

# Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models

## ABSTRACT

Latent variable models are common in the social sciences - to measure ideal points of U.S. Senators, countries' "level of democracy" or the relationships between latent attitudes and values across countries, for instance. Because differences in measurement parameters can be confounded with substantively interesting differences, measurement invariance or "equivalence" is a prerequisite for cross-group comparisons of parameters of interest. The practice of "invariance testing" attempts to rule out confounding by testing equality-constrained models. However, some tests may be rejected due to slight violations of invariance that are inconsequential for the comparison of interest. Conversely, even when the invariance hypothesis fits "closely", measurement inequivalence may still bias comparisons of interest substantially.

This article explores an alternative approach: evaluating directly whether parameters of interest are affected by possibly misspecified measurement invariance restrictions. A sensitivity measure, the "EPC-interest", is shown to provide valuable insight in whether groups can be considered equivalent

for the substantive comparison at hand. A simulation study demonstrates that the EPC-interest provides accurate estimates of the change in parameters of interest that would occur if particular measurement invariance restrictions were freed. We demonstrate the EPC-interest with two examples from the literature: the measurement of Democracy, and the relationship between values and anti-immigration attitudes.

R code and data for the examples discussed in this article are provided in the electronic appendix.

## 1. INTRODUCTION

Latent variable models are a common tool across the social sciences to model unobserved traits (Lord and Novick, 1968; Bollen, 1989; Bartholomew, Knott and Moustaki, 2011). In political science, for example, Jackman (2001) and Clinton, Jackman and Rivers (2004) applied ideal point models to roll calls in the U.S. Supreme Court, Senate, and House; Treier and Jackman (2008) and Armstrong (2011) discussed measurement models for the level of democracy based on the Polity and Freedom House indicators; and Davidov (2009) studied the measurement of national identity and constructive patriotism in the cross-national ISSP survey. The final goal of such analyses is often a comparison of estimated levels of the latent variable across groups; for instance, time series analyses of the level of democracy (Armstrong, 2011), a regression of risk of civil war on the level of democracy (Treier and Jackman, 2008), or a comparison across 34 countries of patriotism and nationalism (Davidov, 2009). Comparing relationships between latent variables may also be of interest - Davidov et al. (2008), for example, compared the regression coefficients between “human values” (Schwartz and Bilsky,

1987) and latent anti-immigration attitudes across 19 European countries.

Comparisons of latent variable means, relationships, or other parameters of interest are only valid, however, when the measurement parameters are equal across the groups to be compared (e.g. Steenkamp and Baumgartner, 1998). Otherwise, between-group differences in the parameters of interest may be confounded with differences in the measurement parameters (Oberski, 2012). In the literature on structural equation modeling (SEM), this observation is known as the requirement of “measurement equivalence” or “invariance” (Meredith, 1993); psychometricians use the terms “item bias”, also known as “differential item functioning” or DIF (Mellenbergh, 1989); and in the measurement error literature the corresponding term is “differential measurement error” (Carroll et al., 2006). We focus on structural equation modeling – a general class of linear latent variable models that has given rise to a large literature on measurement invariance testing (for reviews, see Millsap and Everson, 1993; Vandenberg and Lance, 2000; Schmitt and Kuljanin, 2008).

Invariance testing in SEM proceeds by comparing a model with cross-group equality restrictions on the measurement parameters to a model in which some or all of the measurement parameters are allowed to vary over groups. This comparison may take the form of a chi-square difference test (Steenkamp and Baumgartner, 1998; French and Finch, 2006), of a test of “close fit” involving SEM fit measures such as CFI and RMSEA (Cheung and Rensvold, 2002; Chen, 2007), or of the examination of expected parameter changes (EPC) and modification indices (MI) (Byrne, Shavelson and Muthén, 1989; Yoon and Millsap, 2007). A commonly recognized problem is that the chi-square difference test may be very sensitive to “small” differences in measurement parameters that do not substantially change the parameters of interest. Indeed this has been one of the driving arguments behind the development of “close fit” measures such as CFI and RMSEA (Hu and Bentler, 1998). The converse possibility also ex-

ists, however: both the chi-square difference test and (differences in) fit measures such as CFI and RMSEA may indicate a good model fit even though the “closely” fitting between-group differences in measurement parameters produce large biases in the parameters of interest (Kolenikov, 2009; Saris, Satorra and Van der Veld, 2009). As a partial solution, Saris, Satorra and Van der Veld (2009) suggested examining the MI, EPC, and power of the MI test. However, unless the power is high and no modification indices statistically significant, similar problems occur with this procedure: the MI and EPC may indicate a “substantial misspecification”, i.e. between-group differences in measurement parameters, even though this difference does not produce any substantial bias in the parameters of interest - and vice versa. Therefore, when invariance testing indicates violations of invariance, such violations do not necessarily lead to substantial biases in the parameters of interest. Moreover, even when fit measures indicate that measurement invariance restrictions yield a “close fit”, there may still be substantial bias in the parameters of interest such as differences in latent factor means and relationships, threatening the validity of the comparison.

In this article, we supplement the investigation of invariance testing in SEM with a sensitivity analysis. We suggest using the “EPC-interest” for this purpose: a measure of the expected change in the parameter of interest when freeing a particular equality restriction (Satorra, 1989; Bentler and Chou, 1992). Given parameters of interest to be compared across groups, the EPC-interest allows the researcher to establish more directly whether such a comparison is valid.

The EPC-interest is similar to the existing EPC measures in SEM in that it can be applied to models involving equality restrictions, and also estimates the expected change in a parameter after freeing each restriction. However, the EPC estimates change in the restricted measurement parameter itself, while the EPC-interest estimates change in the parameter(s) of interest. In this sense, it evaluates the sensitivity

of the substantive model of interest to the invariance restrictions, and is similar in spirit to the approach for causal inference discussed by Imai and Yamamoto (2010, pp. 552-3). The EPC-interest approach differs somewhat from the purely derivative-oriented approach to sensitivity analysis common in econometrics (Magnus and Vasnev, 2007, p. 168) and applied to SEM by Yuan, Marshall and Bentler (2003): both direction and magnitude of the misspecification are combined into the same measure here. A contingent hypothesis test of no change in the parameters of interest is, however, equivalent to classic econometric specification tests (Yuan, Marshall and Bentler, 2003; Hausman, 1978). Changes in parameters of interest for specific combinations of measurement and structural models were derived by Millsap (1997), Millsap and Kwok (2004), Millsap (2007), and Meuleman (2012). The EPC-interest can be seen as a general method of obtaining such results applicable to all structural equation models with equality-constrained measurement parameters.

The following section defines the EPC-interest for structural equation models with equality constraints. Subsequently, a simulation study evaluates the finite sample performance of the EPC-interest as an estimate of the shift in parameters of interest when freeing misspecified equality restrictions, as well as its robustness to misspecification in the alternative model. Sections 4 and 5 demonstrate the use of the EPC-interest on two latent variable analyses from the literature where comparisons across groups were of interest (see the digital appendix for R code and data). The final section summarizes the findings and discusses some limitations and future work on the use of the EPC-interest for evaluating invariance hypotheses.

