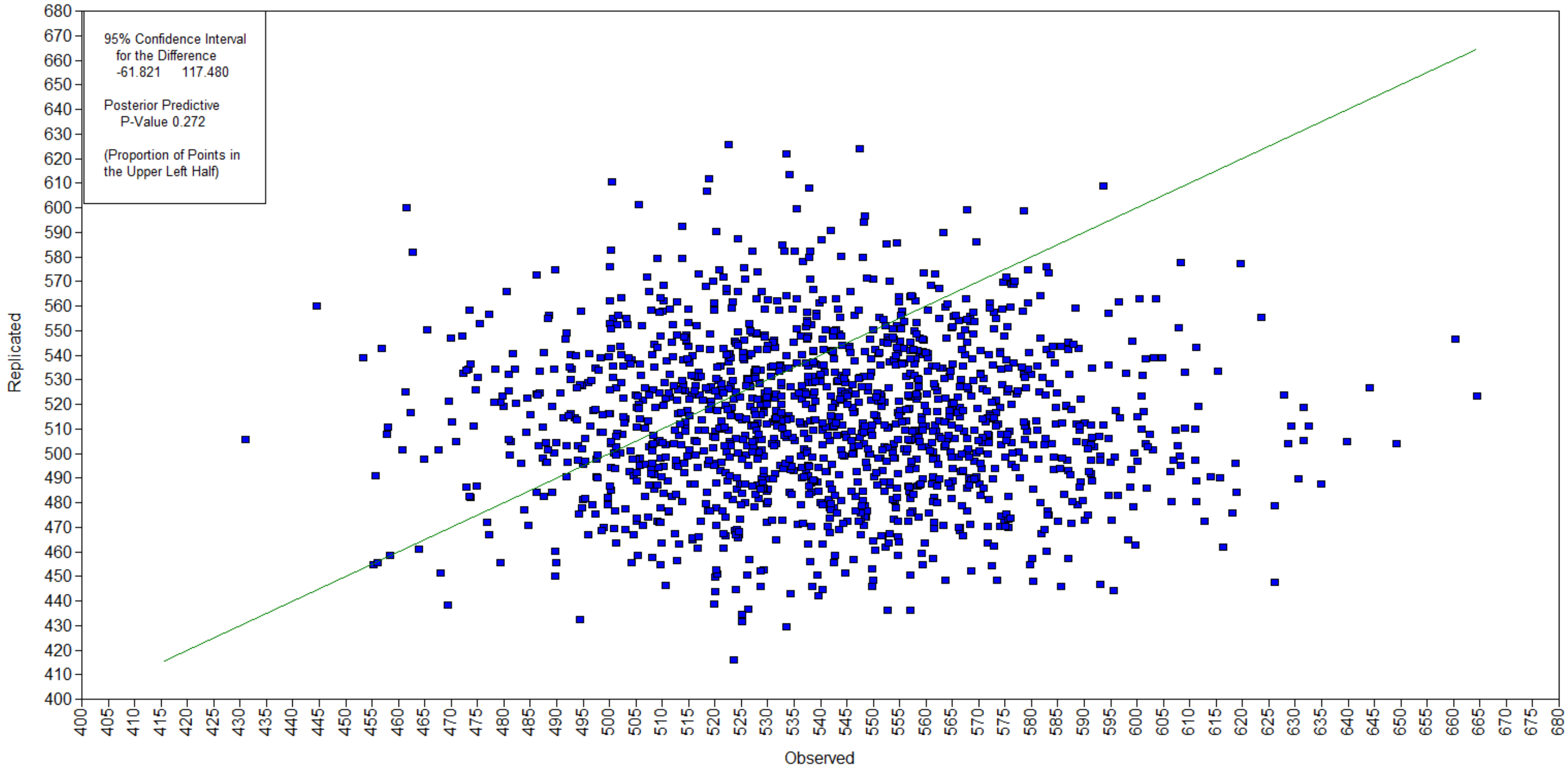
A general framework for MCMC estimation

- Set starting values (often it can be done automatically) and choose appropriate sampler
- Define number of iterations, number of burn-in iterations, and thinning interval.
- Check for autocorrelation and convergence
- Check predictive power of the model.

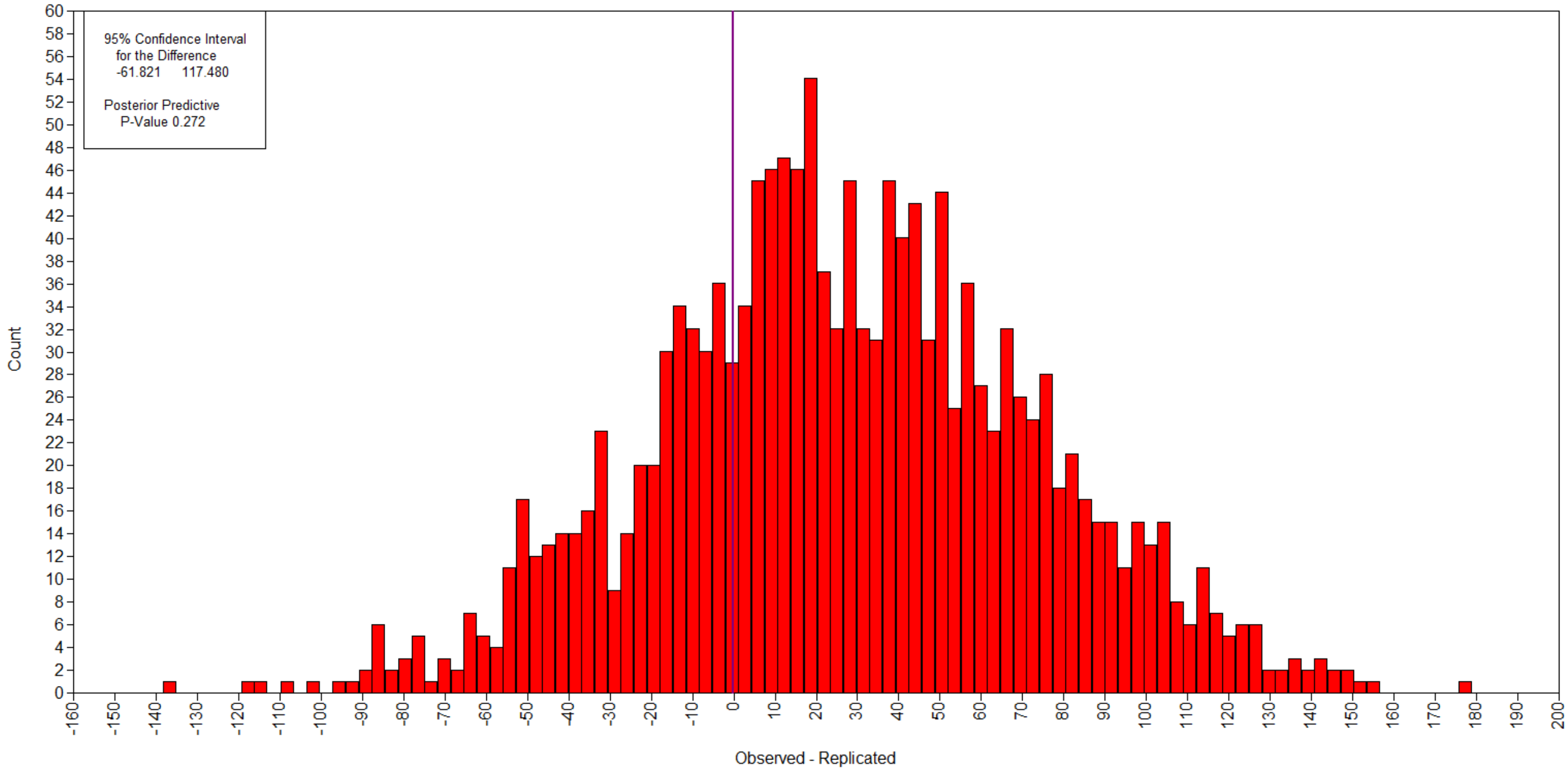
Model selection I
Posterior Predictive Checking