## 2. ASSESSING THE EFFECT OF MISSPECIFIED INVARIANCE RESTRICTIONS ON SEM PARAMETERS OF INTEREST

A structural equation model is any model  $\Sigma(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta})$  that imposes a structure on the population covariance matrix  $\Sigma$  and mean vector  $\boldsymbol{\mu}$  of observed variables  $\mathbf{y}$  as a function of a vector  $\boldsymbol{\theta}$  of unknown model parameters (Bollen, 1989). A common parameterization of SEM is the LISREL “all-y” model for group  $g$ ,

$$\mathbf{y}_g = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_g + \boldsymbol{\epsilon}_g, \quad (1)$$

$$\boldsymbol{\eta}_g = \boldsymbol{\alpha}_g + \mathbf{B}_g \boldsymbol{\eta}_g + \boldsymbol{\zeta}_g, \quad (2)$$

where  $\boldsymbol{\eta}_g$  is a vector of latent variables, and  $\boldsymbol{\epsilon}_g$  and  $\boldsymbol{\zeta}_g$  are observed and latent variable residuals. The “measurement model” consists of the first equation involving as parameters the vector of intercepts  $\boldsymbol{\nu}_g$ , the loading matrix  $\boldsymbol{\Lambda}_g$ , and the residual variance matrix  $\text{Var}(\mathbf{y}_g | \boldsymbol{\eta}_g) = \text{Var}(\boldsymbol{\epsilon}_g) := \boldsymbol{\Psi}_g$ . The second equation is the “structural” part of the model with latent intercepts (or means)  $\boldsymbol{\alpha}_g$ , latent variable regression coefficients  $\mathbf{B}_g$ , and the (residual) variance matrix  $\text{Var}(\boldsymbol{\zeta}_g) := \boldsymbol{\Phi}_g$  as group-specific parameters. Assuming  $\boldsymbol{\eta}$ ,  $\boldsymbol{\epsilon}$ , and  $\boldsymbol{\zeta}$  are mutually uncorrelated, this model produces as moment structure implications

$$\Sigma(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Phi}(\mathbf{I} - \mathbf{B})^{-T} \boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \quad (3)$$

for the covariances and

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\alpha} \quad (4)$$

for the means, where for notational convenience the group-specific parameters have been stacked to obtain the block-diagonal covariance matrix over all groups.

Estimation of the parameters  $\mathbf{a} = [\boldsymbol{\nu}', \boldsymbol{\alpha}', (\text{vec } \boldsymbol{\Lambda})', (\text{vech } \boldsymbol{\Phi})', (\text{vec } \mathbf{B})', (\text{vech } \boldsymbol{\Psi})']'$  then proceeds by minimizing a nonnegative fitting function  $F(\mathbf{S}, \mathbf{m}; \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}))$  so that  $\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} F(\mathbf{S}, \mathbf{m}; \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}))$  (e.g. Satorra, 1989). A common choice of fitting function is that corresponding to maximum likelihood estimation under the assumption that  $\mathbf{y}$  is independently and identically distributed as multivariate normal (Bollen, 1989). Under the model, however, such distributional assumptions do not affect consistency of the parameter estimates (Satorra, 1989).

Models freely estimating all parameters in  $\mathbf{a}$  are generally not identifiable. Therefore, restrictions  $\mathbf{a} = \mathbf{a}(\boldsymbol{\theta})$  are imposed, such as setting certain loadings to zero, restricting the residual variance matrices to be diagonal, or allowing only recursive latent variable regression coefficients. Any subset of  $\mathbf{a}$  may play the role of the parameter interest; attention may focus on differences in the latent means  $\boldsymbol{\alpha}_g$  over groups  $g$ , for instance, or on differences in latent variable regressions  $\mathbf{B}_g$ . Although there is in principle no restriction on what may be defined as a parameter of interest, the structural parameters are more usually of direct interest than the measurement parameters.

To identify differences over groups in structural parameters pertaining to latent variables, however, it is necessary to impose cross-group equality restrictions on the measurement parameters: the measurement invariance restrictions. Even when no explicit restrictions are made, but, for instance, one loading is set to unity in each group to identify regression coefficients in each group, the assumption of invariance of such reference loadings is implicit (Hancock, Stapleton and Arnold-Berkovits, 2009). For identifiability of the comparison of interest, then, it is necessary to include in the restrictions  $\mathbf{a} = \mathbf{a}_0(\boldsymbol{\theta})$  equality restrictions on the measurement model. In general, equality of loadings  $\boldsymbol{\Lambda}_{g'} = \boldsymbol{\Lambda}_g \forall g \neq g'$  (“metric invariance”) is required for identification of parameters pertaining to the covariance matrix of the latent variables such as  $\mathbf{B}_g$ , while for identification of latent mean differences in  $\boldsymbol{\alpha}_g$ , equality of both load-

ings and intercepts  $\Lambda_{g'} = \Lambda_g; \nu_{g'} = \nu_g \forall g \neq g'$  (“scalar invariance”) is required. Of course, such restrictions may be misspecified, and the misspecifications may cause bias in the parameters of interest (Yuan, Marshall and Bentler, 2003; Millsap, 2007; Kolenikov, 2009). It is not necessary that the full measurement parameter vectors be equal however: partial invariance of at least two indicators per latent concept suffices for identification of the parameters of interest (Byrne, Shavelson and Muthén, 1989). This suggests that after estimation of a fully invariant model with restrictions  $\mathbf{a}_0(\boldsymbol{\theta})$ , a partially invariant alternative model  $\mathbf{a}_a(\boldsymbol{\theta})$  can be considered which frees one equality restriction. Alternatively,  $\mathbf{a}_0(\boldsymbol{\theta})$  may itself be a partially invariant model, assuming that it and the alternative model remain identifiable.

The EPC-interest of an equality-constrained measurement parameter with respect to a parameter of interest in the  $\mathbf{a}_0$  model is defined as a consistent estimate of the expected change in the parameter of interest if the equality constraint were freed in the  $\mathbf{a}_a$  model. Let the parameters of interest  $\boldsymbol{\pi}$  be defined as  $\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\theta}$ . Typically  $\mathbf{P}$  is a selection matrix, but  $\mathbf{P}$  may also produce a linear combination or contrast of free model parameters, for example the cross-group differences in regression coefficients. Let  $\mathbf{A}_a = \partial\mathbf{a}_a/\partial\boldsymbol{\theta}'$ . Then  $\mathbf{A}_a$  is a logical (0/1) matrix corresponding to the alternative model including the possible misspecification under consideration as though it were a free parameter. As shown in the appendix,

$$\text{EPC-interest} := \boldsymbol{\pi} - \hat{\boldsymbol{\pi}} \approx \mathbf{P}(\mathbf{A}'_a \mathbf{J}(\hat{\boldsymbol{\theta}}) \mathbf{A}_a)^{-1} \mathbf{g}(\hat{\boldsymbol{\theta}}) \mathbf{A}_a. \quad (5)$$

where  $\mathbf{g}(\hat{\boldsymbol{\theta}})$  and  $\mathbf{J}(\hat{\boldsymbol{\theta}})$  are consistent estimates of respectively the gradient and the hessian of the fitting function with respect to the unrestricted parameter vector ( $\mathbf{a}$ ), evaluated at the sample parameter estimates under the  $\mathbf{a}_0$  model (Satorra, 1989; Bentler and Chou, 1992). For the LISREL all-y model, Neudecker and Satorra (1991) provided

$\mathbf{g}$  and  $\mathbf{J}$  as a function of the parameter estimates: the EPC-interest depends only on the parameter estimates from the restricted  $\mathbf{a}_0$  model.

A key assumption in deriving the EPC-interest is that the hessian  $\mathbf{J}$  is approximately constant between the null and alternative model population parameter values (Satorra, 1989). This implies that the alternative model  $\mathbf{a}_a$  should not itself be strongly misspecified. Some degree of misspecification in the alternative model is allowed for; in this case the EPC-interest becomes an approximation to the shift in the parameter of interest if the equality restriction were freed. The following section will study the influence of alternative model misspecification on the EPC-interest statistic in an example, and suggests that the approximation may be rather robust to this assumption.

The EPC-interest (equation 5) allows the researcher to fit the invariance model and obtain consistent estimates of the effect of various restrictions on the parameters of interest. Thus, it is similar to the more familiar EPC (Saris, Satorra and Sörbom, 1987) in the sense that it gives an estimated shift in a parameter when freeing a restriction, the difference being that this shift is not in the restricted parameter itself but in the parameter(s) of interest<sup>1</sup>. To avoid confusion with the EPC, we will denote that measure as the “EPC-self” and its standardized version (Kaplan, 1989; Chou and Bentler, 1993) as the “SEPC-self”. Since partial invariance testing is done precisely because differences in measurement parameters may affect the parameters of interest, the EPC-interest should prove useful when evaluating whether particular equality restrictions should be maintained or not.

---

<sup>1</sup>If the parameter of interest were defined to be the restricted parameter itself, EPC-interest will equal EPC-self.

### 3. ACCURACY OF THE EPC-interest: MONTE CARLO SIMULATION AND POPULATION ROBUSTNESS

A small simulation study evaluates the performance of the EPC-interest in small samples as an estimate of the shift in a parameter of interest when freeing a measurement parameter. In this simulation, a two-group, one-factor model with three indicators was formulated. The parameter of interest was taken to be the latent mean difference between the two groups, which was set to 0.2. The unstandardized factor loadings were chosen to equal 1 for all indicators and in both groups, the latent variables were chosen to have variance equal to 1 in both groups, and the error variances of the three indicators were set to 0.5. Two indicators' intercepts were set to zero in both groups, but the third indicator's intercept violated scalar invariance.

The cross-group difference in the intercept (misspecification) was varied across simulation conditions ( $\Delta\tau_1 \in \{0.1, 0.3, 0.8\}$ ). The conditions also varied the number of observations for each of the two groups ( $n_g \in \{50, 100, 500\}$ ). For each of the nine resulting conditions, 200 samples were drawn from multivariate normal distributions based on the population model. In each sample, the misspecified full scalar invariance model was fit to the data. The EPC-self was then calculated for the misspecified parameter, as well as the EPC-interest of the misspecified parameter with respect to the latent mean difference. Table 1 shows the results of this simulation for each of the nine conditions.

Table 1 shows that the bias in the latent mean difference,  $\Delta\hat{\kappa}$ , is affected differently in the different conditions. As one would expect, the bias is larger in conditions with larger misspecifications. The EPC-interest statistic is meant to estimate this bias as a result of the misspecification in the intercept  $\tau_1$ . The average over repeated samples of EPC-interest is indeed very close to the actual bias in the latent mean difference.

Table 1: Monte Carlo simulation results show that the EPC-interest approximates the true latent mean bias well, even in small samples.

$\Delta\tau_1$	$n_g$	Average over 200 replications				
		EPC-self	$\Delta\hat{\kappa}$	$\Delta\hat{\kappa}$ bias	EPC-interest	EPC-interest bias
0.1	50	0.064	0.240	-0.040	-0.034	0.005
0.3	50	0.213	0.313	-0.113	-0.113	-0.001
0.8	50	0.657	0.505	-0.305	-0.401	-0.096
0.1	100	0.058	0.231	-0.031	-0.031	0.000
0.3	100	0.203	0.323	-0.123	-0.109	0.014
0.8	100	0.619	0.492	-0.292	-0.370	-0.077
0.1	500	0.063	0.233	-0.033	-0.033	0.000
0.3	500	0.208	0.307	-0.107	-0.112	-0.005
0.8	500	0.598	0.501	-0.301	-0.349	-0.048

This means that the EPC-interest gives a close estimate of the shift in latent mean difference estimate if the misspecified  $\tau_1$  parameter were freed. The estimate is close to the true bias even for the small sample size condition in which each group contains only 50 observations. The EPC-self appears to be uniformly underestimated.

The Monte Carlo simulation shows that the finite-sample estimate of EPC-interest is close to the population shift in the parameter of interest, even for small samples. However, the quality of the population approximation is affected by the degree of misspecification in the alternative model. If the alternative model is strongly misspecified, the EPC-interest becomes only an approximation to the actual population shift in the parameter of interest. Asymptotic robustness to the effect of misspecification in the alternative model can be studied by first calculating the population implied covariance matrix and then adding some amount of random variation to these population covariances to reflect the effect of misspecification. The scalar invariance model is then fitted to this shifted population covariance matrix and mean vector, and the EPC-interest calculated on it.

Table 2 shows the asymptotic robustness of the EPC-interest approximation to misspecification in the alternative model for increasing degrees of misspecification.

Table 2: Asymptotic robustness of the EPC-interest to misspecification of the alternative model. Shown is the difference between the EPC-interest and the true bias it is meant to approximate.

Misspecification	Mean	Std. dev.	First quartile	Median	Third quartile
Unif(-0.05, 0.05)	-0.005	0.004	-0.008	-0.005	-0.002
Unif(-0.10, 0.10)	-0.006	0.008	-0.012	-0.006	-0.001
Unif(-0.20, 0.20)	-0.006	0.020	-0.019	-0.007	0.005

Again the two-group, single factor three indicator model was used, with a difference in one of the intercepts of 0.3 and a true latent mean difference of 0.2. The population covariance matrix resulting from this model was then perturbed 200 times with uniform random numbers. The minimum and maximum of the perturbations increased in three conditions from 0.05 to 0.20. The scalar invariance model was fitted to each of the resulting perturbed matrices, and the EPC-interest calculated as well as the “estimate” of the latent mean difference of interest. If the EPC-interest is robust to misspecification of the alternative model, it should be close to the bias in the latent mean difference resulting from fitting the scalar invariance model to the perturbed matrix.

Table 2 shows the mean, standard deviation, and quartiles over the 200 perturbations of the difference between EPC-interest and the bias in the latent mean difference. It can be seen from the increase in standard deviation that this difference can increase with the amount of misspecification in the alternative model. However, the error in the approximation of the EPC-interest is in general small compared to the latent mean difference bias, which ranged between -0.30 and +0.05. This shows that even though the quality of the approximation in principle depends on the closeness of the alternative model to the true model, at least in the example studied this effect is minimal and the EPC-interest appears asymptotically robust to this assumption.

Table 3: Intercept estimates and standardized loadings from the scalar invariance confirmatory factor analysis. Chi-square: 104,  $df = 70$  ( $p = 0.005$ ), CFI = 0.997, RMSEA = 0.050 ( $p = 0.468$ ), SRMR = 0.010.

Indicator	Intercept	(s.e.)	Loading	(s.e.)
<i>Political rights</i>				
Electoral Process (A)	7.7	(0.303)	4.14	(0.218)
Political Pluralism (B)	10.2	(0.369)	5.06	(0.264)
Functioning of Government (C)	6.6	(0.251)	3.40	(0.182)
<i>Civil liberties</i>				
Freedom of Expression (D)	11.6	(0.307)	4.21	(0.220)
Associational Rights (E)	8.0	(0.267)	3.67	(0.192)
Rule of Law (F)	8.7	(0.320)	4.30	(0.233)
Personal Autonomy (G)	9.8	(0.267)	3.57	(0.196)

#### 4. EXAMPLE 1: MEAN LEVELS OF DEMOCRACY FACTORS

Armstrong (2011) presented a confirmatory factor analysis of seven indicators of the “level of democracy” obtained from Freedom House. Values for these seven indicators were observed for 193 countries in four subsequent years (2006–2009). Based on substantive decisions made by Freedom House, Armstrong (2011) estimated a model with two separate factors; a maximum-likelihood analysis of this model is shown in Table 3.