- $PPP$is for posterior predictive P-value
- Consider an arbitrary discrepancy function $f(Y, X, \theta)$ (e.g., chi-square). It is computed at each MCMC iteration t, and then compared to another discrepancy function ,$f(\tilde{Y}, X, \theta)$ ($\tilde{Y}$ is a data set of the same size as Y, but generated at t with the use of current parameter estimates θ_t.
- $PPP = P\left(f(Y, X, \theta) < f(\tilde{Y}, X, \theta)\right) \approx \frac{1}{m}\sum_{t=1}^{m} \delta_t$, where $\delta_t = 1$ when $f(Y, X, \theta) < f(\tilde{Y}, X, \theta)$ and zero otherwise.
- In the ideal universe, PPP should be close to 0.5. In the ideal universe…
- In BSEM context, $PPP < 0.05$ indicates poor model fit.

An example of PPP scatterplot



An example of PPP histogram



Model selection II
Deviance Information Criterion

- $DIC = E^{\theta}[D(\theta)] + pV$, where $mean(D)$ is the mean model-level deviance and pV is the model complexity.
- $D(\theta) = -2\log[p\,(\boldsymbol{y}|\theta)]$ - deviance for
- $pV = var[D(\theta)]/2$
- DIC is similar to BIC: $BIC = E^{\theta}[D(\theta)] + k * log(N)$, where k is the number of model parameters and N is the sample size. But DIC is less sensitive to the sample size and the number of model parameters
- Smaller DIC (BIC) indicates better model
- DIC may be compared across different models and even different methods, as long as the dependent variable remain the same.
- *** DIC is only valid when the posterior distribution is approximately multivariate normal.

Model selection III
Bayes Factor

- Bayes factor is the ratio of the marginal likelihood of the data in the models M_1 and M_2
- $BF = \frac{\Pr(data \mid M_1)}{\Pr(data \mid M_2)} = \frac{\int \Pr(\theta_1 \mid M_1) \Pr(data \mid \theta_1, M_1) d\theta_1}{\int \Pr(\theta_2 \mid M_2) \Pr(data \mid \theta_2, M_2) d\theta_2}$
- $BF = \frac{Exp(-0.5\ BIC_{M1})}{Exp(-0.5\ BIC_{M2})}$
- BF = 2 means that the data favor $M1$ over $M2$ with 2:1 odds
- Values of $BF > 3$ indicate substantial empirical support for $M1$.
- Bayes factor can be used to compare both nested and non-nested models

Part II. Bayesian structural equation modeling in MPLUS

Bayesian Analysis for Comparative Social Research

Why use MPLUS for Bayesian SEM? BUGS/JAGS code for two-factor CFA model

```
model{
 for (i in 1:N){
    #latent 1
    for (t in 1:12){
      y[i,t] ~ dnorm(condmn[i,t], invsig2[t])
      condmn[i,t] <-mu[t] + fload[t]*fscore[i,1]
    }
    #latent 2
    for (t in 13:24){
      y[i,t] ~dnorm(condmn[i,t], invsig2[t])
      condmn[i,t] <-mu[t] + fload[t]*fscore[i,2]
    }
 fscore[i,1:2]~dmnorm(mn.fs[], siginv.fs[,])
 }
 mn.fs[1]<-0
 mn.fs[2]<-0
 siginv.fs <- inverse(sig.fs)
 sig.fs[1,1]<-1
 sig.fs[2,2]<-1
 sig.fs[1,2]<-phi
 sig.fs[2,1]<-phi
 #Prior distribution
 phi ~ dunif(-1,1)
 for (t in 1:24){
   fload[t] ~ dnorm(0,1.0E-3)I(0,)
   invsig2[t] <- 1/psi[t]
   psi[t] ~ dunif(0,400)
   mu[t] ~ dnorm(20,.05)
 }
}
```

Why use MPLUS for Bayesian SEM? LaplacesDemon code for two-factor CFA model

```
Model <- function(parm, Data)
{
### Parameters
alpha <- parm[grep("alpha", Data$parm.names)]
lambda <- parm[grep("lambda", Data$parm.names)]
sigma <- exp(parm[grep("log.sigma", Data$parm.names)])
F <- matrix(parm[grep("F", Data$parm.names)], Data$N, Data$P)
Omega <- as.parm.matrix(Omega, Data$P, parm, Data)
parm[grep("Omega", Data$parm.names)] <- upper.triangle(Omega,
diag=TRUE)
### Log(Prior Densities)
alpha.prior <- sum(dnormv(alpha, 0, 1000, log=TRUE))
lambda.prior <- sum(dnormv(lambda, 0, 1000, log=TRUE))
sigma.prior <- sum(dhalfcauchy(sigma, 25, log=TRUE))
Omega.prior <- dwishart(Omega, Data$N, Data$S, log=TRUE)
F.prior <- sum(dmvnp(F, Data$gamma, Omega, log=TRUE))
### Log-Likelihood
mu <- matrix(alpha, Data$N, Data$M, byrow=TRUE) + F[,Data$f] *
matrix(lambda, Data$N, Data$M, byrow=TRUE)
LL <- sum(dnorm(Data$Y, mu, sigma, log=TRUE))
### Log-Posterior
LP <- LL + alpha.prior + lambda.prior + sigma.prior + F.prior +
Omega.prior
Modelout <- list(LP=LP, Dev=-2*LL, Monitor=LP, yhat=mu, parm=parm)
return(Modelout)
}
```

Why use MPLUS for Bayesian SEM? MPLUS code for two-factor CFA model

```
Model:
F1 by X1 X2 X3;
F2 by X4 X5 X6;
```

Bayesian Estimation In Mplus

Modelling

- Single-level, multilevel, and mixture latent variable models
- Continuous and categorical outcomes
- Multiple Imputation

Estimation

- Default non-informative priors or user-specified informative priors
- Gibbs (default) and Hastings-Metropolis samplers
- Multiple chains using parallel processing
- Posterior parameter distributions with means, medians, modes, and credibility intervals

Diagnostics

- Convergence assessment using Gelman-Rubin potential scale reduction factors (PSR) and Kolmogorov-Smirnoff test (TECH 8)
- Posterior parameter trace plots
- Autocorrelation plots
- Posterior predictive checking plots

Model comparisons

- PPP
- DIC and BIC (for continuous dependent variables)
- Bayes Factor (for continuous dependent variables)
- Informative Hypothesis Testing