The authors allowed for differences over time in the loadings and intercepts. If intercepts and loadings are not equal over time, latent mean differences over time would not be identifiable (Steenkamp and Baumgartner, 1998). Therefore invariance testing is performed by estimating a model constraining both intercepts and loadings to be equal over time: the so-called “scalar invariance” model. Measurement invariance testing then consists of comparing the scalar invariance model fit with the fit for a model in which intercepts are allowed to vary over time, but loadings are constrained to be equal (“metric invariance”), and the model in which both loadings and intercepts are freed. None of the chi-square difference tests between these models are statistically significant ( $\Delta\chi^2 = 6.21$ ,  $df = 30$ ), differences in CFI and RMSEA are below the cutoffs

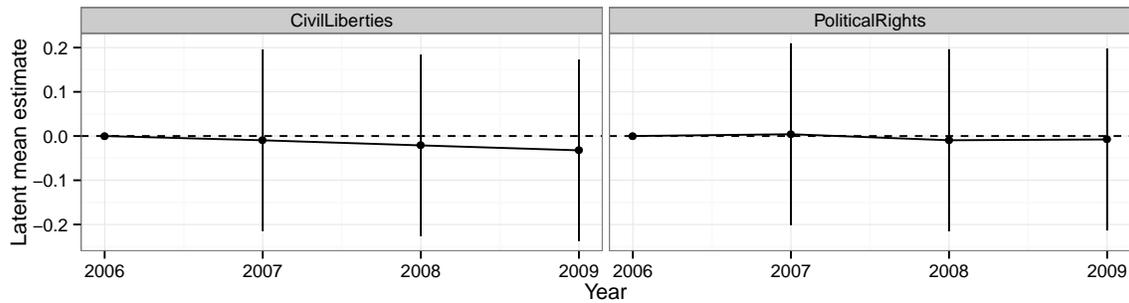


Figure 1: Change in “level of democracy” over time based on scalar invariance model of Freedom House data from 193 countries using the year as grouping variable. Latent means of factors Civil Liberties and Political Rights with two standard error-intervals. The dotted reference line indicates no average change relative to 2006.

recommended by Chen (2007), and the overall model fit, shown in Table 3, would be judged adequate.

The differences over time in the latent means of Political Rights and Civil Liberties may be of interest. These changes are plotted in Figure 1 based on the scalar invariance model. In Figure 1, the year 2006 is taken as the reference group by setting its estimate to zero. Figure 1 shows that, assuming measurement invariance over time, no changes in either of the factors are observed in this period. The EPC-interest can now be applied to assess the sensitivity of these changes to the scalar invariance assumption. All the  $|EPC\text{-interest}| < 0.008$ . This means that none of the estimates of change in the latent mean differences over time would change by more than 0.008 in absolute value if a scalar invariance restriction on loadings or intercepts were freed. The EPC-interest therefore yields much the same conclusion as the measurement invariance tests. It also provides a reasoning behind selecting the scalar invariance model: under the scalar invariance model, the parameters plotted in Figure 1 are identifiable and any effects of misspecification in the scalar invariance restrictions is too slight to substantially change Figure 1.

Another analysis of interest to Armstrong (2011) was the comparison of democracy

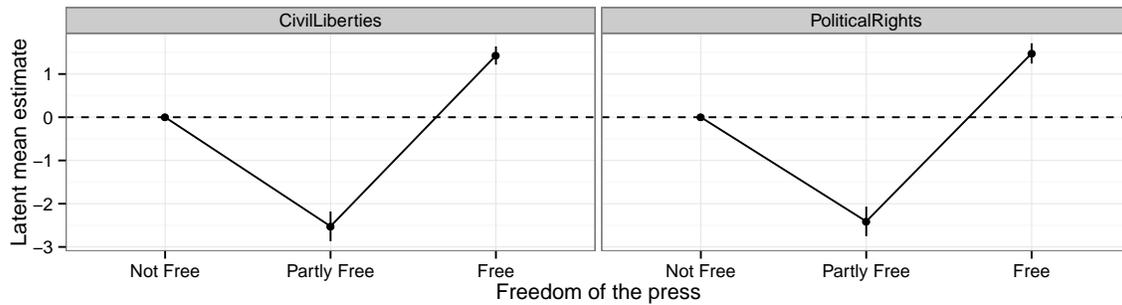


Figure 2: Latent means of Civil Liberties and Political Rights factors by level of freedom of the press based on the scalar invariance model. The dotted reference line indicates no difference relative to the reference “not free” group.

levels across countries with different levels of press freedom (low, middle, high). In this case the relevant grouping variable is not the year, but the press freedom variable. The same logic applies to this comparison as to the over-time comparison: measurement parameters (intercepts and loadings) should be the same for countries with low, middle, and high amounts of press freedom if these groups are to be comparable. Contrary to the across-time invariance test, however, imposing scalar invariance restrictions leads to a significantly worse model fit in terms of chi-square ( $\Delta\chi^2 = 747, df = 21$ ). The CFI’s for the free, metric invariance, and scalar invariance models are 0.900, 0.848, and 0.682 respectively, while the corresponding RMSEA’s are 0.232, 0.251, 0.322 respectively. This indicates that the scalar and metric invariance restrictions fit badly, suggesting that, strictly speaking, the regression of the Civil Liberties and Political Rights scores on Freedom of the Press is not valid, since the comparison over press freedom groups is possibly confounded with measurement differences. Figure 2 plots the latent mean estimates ignoring the lack of model fit of the scalar invariance model which allows for identification of these differences.

Invariance testing has indicated that Figure 2 may not provide a valid comparison, because it is based on a misspecified measurement invariance model. The EPC-interest

Table 4: EPC-interest of equality-constrained parameters with respect to latent mean differences shown in Figure 2.

Mean estimate of...	Group	EPC-interest when freeing...	
		F~1 in group “Free”	D~1 in group “Partly free”
PoliticalRights	Free	-0.247	0.064
CivilLiberties	Free	0.021	0.095
PoliticalRights	Partly free	-1.125	-1.037
CivilLiberties	Partly free	-0.681	-0.561

allows us to investigate whether these misspecifications are capable of changing the substantive conclusion of interest shown in Figure 2, namely that there appears to be a nonlinear relationship between freedom of the press and the level of democracy. Figure 2 shows that an EPC-interest in the “partly free” group of at least 2, and an EPC-interest in the “free” group of at least 1 in absolute value would be required to change the substantive conclusion<sup>2</sup>. Table 4 shows both columns of EPC-interest values that involve a EPC-interest of at least 1 in absolute value. Although the EPC-interest values are much larger than those found for the scalar invariance model with respect to time, none of the EPC-interest values are large enough to change the substantive conclusions of interest. In spite of the obvious model misspecification, therefore, the comparison between groups representing different levels of press freedom does not appear to be threatened by differences in measurement parameters.

## 5. EXAMPLE 2: REGRESSION COEFFICIENTS IN 19 COUNTRIES

While the comparison of latent means requires scalar invariance, the cross-group comparison of regression coefficients among latent variables calls for metric invariance (e.g. Steenkamp and Baumgartner, 1998). A study by Davidov et al. (2008) on the effect of human values on attitudes toward immigration illustrates the application of

<sup>2</sup>Note that this requirement is very similar to the bounds derived by Imai and Yamamoto (2010).

the EPC-interest to this more complex situation. The authors compare 19 European countries on four regression coefficients between latent variables measured by 17 items. The effect of  $19 \times 17 = 323$  partial metric invariance restrictions' potential effects on  $19 \times 4 = 76$  country-specific regression coefficients needs to be assessed:  $323 \times 76 = 24,548$  EPC-interest statistics. From this apparently daunting task, the EPC-interest applied to this example produces a surprisingly simple picture: it turns out to be inconsequentially small in all but four cases.

Figure 3 reproduces Davidov et al. (2008)'s model along with the metric invariance model's cross-country average structural regression coefficients. Interest focuses on the structural regression coefficients representing the effect of the two value dimensions "Self-transcendence" and "Conservation" advanced by Schwartz and Bilsky (1987) on two different dimensions of attitudes toward immigration, "Allow" and "No conditions". In this parametrization of the metric invariance model, the first country's latent variables have been standardized and the other countries' latent variable variances allowed to vary to take heteroskedasticity into account. This parametrization is equivalent to the more common practice of fixing one loading per latent variable to unity in each country and facilitates the interpretation and comparison of the regression coefficients over countries.

Data were obtained from the European Social Survey 2002, a high quality cross-national probability survey (Jowell et al., 2007). Each latent variable is measured with multiple observed indicators. Davidov et al. (2008) compared 19 countries on these regression coefficients: Austria ( $n = 2,257$ ), Belgium (1,899), Czech Republic (1,360), Denmark (1,506), Finland (2,000), France (1,503), Germany (2,919), Great Britain (2,052), Greece (2,566), Hungary (1,685), Ireland (2,046), Netherlands (2,364), Norway (2,036), Poland (2,110), Portugal (1,510), Slovenia (1,519), Spain (1,729), Sweden (1,999), and Switzerland (2,037). For the original data, precise wording of the ques-

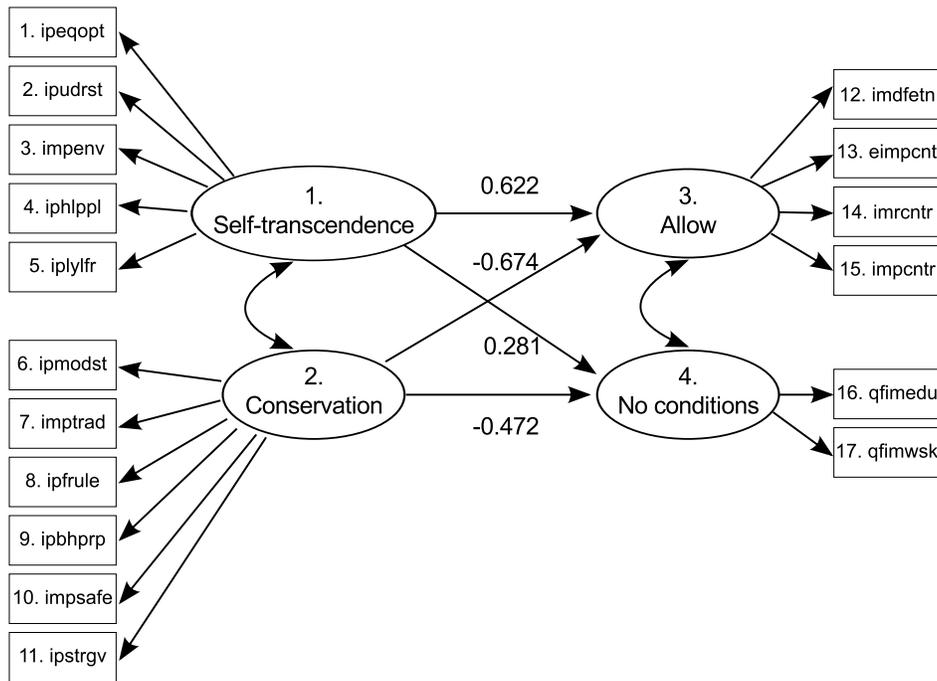


Figure 3: Structural equation model of relationships of interest between four latent variables. The regression coefficient estimates shown are averages over all 19 countries under the full metric invariance model (see Figure 4 for country estimates).

Table 5: Left: descriptive statistics for the observed variables. Overall means over countries, within-country standard deviation, and between-country standard deviation. Right: loading estimates in the full metric invariance model.

		Overall mean	Within-country sd.	Between-country sd.	SelfTransConserv 1.	Allow 2.	NoCond 3.	4.
1.	ipeqopt	2.07	(1.04)	(0.18)	0.648			
2.	ipudrst	2.41	(1.05)	(0.18)	0.711			
3.	impenv	2.13	(0.99)	(0.19)	0.697			
4.	iphlppl	2.31	(0.99)	(0.16)	0.734			
5.	iplylfr	1.97	(0.89)	(0.19)	0.647			
6.	ipmodst	2.83	(1.24)	(0.42)		0.677		
7.	imptrad	2.76	(1.32)	(0.35)		0.768		
8.	ipfrule	3.14	(1.34)	(0.37)		0.847		
9.	ipbhprp	2.70	(1.23)	(0.31)		0.920		
10.	impsafe	2.37	(1.18)	(0.36)		0.801		
11.	ipstrgv	2.43	(1.19)	(0.37)		0.838		
12.	imdfetn	2.54	(0.78)	(0.27)			0.589	
13.	eimpent	2.45	(0.76)	(0.28)			0.612	
14.	imrcntr	2.48	(0.81)	(0.21)			0.517	
15.	impcntr	2.51	(0.77)	(0.28)			0.632	
16.	qfimedu	6.22	(2.64)	(0.64)				1.732
17.	qfimwsk	6.75	(2.65)	(0.78)				2.170

tions, and further information on data collection procedures, we refer to the ESS website<sup>3</sup>. The left-hand side of Table 5 presents descriptive statistics for the 17 observed variables. The within-country and between-country standard deviations call attention to the considerable between-country variation in the means, which could reflect substantive differences between the 19 countries, but could also originate in cross-country loading (or intercept) differences.