MPLUS language options related to Bayesian estimation

- POINT (mean, median, mode; default is median) – defines which parameter of the posterior sample is used as an estimate for a given model parameter
- CHAINS (default is 2) - defines the number of independent MCMC chains
- BSEED – sets seed for random number generator
- STVALUES (= ML, PERTURBED, UNPERTURBED) – sets a method for computing starting values
- ALGORITHM (GIBBS, MH; default is GIBBS) – sets an MCMC algorithm
- BCONVERGENCE (related to PSR) – sets a threshold for PSR (0.05 is default, which means that PSR > 1.05 indicates non-convergence)
- BITERATIONS (to go beyond 50K iterations)
- FBITERATIONS (fixed number of iterations)
- THIN (every k-th iteration recorded; default is 1) – set thinning interval to avoid high amount of autocorrelation in the posterior distribution

MPLUS Default Priors

- Intercepts, regression slopes, loadings: N(0, infinity), unless these parameters are in a probit regression in which case N(0, 5)
- Variances: IG(0, -1) – Inversed Gamma distribution
- Covariance matrices: IW(0, -p-1), unless the elements include parameters from a probit regression in which case IW(I, p+1) is used
- Thresholds: N(0, infinity)
- Class proportions: Dirichlet prior D(10, 10, …, 10)

Example I: Second-order CFA for the Index of Emancipative Values

- The 6th wave of WVS (2010-2014)
- 67 countries and 69137 individuals (NAs were listwise deleted
- 13 manifest variables, 4 first-order factors (Equity, Liberty, Autonomy, Expression)

The Index of Emancipative Values

Equity:

- V45. Men have more right for job (three-category item)
- V51. Men are better political leaders (four-category item)
- V52. Education is more important for boys (four-category item)

Autonomy (child qualities, mentioned/not mentioned):

- V12. Independence
- V15. Imagination
- V19. Religious faith
- V21. Obedience

Liberty:

- V203. Homosexuality can be justified (1-10 scale)
- V204. Abortion can be justified (1-10 scale)
- V205. Divorce can be justified (1-10 scale)

Expression (V60-V63):

- E1. =2 if "*giving people more say in government*" is the first choice, =1 if "*giving people more say in government*" is the second choice, and =0 otherwise
- E2. =2 if "*giving people more say in local affairs*" is the first choice, =1 if "*giving people more say in local affairs*" is the second choice, and =0 otherwise
- E3. =2 if "*protecting freedom of speech*" is the first choice, =1 if "*protecting freedom of speech*" is the second choice, and =0 otherwise

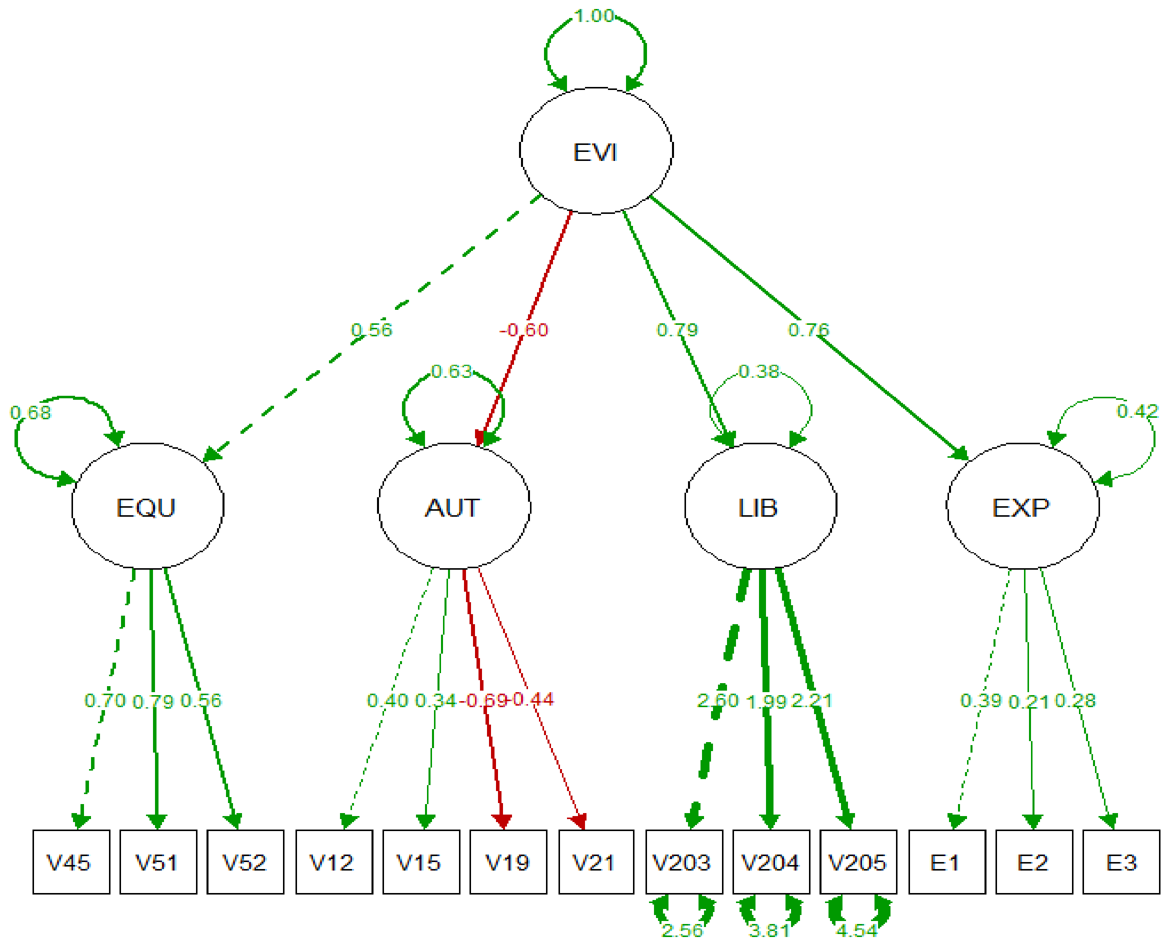
Example I. Second-Order CFA using frequentist approach

```
TITLE: Second-Order CFA model for EVI;
DATA: FILE = "C:/Users/lssi/Desktop/wvs.dat";
VARIABLE:
NAMES = V2 V45 V51 V52 V12 V15 V19 V21 V203
V204 V205 V60 V61 V62 V63 E1 E2 E3;
USEVARIABLES ARE V45 V51 V52 V12 V15 V19 V21 V203
V204 V205 E1 E2 E3;
CATEGORICAL ARE V45 V51 V52 V12 V15 V19 V21 E1 E2 E3;

MODEL:
Equity BY V45 V51 V52;
Autonomy BY V12 V15 V19 V21;
Liberty BY V203 V204 V205;
Expression BY E1 E2 E3;
EVI BY Equity Autonomy Liberty Expression;

OUTPUT:
STANDARDIZED;
```