Following standard practice, the original authors fit the full metric invariance model to the ESS data: a multiple group structural equation model in which the loadings in

<sup>3</sup><http://ess.nsd.uib.no/ess/round1/>

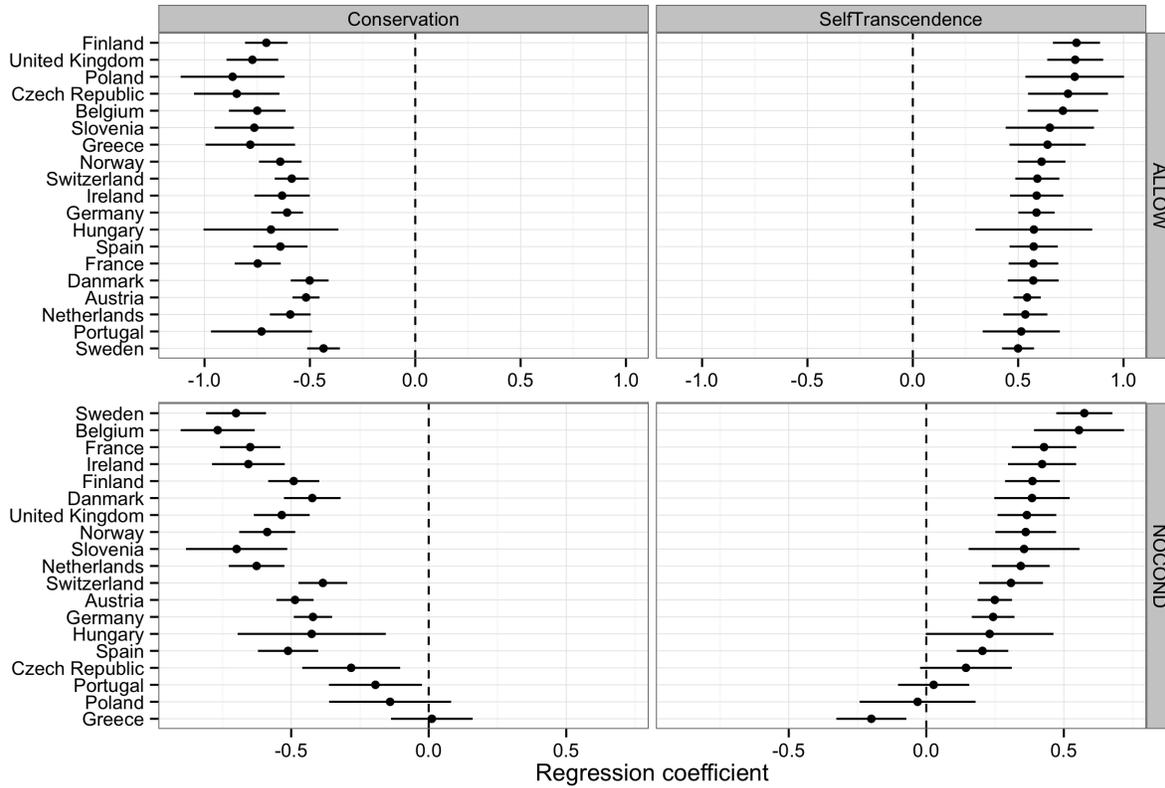


Figure 4: Estimates  $\pm 2$  s.e. for the four regression coefficients between latent variables of interest under the metric invariance model in 19 countries.

Figure 3 are constrained to be equal across the 19 countries, while the structural regression coefficients and other parameters are allowed to vary. The right-hand side of Table 5 gives the resulting estimates of the loadings (under the variance parameterization). Parameter estimates of interest for the 19 countries are shown in Figure 4.

Davidov et al. (2008, 589) test for (partial) metric equivalence by examining overall fit measures as well as MI and SEPC-self. They decide that “the overall fit measures (CFI = 0.95, NFI = 0.94, RMSEA = 0.01, Pclose = 1.0) suggest that [the full metric invariance model] is acceptable. However, modification indices pointed to misspecifications in the model. Therefore, in model 2 we (...) had to relax the measurement invariance constraints for some items.” Based on the MI and (S)EPC-self, the author’s final model frees four out of the possible 323 loading equalities in three different

Table 6: EPC-interest statistics of at least 0.1 in absolute value with respect to the latent variable regression coefficients. For reference, the standardized expected parameter change (SEPC-self) is also given.

	Loading "Conditions → qfimwsk" in...			
	Slovenia	France	Hungary	Ireland
<i>EPC-interest with respect to:</i>				
Conditions → SelfTranscendence	-0.073	-0.092	-0.067	0.073
Conditions → Conservation	0.144	0.139	0.123	-0.113
SEPC-self	0.610	0.692	0.759	-0.514

countries, namely "Conservation → ipmodst" in Portugal, "Conservation → ipstrgv" in Ireland, and "SelfTranscendence → iplylfr" in Portugal and Denmark<sup>4</sup>.

The EPC-interest statistic provides an alternative way to assess whether a loading equality restriction is substantively important or not. Saris, Satorra and Van der Veld (2009) suggested a cutoff of 0.1 in absolute value for correlations and standardized regression coefficients. Given the parameterization used, this appears to be a reasonable criterion for the EPC-interest with respect to regression coefficients as well. Although there are potentially 76 affected parameters of interest for each of the 323 equality restrictions, as it happens only the four EPC-interest statistics shown in Table 6 meet this criterion. The form of the model plays an important role here: equality constraints on loadings in one country hardly affect structural parameters in another, misspecified constraints on one dependent latent variables' loadings hardly affect the regression coefficients of the other, and latent variables with many indicators are generally less affected by misspecifications.

In summary, even though the full metric invariance model shows "close fit" in terms of CFI and RMSEA, this model still contains misspecified equality restrictions on loadings that threaten the comparison of the regression coefficients of interest. Freeing these loadings takes this possibility into account while still allowing for the identification of

<sup>4</sup>Davidov (2012, personal communication).

the parameters of interest. The misspecified restrictions on loadings in question were not detected with the SEPC-self and MI: these measures detected other misspecifications that were large in the sense of having strongly differing loadings.

## 6. DISCUSSION AND CONCLUSION

We explored the use of the EPC-interest as a measure of whether misspecifications in invariance constraints are substantively relevant. The EPC-interest approximates the change in parameters of interest if the equality-constrained measurement parameters were freed. The EPC-interest follows more closely the logic that invariance testing matters because of its consequences for the substantive comparisons of interest. Another advantage of the EPC-interest is that its calculation requires only the results obtained by fitting the invariance model. Although we discussed the EPC-interest in the context of structural equation modeling, Equation 5 is more general, and easily extended to other latent variable models - including those for discrete observed and latent variables (Bartholomew, Knott and Moustaki, 2011).

The EPC-interest is an approximation to the actual change in parameters of interest. A robustness study investigated the influence of alternative model misspecification on population bias in the EPC-interest, finding that it was generally small. A Monte Carlo simulation then evaluated the finite sample performance of the EPC-interest. Estimates of the EPC-interest were close to the population shift in parameters of interest, even under the condition with sample sizes of 50 observations per group and a strong misspecification in the null model. The EPC-interest statistic therefore appears to have reasonable asymptotic and finite-sample properties.

Two examples of the use of the EPC-interest in cross-group latent variable applications from the literature demonstrated its use. The first example analyzed a scalar

invariance model in which the goal was to compare latent means across groups, while the second example's goal of comparing structural regression coefficients over 19 countries compelled its authors to appraise metric invariance.

The first example compared countries' levels of democracy across time and levels of press freedom. Applying the EPC-interest to the model with scalar invariance (both intercept and loading) restrictions across time led to much the same conclusion as drawn by invariance testing procedures as well as the original authors. The EPC-interest was then applied to scalar invariance restrictions across levels of press freedom. Here measurement invariance testing indicated measurement inequivalence; based on the EPC-interest, this lack of invariance does not appear to threaten the substantive conclusions regarding the relationship between press freedom and levels of democracy, however.

In the second example, the metric invariance (loading) equalities that are in principle required for comparison of regression coefficients fit the data according to the "close fit" measures CFI and RMSEA. In spite of this adequate overall fit, the EPC-interest indicated for four (out of 323 possible) equality restrictions that freeing these loadings would substantially change the coefficients in the latent variable regressions of interest. This is possible because the "closeness" of the fit in measures such as CFI and RMSEA is not defined in terms of changes in the parameters of interest. These four influential restrictions were not the largest in terms of SEPC and MI, but, due to the form of the model and the values of the parameters, did have the largest influence on parameters of interest. This example therefore demonstrates that large misspecifications in invariance restrictions (SEPC) need not correspond to substantively large changes in parameters of interest. Conversely, smaller changes in loading estimates may produce substantially relevant changes in the parameters of interest.