Example I. Second-Order CFA using Bayesian approach

```
TITLE: Second-Order CFA model for EVI;
DATA: FILE = "C:/Users/lssi/Desktop/wvs.dat";
VARIABLE:
NAMES = V2 V45 V51 V52 V12 V15 V19 V21 V203
V204 V205 V60 V61 V62 V63 E1 E2 E3;
USEVARIABLES ARE V45 V51 V52 V12 V15 V19 V21 V203
V204 V205 E1 E2 E3;
CATEGORICAL ARE V45 V51 V52 V12 V15 V19 V21 E1 E2 E3;

DEFINE:
STANDARDIZE V203 V204 V205; ! It is recommended to standardize all continuous variable prior to MCMC estimation

ANALYSIS:
ESTIMATOR=BAYES; ! Choose Bayesian approach
CHAINS = 2;
PROCESS = 2;
FBITERATIONS = 15000; ! Choose maximum number of MCMC iteration

MODEL:
Equity BY V45 V51 V52;
Autonomy BY V12 V15 V19 V21;
Liberty BY V203 V204 V205;
Expression BY E1 E2 E3;
EVI BY Equity Autonomy Liberty Expression;

OUTPUT:
STAND(STDYX);
TECH1 TECH8; !TECH8 produce information about PSR and Kolmogorov-Smirnoff tests for convergence

PLOT:
TYPE = PLOT2; !Produce traceplots, autocorrelation plots, plots for prior and posterior distributions, and posterior predictive scatterplots and
    histograms
```

Example I. WLSMV estimates



Example I. CFA and BSEM comparison

WLSMV ESTIMATES

	Estimate	SE	t	p
EQUITY BY				
V45	0.702	0.004	179.693	0.000
V51	0.790	0.004	212.586	0.000
V52	0.564	0.004	149.967	0.000
AUTONOMY BY				
V12	0.404	0.007	60.126	0.000
V15	0.345	0.008	44.571	0.000
V19	-0.694	0.007	-93.924	0.000
V21	-0.439	0.007	-66.933	0.000
LIBERTY BY				
V203	0.852	0.004	225.502	0.000
V204	0.713	0.003	208.318	0.000
V205	0.720	0.004	204.017	0.000
EXPRESSI BY				
E1	0.388	0.008	47.530	0.000
E2	0.214	0.007	30.081	0.000
E3	0.280	0.008	34.207	0.000
EVI BY				
EQUITY	0.564	0.006	98.246	0.000
AUTONOMY	-0.605	0.008	-78.615	0.000
LIBERTY	0.786	0.007	119.515	0.000
EXPRESSION	0.764	0.015	51.537	0.000

BAYESIAN ESTIMATES

	Estimate	S.D.	p	2.5%	97.5%	
EQUITY BY						
V45	0.684	0.004	0.000	0.675	0.692	*
V51	0.784	0.004	0.000	0.777	0.792	*
V52	0.593	0.004	0.000	0.585	0.601	*
AUTONOMY BY						
V12	0.429	0.007	0.000	0.416	0.442	*
V15	0.308	0.007	0.000	0.294	0.322	*
V19	-0.666	0.007	0.000	-0.680	-0.653	*
V21	-0.472	0.007	0.000	-0.485	-0.459	*
LIBERTY BY						
V203	0.748	0.002	0.000	0.743	0.753	*
V204	0.786	0.002	0.000	0.782	0.791	*
V205	0.765	0.002	0.000	0.760	0.769	*
EXPRESSI BY						
E1	0.563	0.013	0.000	0.539	0.587	*
E2	0.283	0.008	0.000	0.267	0.299	*
E3	0.191	0.009	0.000	0.172	0.209	*
EVI BY						
EQUITY	0.566	0.006	0.000	0.553	0.578	*
AUTONOMY	-0.634	0.008	0.000	-0.648	-0.618	*
LIBERTY	0.777	0.007	0.000	0.764	0.791	*
EXPRESSION	0.527	0.015	0.000	0.499	0.556	*