R code (R Core Team, 2012) implementing the EPC-interest for the SEM library

`lavaan` (Rosseel, 2012) and allowing for reproduction of the examples discussed in this paper is available online at [http://\[author's homepage\]/](http://[author's homepage]/).

A disadvantage of the univariate EPC-interest measure used is that the measurement parameters in latent variable models are often dependent, meaning that the EPC-interest suffers from similar problems of stepwise model improvement and possible capitalization on chance as the EPC-self and MI (MacCallum, Roznowski and Necowitz, 1992). Though this problem is not specific to the EPC-interest, it remains a concern. A possible extended method is to evaluate the effect of freeing several restrictions at a time, i.e., a multivariate version of the EPC-interest. It is not known, however, to what degree such a statistic continues to perform well both asymptotically and in finite samples. More work is therefore needed to evaluate the performance of the EPC-interest, EPC-self, MI statistics, and possibly their multivariate versions for the particular purpose of partial invariance testing.

Interesting avenues for future study include extensions to Bayesian analysis (Fox and Glas, 2005; Muthén and Asparouhov, 2012), and to other types of conclusions of interest such as ranking groups; the investigation of configural invariance (cross-loadings and error covariances); and an investigation of the robustness of the EPC-interest to misspecification of the alternative model under a wider set of conditions than could be studied here. These issues fall beyond the scope of this paper.

In the sensitivity analysis approach, partial measurement invariance is no longer a property of a measurement model, but becomes a property holding only with respect to a particular analysis of interest. Use of the EPC-interest for measurement invariance evaluation requires more substantive knowledge and input from the researcher. Whether this is a drawback or an advantage is up for debate: arguably, invariance is a requirement in comparative research because of its substantive implications, so sub-

stantive implications should guide decisions on whether or not equality restrictions are warranted. As demonstrated by the second example, invariance tests do not guarantee that “closely fitting” but misspecified restrictions are inconsequential for parameters of interest, nor that “badly fitting” equality restrictions are substantively relevant. The EPC-interest thus represents a trade-off between higher analysis complexity and a greater degree of confidence that the groups are comparable for the purposes at hand.

## REFERENCES

- Armstrong, D.A. 2011. “Stability and change in the Freedom House political rights and civil liberties measures.” *Journal of Peace Research* 48:653–662.
- Bartholomew, D.J., M. Knott and I. Moustaki. 2011. *Latent variable models and factor analysis: a unified approach*. New York: John Wiley & Sons.
- Bentler, PM and C.P. Chou. 1992. “Some new covariance structure model improvement statistics.” *Sociological Methods & Research* 21:259–282.
- Bentler, P.M. and T. Dijkstra. 1984. “Efficient estimation via linearization in structural models.” In *Multivariate Analysis VI*, ed. P.R. Krishnaiah. Amsterdam: North-Holland pp. 9–42.
- Bollen, Kenneth. 1989. *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- Byrne, B.M., R.J. Shavelson and B. Muthén. 1989. “Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance.” *Psychological Bulletin* 105:456.

- Carroll, R.J., D. Ruppert, L.A. Stefanski and C.M. Crainiceanu. 2006. *Measurement error in nonlinear models: a modern perspective*. Vol. 105 Boca Raton, FL: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- Chen, F.F. 2007. "Sensitivity of goodness of fit indexes to lack of measurement invariance." *Structural Equation Modeling* 14:464–504.
- Cheung, G.W. and R.B. Rensvold. 2002. "Evaluating goodness-of-fit indexes for testing measurement invariance." *Structural Equation Modeling* 9:233–255.
- Chou, C.P. and PM Bentler. 1993. "Invariant standardized estimated parameter change for model modification in covariance structure analysis." *Multivariate behavioral research* 28:97–110.
- Clinton, J., S. Jackman and D. Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98:355–370.
- Davidov, E. 2009. "Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective." *Political Analysis* 17:64–82.
- Davidov, E., B. Meuleman, J. Billiet and P. Schmidt. 2008. "Values and support for immigration: a cross-country comparison." *European Sociological Review* 24:583–599.
- Fox, Jean-Paul and Cees AW Glas. 2005. "Bayesian modification indices for IRT models." *Statistica Neerlandica* 59:95–106.
- French, B.F. and W.H. Finch. 2006. "Confirmatory factor analytic procedures for the determination of measurement invariance." *Structural Equation Modeling* 13:378–402.

- Hancock, Gregory R., Laura M. Stapleton and Ilona Arnold-Berkovits. 2009. "The tenuousness of invariance tests within multisample covariance and mean structure models." In *Structural Equation Modeling in Educational Research: Concepts and Applications*, ed. T. Teo and M.S. Khine. Sense Publishers pp. 137–174.
- Hausman, J.A. 1978. "Specification tests in econometrics." *Econometrica: Journal of the Econometric Society*.
- Hu, L. and P.M. Bentler. 1998. "Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification." *Psychological methods* 3:424.
- Imai, K. and T. Yamamoto. 2010. "Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis." *American Journal of Political Science* 54:543–560.
- Jackman, S. 2001. "Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference, and model checking." *Political Analysis* 9:227–241.
- Jowell, Roger, Caroline Roberts, Rory Fitzgerald and Gillian Eva. 2007. *Measuring Attitudes Cross-nationally: Lessons from the European Social Survey*. SAGE.
- Kaplan, D. 1989. "Model modification in covariance structure analysis: Application of the expected parameter change statistic." *Multivariate Behavioral Research* 24:285–305.
- Kolenikov, S. 2009. "Biases of parameter estimates in misspecified structural equation models." *Sociological Methodology*.
- Lord, F. M. and M. R. Novick. 1968. "Statistical theories of mental scores." *Reading, Addison-Wesley*.

- MacCallum, R.C., M. Roznowski and L.B. Necowitz. 1992. "Model modifications in covariance structure analysis: the problem of capitalization on chance." *Psychological Bulletin*; *Psychological Bulletin* 111:490.
- Magnus, J.R. and A.L. Vasnev. 2007. "Local sensitivity and diagnostic tests." *The Econometrics Journal* 10:166–192.
- Mellenbergh, G.J. 1989. "Item bias and item response theory." *International Journal of Educational Research* 13:127–143.
- Meredith, W. 1993. "Measurement invariance, factor analysis and factorial invariance." *Psychometrika* 58:525–543.
- Meuleman, B. 2012. "When are item intercept differences substantively relevant in measurement invariance testing?" *Methods, Theories, and Empirical Applications in the Social Sciences*.
- Millsap, R.E. 1997. "Invariance in measurement and prediction: Their relationship in the single-factor case." *Psychological Methods* 2:248.
- Millsap, R.E. 2007. "Invariance in measurement and prediction revisited." *Psychometrika* 72:461–473.
- Millsap, R.E. and H.T. Everson. 1993. "Methodology review: Statistical approaches for assessing measurement bias." *Applied Psychological Measurement* 17:297–334.
- Millsap, R.E. and O.M. Kwok. 2004. "Evaluating the impact of partial factorial invariance on selection in two populations." *Psychological Methods* 9:93.
- Muthén, B. and T. Asparouhov. 2012. "Bayesian SEM: A more flexible representation of substantive theory." *Psychological Methods*.