Example I. CFA and BSEM comparison: model fit

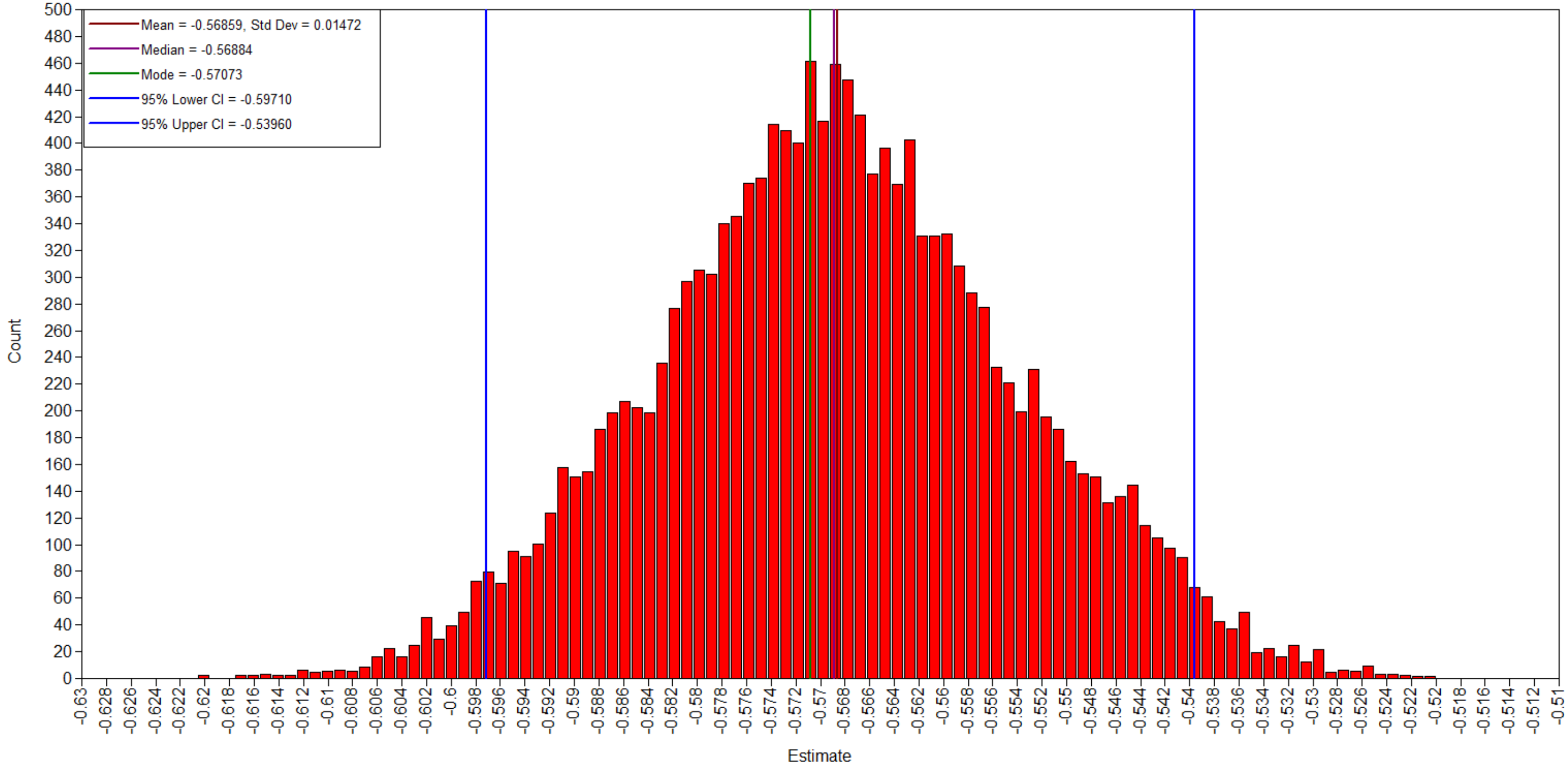
FREQUENTIST

RMSEA	0.048
P RMSEA <= .05	1.000
CFI	0.905
TLI	0.878

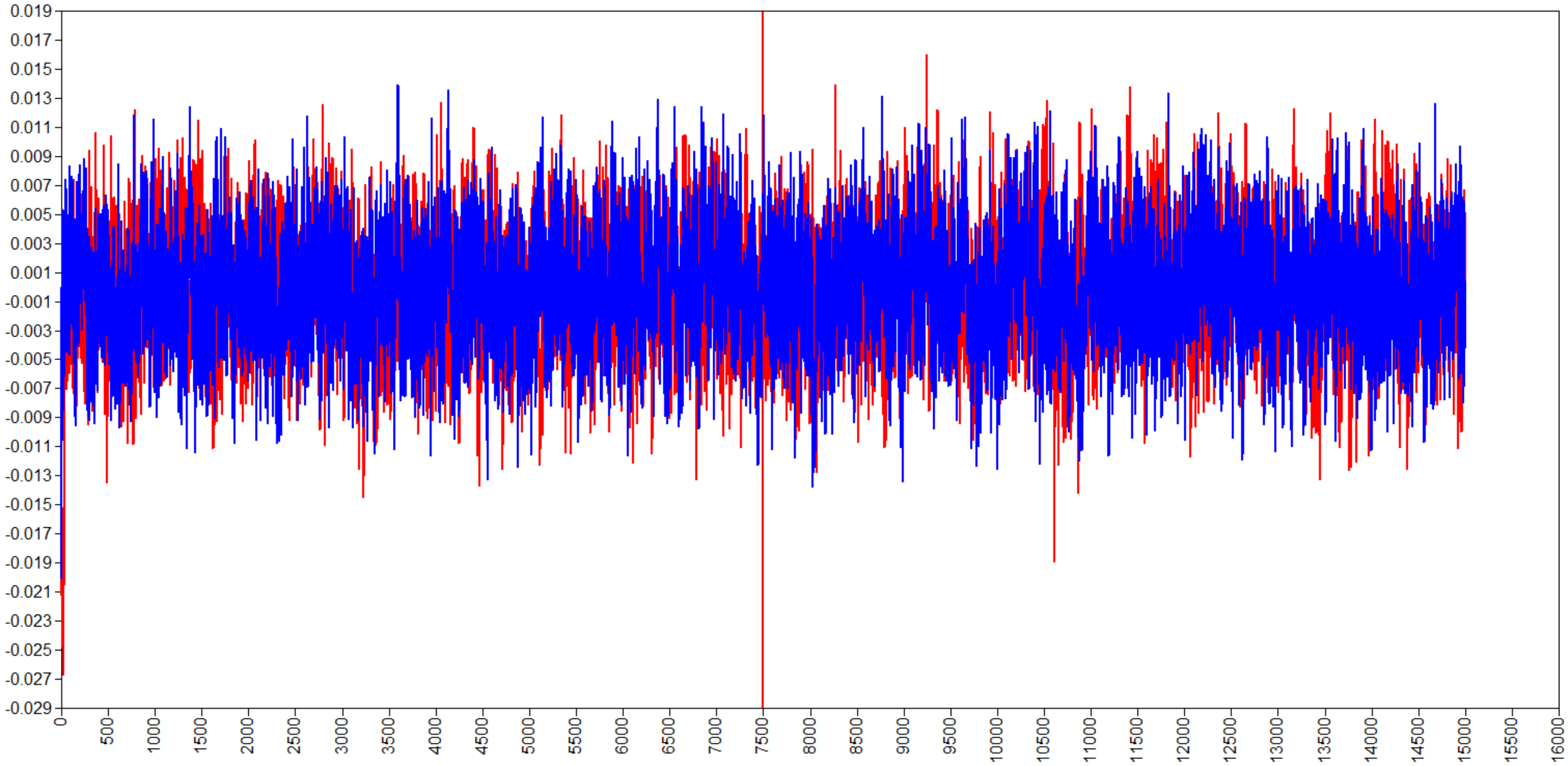
BAYESIAN

Posterior Predictive P-Value	0.000

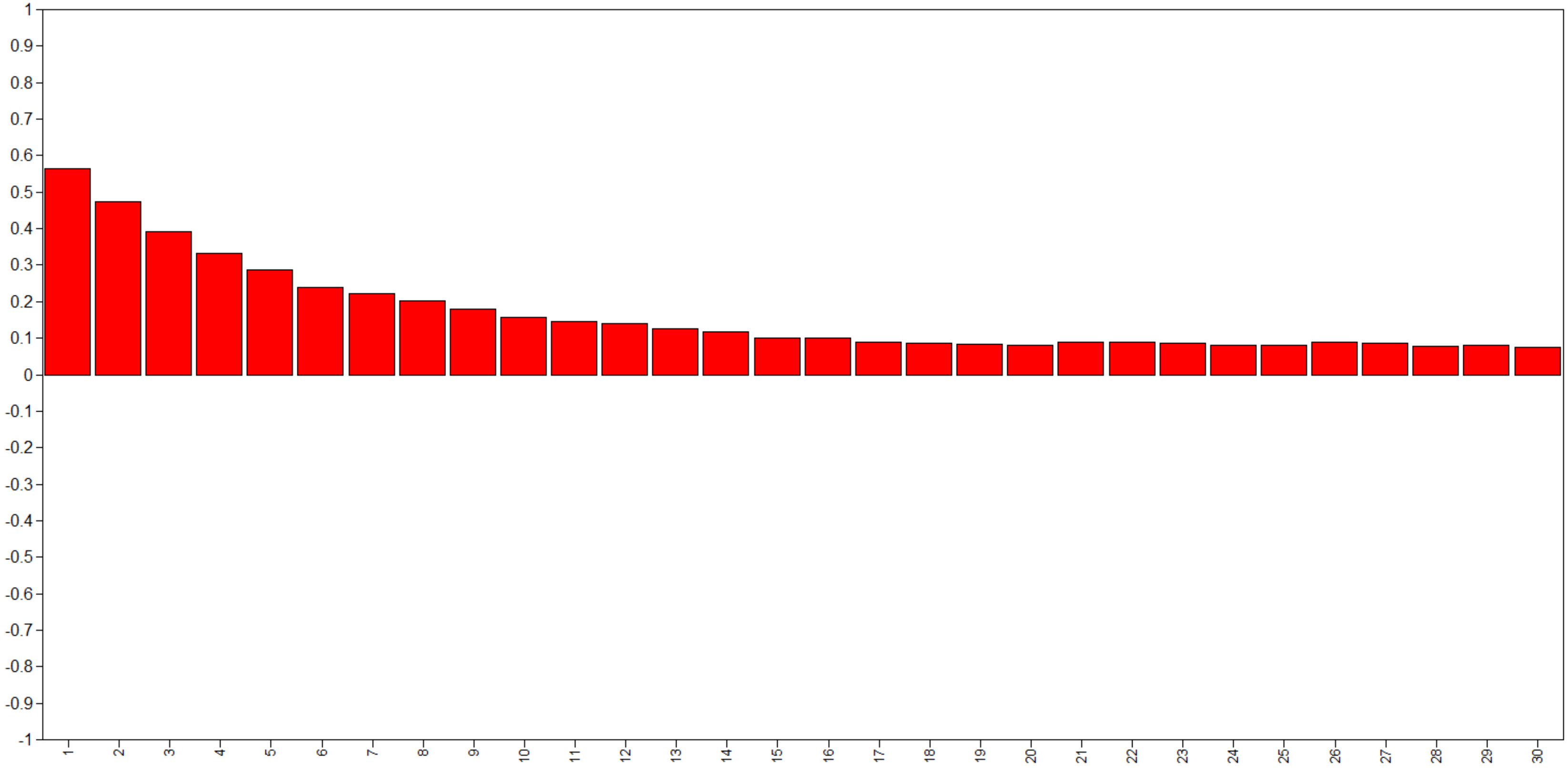
MPLUS histogram of the posterior distribution of parameter estimate Second-Order Factor Loading for Autonomy



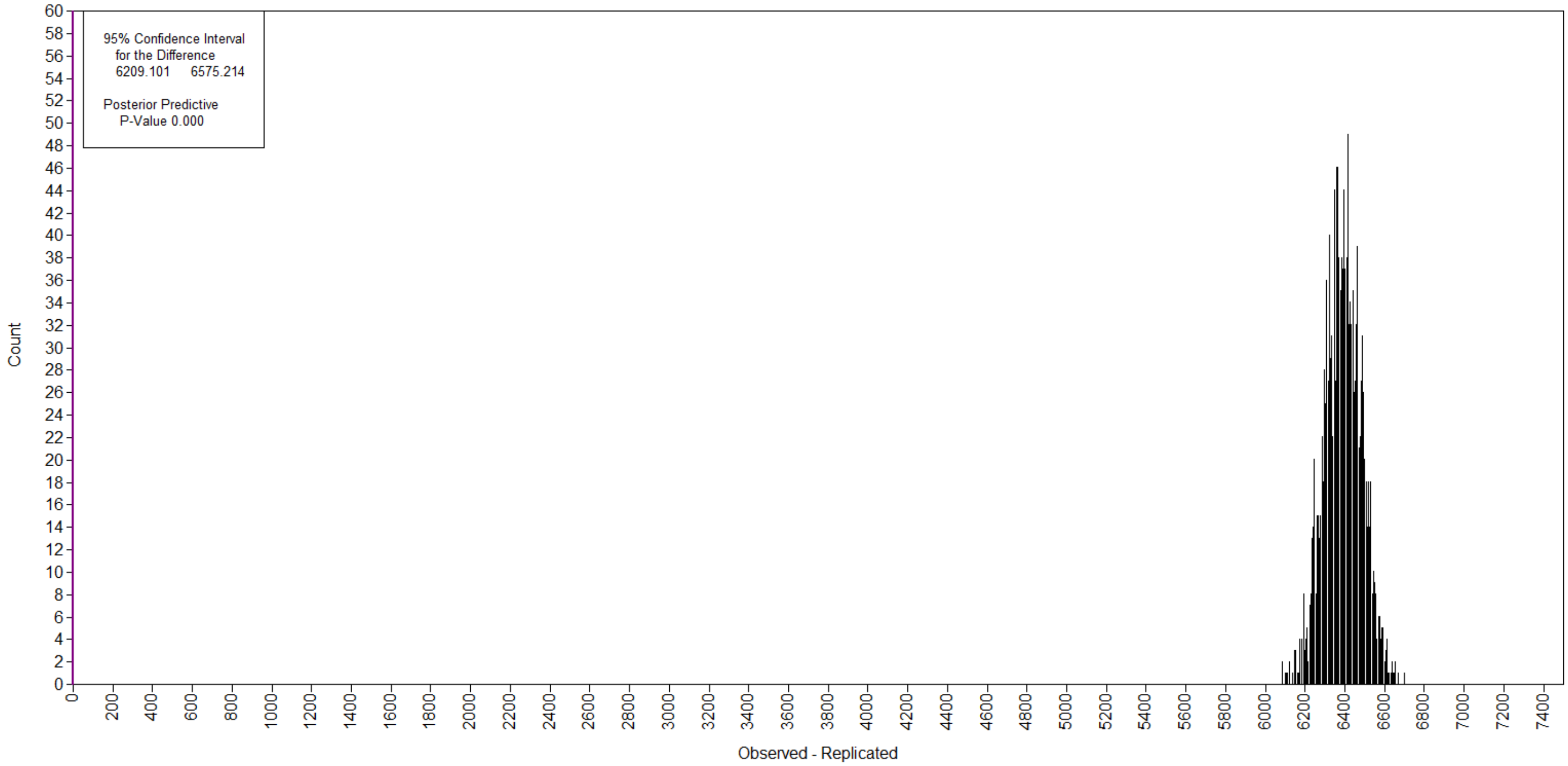
An example of MPLUS trace plot (for two chains)



An example of MPLUS autocorrelation plot



An example of MPLUS PPP histogram



Example II. Informative Priors for Cross-Loadings in CFA

- Both frequentist and Bayesian CFA for the construct behind the Index of Emancipative Values indicate lack of model fit.
- Misspecified cross-loadings may be a possible source of misfit (Muthen and Asparouhov 2015)

Example II MPLUS input

```
MODEL:
Equity BY V45* V51 V52
V12 V15 V19 V21 V203 V204 V205 E1 E2 E3(xload1-xload10); !specify cross-loadings and set labels for them
Autonomy BY V12* V15 V19 V21
V45 V51 V52 V203 V204 V205 E1 E2 E3 (xload11-xload19);
Liberty BY V203* V204 V205
V45 V51 V52 V12 V15 V19 V21 E1 E2 E3(xload20-xload29);
Expression BY E1* E2 E3
V45 V51 V52 V12 V15 V19 V21 V203 V204 V205 (xload30-xload39);
EVI BY Equity* Autonomy Liberty Expression;
MODEL PRIORS:
xload1-xload12~N(0, 0.01); ! set prior variance of 0.01 for cross-loadings
```

Example II MPLUS output: EQUITY

Variable	Estimate	SE	
V45	**0.816**	**0.108**	*
V51	**1.242**	**0.159**	*
V52	**0.788**	**0.101**	*
V12	0.025	0.038	
V15	-0.016	0.018	
V19	0.081	0.038	*
V21	-0.042	0.044	
V203	0.145	0.050	*
V204	-0.059	0.050	
V205	0.035	0.050	
E1	0.179	0.027	*
E2	0.010	0.106	
E3	0.189	0.026	*

Example II MPLUS output: AUTONOMY

Variable	Estimate	SE	
V12	**-1.249**	**0.615**	*
V15	**-0.475**	**0.241**	*
V19	**0.985**	**0.481**	*
V21	**1.453**	**0.693**	*
V45	0.045	0.049	
V51	0.082	0.071	
V52	-0.200	0.136	
V203	0.110	0.063	
V204	0.066	0.066	
V205	-0.123	0.116	
E1	0.020	0.023	
E2	0.005	0.103	
E3	0.085	0.049	