- Neudecker, H. and A. Satorra. 1991. "Linear Structural Relations: Gradient and Hessian of the Fitting Function." *Statistics and Probability Letters* 11:57–61.
- Oberski, D.L. 2012. "Comparability of Survey Measurements." In *Handbook of Survey Methodology for the Social Sciences*, ed. Lior Gideon. New York: Springer-Verlag pp. 477–498.
- R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org/>
- Rosseel, Y. 2012. "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48:1–36.
- Saris, W.E., A. Satorra and D. Sörbom. 1987. "The Detection and Correction of Specification Errors in Structural Equation Models." *Sociological Methodology* 17:105–129.
- Saris, W.E., A. Satorra and W.M. Van der Veld. 2009. "Testing structural equation models or detection of misspecifications?" *Structural Equation Modeling* 16:561–582.
- Satorra, A. 1989. "Alternative Test Criteria in Covariance Structure Analysis: A Unified Approach." *Psychometrika* 54:131–151.
- Schmitt, N. and G. Kuljanin. 2008. "Measurement invariance: Review of practice and implications." *Human Resource Management Review* 18:210–222.
- Schwartz, S.H. and W. Bilsky. 1987. "Toward a universal psychological structure of human values." *Journal of personality and social psychology* 53:550.
- Steenkamp, JBEM and H. Baumgartner. 1998. "Assessing measurement invariance in cross-national consumer research." *Journal of consumer research* 25:78–107.

- Treier, S. and S. Jackman. 2008. “Democracy as a latent variable.” *American Journal of Political Science* 52:201–217.
- Vandenberg, R.J. and C.E. Lance. 2000. “A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research.” *Organizational research methods* 3:4–70.
- Yoon, M. and R.E. Millsap. 2007. “Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study.” *Structural Equation Modeling* 14:435–463.
- Yuan, K.H., L.L. Marshall and P.M. Bentler. 2003. “Assessing the effect of model misspecifications on parameter estimates in structural equation models.” *Sociological methodology* 33:241–265.

#### A. APPENDIX: DERIVATION OF THE EPC-INTEREST

Assume both the restricted and alternative model are identified, i.e.  $\mathbf{A}'_a \mathbf{J}(\hat{\boldsymbol{\theta}}) \mathbf{A}_a$  and  $\mathbf{A}'_0 \mathbf{J}(\hat{\boldsymbol{\theta}}) \mathbf{A}_0$  are full-rank, and that standard regularity conditions apply (see Satorra, 1989, assumptions 1 through 6\* inclusive). By a Taylor expansion at the restricted solution,

$$\begin{aligned}
 F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) &\approx \hat{F} + [\mathbf{a}_0(\boldsymbol{\theta}) - \mathbf{a}_0(\hat{\boldsymbol{\theta}})]' \mathbf{g}(\hat{\boldsymbol{\theta}}) + \frac{1}{2} [\mathbf{a}_0(\boldsymbol{\theta}) - \mathbf{a}_0(\hat{\boldsymbol{\theta}})]' \mathbf{J}(\hat{\boldsymbol{\theta}}) [\mathbf{a}_0(\boldsymbol{\theta}) - \mathbf{a}_0(\hat{\boldsymbol{\theta}})] \quad (6) \\
 &= \hat{F} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{g}(\hat{\boldsymbol{\theta}}) \mathbf{A}_0 + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{A}'_0 \mathbf{J}(\hat{\boldsymbol{\theta}}) \mathbf{A}_0 (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \quad (7)
 \end{aligned}$$

because the likelihood depends only on the free parameters. Now, to investigate what would happen if a set of restrictions in  $\mathbf{a}_0$  were freed, we find the solution to the

equation  $\partial F/\partial \mathbf{a}_a = 0$ , that is,

$$\mathbf{g}(\hat{\boldsymbol{\theta}})\mathbf{A}_a + \mathbf{A}'_a\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{A}_a(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = 0.$$

Therefore,

$$\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \approx -(\mathbf{A}'_a\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{A}_a)^{-1}\mathbf{g}(\hat{\boldsymbol{\theta}})\mathbf{A}_a, \quad (8)$$

and, since the parameters of interest are defined as  $\boldsymbol{\pi} := \mathbf{P}\boldsymbol{\theta}$ ,

$$\text{EPC-interest} := \boldsymbol{\pi} - \hat{\boldsymbol{\pi}} \approx -\mathbf{P}(\mathbf{A}'_a\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{A}_a)^{-1}\mathbf{g}(\hat{\boldsymbol{\theta}})\mathbf{A}_a. \quad (9)$$

A test can also be constructed of the null hypothesis that the total change in the parameter(s) of interest equals zero. From equation 9, under the null hypothesis the EPC-interest will have variance

$$\text{avar}(\text{EPC-interest}) = \mathbf{P}(\mathbf{A}'_a\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{A}_a)^{-1}\mathbf{P}'$$

(Satorra, 1989, eq. 26, p. 143). Under the null hypothesis, then, the statistic

$$T_a := (\text{EPC-interest})'\text{avar}(\text{EPC-interest})^{-1}(\text{EPC-interest})$$

will be distributed as central chi-square with  $\text{rk}(\mathbf{P})$  degrees of freedom.

As also noted by Yuan, Marshall and Bentler (2003, 253), the  $T_a$  statistic is a generalization of the test developed by Hausman (1978). Indeed, when the hypothesis of invariance holds together with the likelihood specification, the alternative model estimates are consistent but inefficient, while the restricted model estimates will be consistent and most efficient. Lemma 2.1 of Hausman (1978, 1253) will then be directly applicable to  $\hat{\boldsymbol{\theta}}_0$  and  $\hat{\boldsymbol{\theta}}_a$ . Thus, the  $T_a$  and EPC-interest statistics fall under the general

principle of sensitivity analysis in econometrics.

The power of the  $T_a$  test can also be estimated, as follows from Satorra (1989, theorem 5.3) and an extension of the argument in Saris, Satorra and Van der Veld (2009, 570-1). Letting the hypothetical vector of minimum differences in the parameters of interest to be detected equal  $\tilde{\boldsymbol{\delta}}$ , the noncentrality parameter (ncp) then equals

$$\text{ncp} = \tilde{\boldsymbol{\delta}}' \text{avar}(\text{EPC-interest})^{-1} \tilde{\boldsymbol{\delta}}, \quad (10)$$

and the power of the  $T_a$  test to detect combined differences in the parameters of interest as large as or larger than  $\tilde{\boldsymbol{\delta}}$  can be calculated by referring to the noncentral chisquare distribution:

$$\Pr(\chi^2(\text{rk}(P), \text{ncp}) > c_\alpha), \quad (11)$$

where  $c_\alpha$  is the critical value of the central chi-square distribution with  $\text{rk}(P)$  degrees of freedom corresponding to a given  $\alpha$ . For example, for  $\alpha = 0.05$ ,  $c_\alpha \approx 3.84$  for a one-degree of freedom test. This result allows one to investigate whether the procedure suggested here provides enough power to detect differences in the parameters of interest.

The statistic EPC-interest can also be used to obtain an estimate of the values the model parameters would take on if the restriction under investigation were freed. Without fitting the alternative model, the estimate of the parameters of interest under the alternative model can be approximated by

$$\hat{\boldsymbol{\pi}}_a = \hat{\boldsymbol{\pi}} + \text{EPC-interest}$$

(Bentler and Dijkstra, 1984). These “new” parameter estimates can also be obtained for the other parameters of the model.