Example II MPLUS output: LIBERTY

Variable	Estimate	SE	
V203	**0.559**	**0.108**	*
V204	**0.713**	**0.152**	*
V205	**0.673**	**0.144**	*
V12	0.001	0.065	
V15	-0.094	0.024	*
V19	0.214	0.050	*
V21	0.012	0.074	
V45	0.111	0.052	*
V51	-0.009	0.099	
V52	-0.023	0.058	
E1	0.081	0.015	*
E2	0.009	0.094	
E3	0.085	0.014	*

Example II MPLUS output: EXPRESSION

Variable	Estimate	SE	
E1	**0.085**	**0.029**	*
E2	**2.626**	**0.559**	*
E3	**-0.113**	**0.034**	*
V203	0.012	0.026	
V204	-0.012	0.024	
V205	-0.010	0.023	
V12	-0.003	0.017	
V15	-0.028	0.014	*
V19	-0.003	0.022	
V21	-0.004	0.019	
V45	0.020	0.026	
V51	0.015	0.033	
V52	-0.030	0.018	*

Example II. Second-order factor loadings

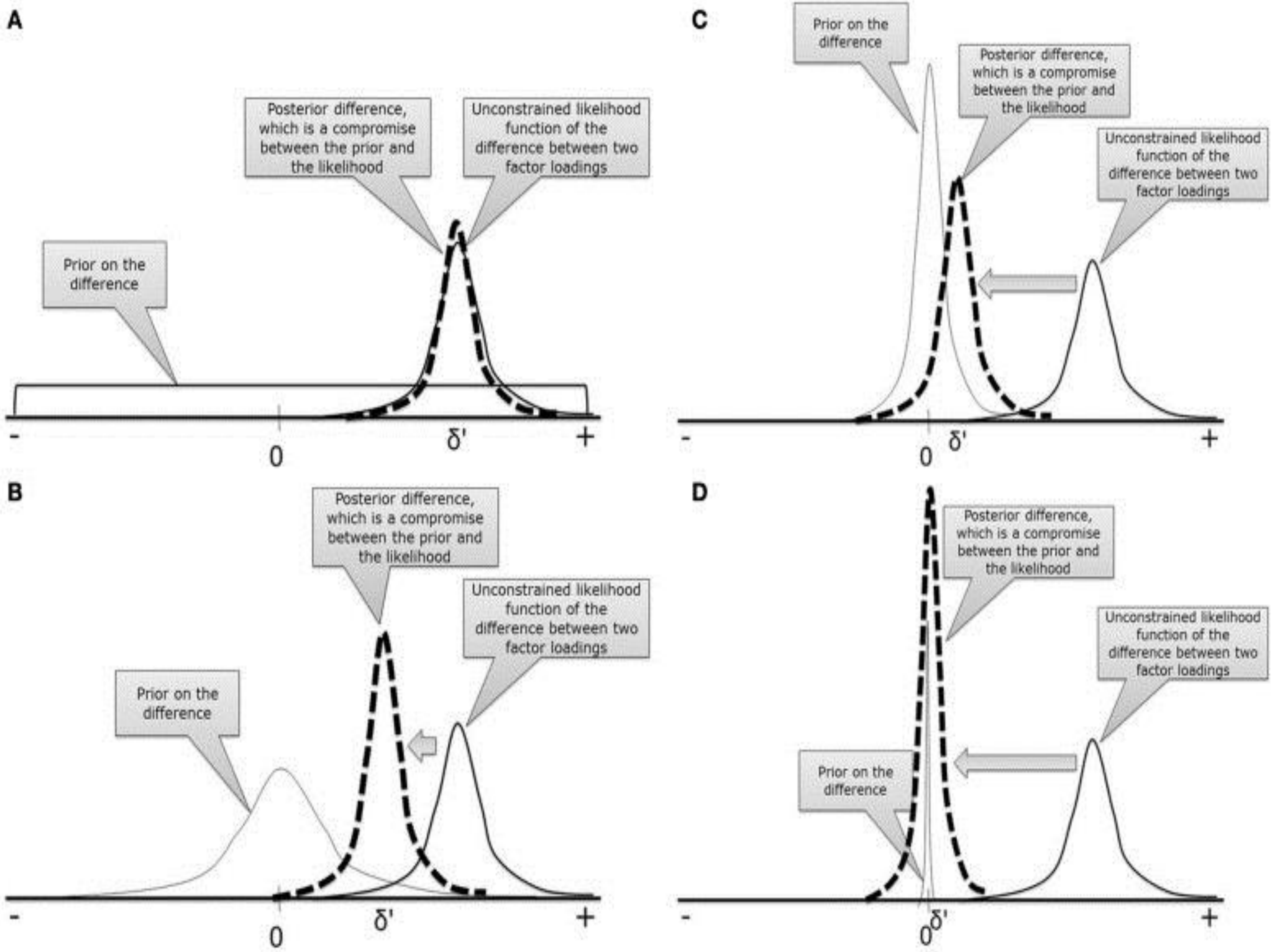
Variable	Estimate	SE	
EQUITY	1.000	0.000	(fixed for identification purposes)
AUTONOMY	0.365	0.285	
LIBERTY	1.792	0.600	*
EXPRESSION	0.845	0.493	*

Example II Interpretation

- PPP = 0.000
- While some cross-loadings are significant, the modified model does not fit data much better than the initial one
- Another possible source of misfit is the presence of small residual covariances (Muthen and Asparouhov 2015)

Example III. Approximate (Bayesian) Measurement Invariance

- Maximum likelihood measurement invariance CFA is suitable for cases with only a few groups and a small number of non-invariant parameters due to computational complexities
- Strict parameter invariance is often an unrealistic assumption
- MCMC is a powerful computational method that handle high-dimensional models better than ML
- Approximate measurement invariance: does not fix across-group parameter differences equal to zero, but treat them as additional parameters with small prior variance


A
Posterior difference, which is a compromise between the prior and the likelihood
Unconstrained likelihood function of the difference between two factor loadings
Prior on the difference
-
0
δ'
+
B
Posterior difference, which is a compromise between the prior and the likelihood
Unconstrained likelihood function of the difference between two factor loadings
Prior on the difference
-
0
δ'
+
C
Prior on the difference
Posterior difference, which is a compromise between the prior and the likelihood
Unconstrained likelihood function of the difference between two factor loadings
-
0
δ'
+
D
Posterior difference, which is a compromise between the prior and the likelihood
Unconstrained likelihood function of the difference between two factor loadings
Prior on the difference
-
0
δ'
+

Example III

- **Liberty** component of the Index of emancipative values
  - Three variables (V203, V204, V205)
  - The 6th wave of WVS (2010-2014)
  - 67 countries and 69137 individuals (NAs were listwise deleted)
  - 626 free parameters
- Comparing four models:
  - Model 1: all loadings and intercepts set to be equal across countries
  - Model 2: weakly informative priors for loadings and intercepts: N(0, 0.01)
  - Model 3: weakly informative priors for loadings and intercepts: N(0, 0.03)
  - Model 4: weakly informative priors for loadings and intercepts: N(0, 0.05)

Example III MPLUS input

```
TITLE: Bayesian  multiple group model with approximate
 measurement invariance;
DATA: FILE = ex5.33.dat;
VARIABLE: NAMES = country V1-V3;
 USEVARIABLES = country V1-V3; !V1, V2, and V3 are for V203, V204 and V206 respectively
 CLASSES = c(67);
 KNOWNCLASS = c(country = 1-67);

ANALYSIS: TYPE = MIXTURE; !set multi-group framework
ESTIMATOR = BAYES;
CHAINS = 2;
 PROCESSORS = 2;
FBITERATIONS = 30000;
 MODEL = ALLFREE; ! changes the mixture default of across-class equality of  model parameters when using BY

MODEL: %OVERALL%
 f1 BY V1-V3* (lam#_1-lam#_3); !set labels for factor loadings in groups  1-67: # is for group
  [V1-V3] (nu#_1-nu#_3); !set labels for intercepts

MODEL PRIORS:
 DO(1,3) DIFF(lam1_#-lam67_#)~N(0,0.01); ! set prior variance for factor loadings  equal to 0.01
 DO(1,3) DIFF(nu1_#-nu67_#)~N(0,0.01); ! set prior variance for intercepts equal to 0.01

OUTPUT: TECH1
TECH8;
PLOT: TYPE = PLOT2;
```

Comparison of the models with different levels of invariance

- Model1: PPP = 0.000
- Model 2: PPP = 0.005
- **Model 3: PPP = 0.272**
- Model 4: PPP = 0.366

- Prior variance of 0. 03 means that 95% of the deviations of loadings and intercepts of their across-group average values lay between -0.34 and 0.34 which, on a standardized variable scale, approaches a reasonable loading size
- It clearly indicates that approximate invariance does not hold even for the single component of the EVI. What about the general index?
- Multilevel Bayesian CFA as a way to explain non-invariance (Davidov et al. 2012)

MPLUS output: ‘difference’ section

- Differences between country-specific loadings and the sample average loading (for the first ten countries)

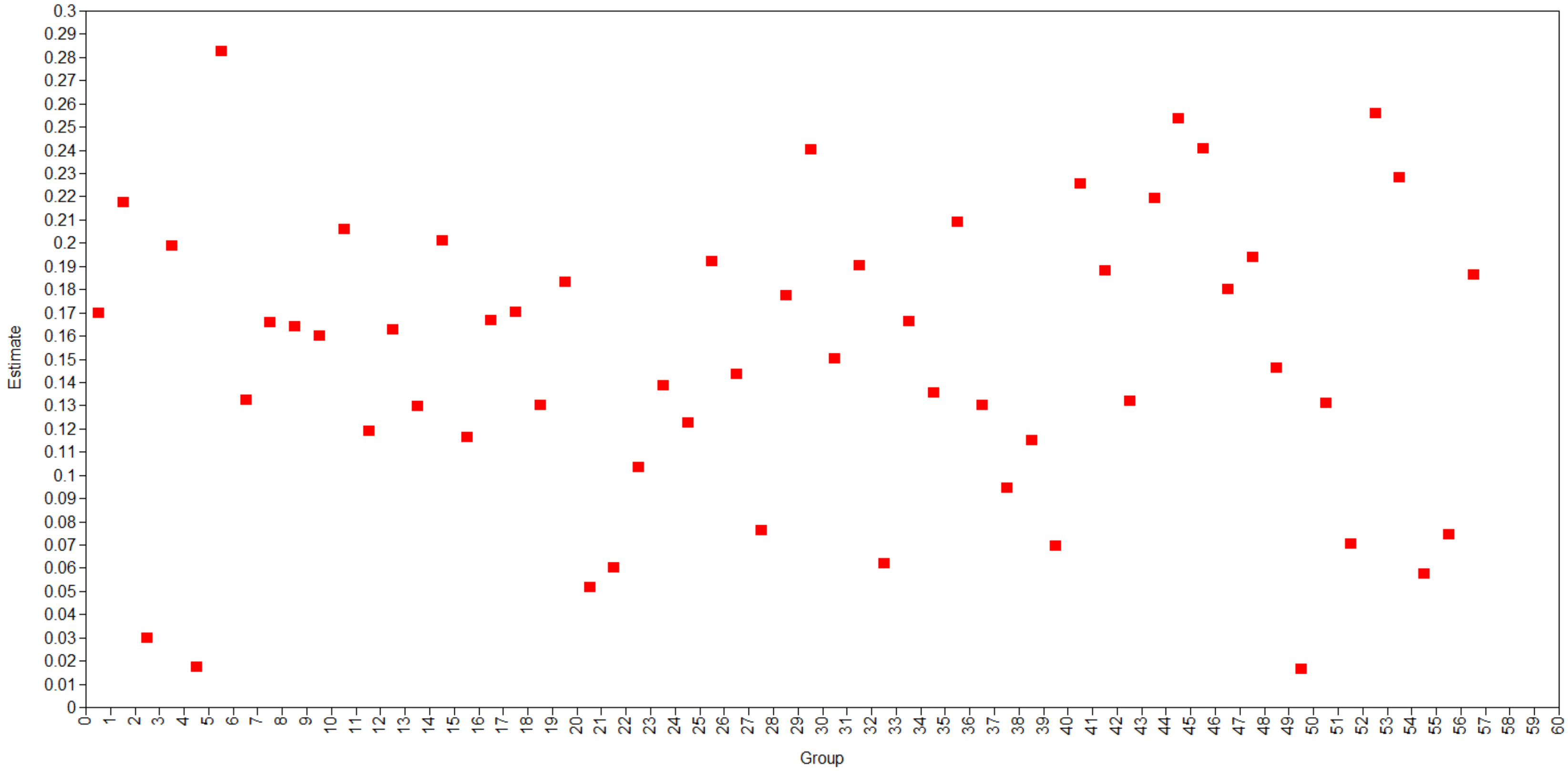
LAM1_1	LAM2_1	LAM3_1	LAM4_1	LAM5_1
0.013	0.063	-0.120*	0.035	-0.136*

LAM6_1	LAM7_1	LAM8_1	LAM9_1	LAM10_1
0.132*	-0.028	0.011	0.020	0.005

....

- Loadings in countries 3, 5 and 6 are significantly different from the average. (3-Argentina, 5- Bahrain, 6 - Armenia)

MPLUS Invariance Plot



Further reading: Bayesian SEM

- **Asparouhov, T. & Muth´en, B. (2010).** Bayesian analysis of latent variable models using Mplus. Technical Report. Version 4. www.statmodel.com.
- **Muthén, B. (2010).** Bayesian analysis in Mplus: A brief introduction. *Unpublished manuscript. www. statmodel. com/download/IntroBayesVersion*, *203*.
- **Muthen, B. & Asparouhov, T. (2012).** Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313- 335.

Further reading: AMI

- **Muthén, B., & Asparouhov, T. (2013)**. BSEM measurement invariance analysis. *Mplus Web Notes*, *17*, 1-48.
- **Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013).** Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770. doi:10.3389/fpsyg.2013.00770
- **Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012).** Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, 43(4), 558-575.

Thank you for attention!

This report was presented at the training methodological workshop
"Economic and Social Changes: values effects across Eurasia”.

March 31 - April 6, 2015 – Turkey.

http://lcsr.hse.ru/en/seminar_m2015

Настоящий доклад был представлен на методологическом учебном семинаре
«Экономические и социальные изменения: оценка эффектов по всей Евразии».

31 марта – 6 апреля 2015 года – Турция.

http://lcsr.hse.ru/seminar_m2015