

How to Compare Means of Latent Variables across Countries and Waves: Testing for Invariance Measurement. An Application Using Eastern European Societies¹

Mircea Comșa²

Department of Sociology and Social Work, Babeș-Bolyai University of Cluj-Napoca, Romania

How to Compare Means of Latent Variables across Countries and Waves: Testing for Invariance Measurement. An Application Using Eastern European Societies.

Comparative social survey research is exponentially growing, more and more data being available at longitudinal and cross-national levels. When unobservable concepts are involved, the analysis of this kind of data poses the issue of measurement invariance (Are the concepts equivalent across countries and longitudinally?). The aim of this paper is to draw a clear picture of measurement invariance (MI) test via multiple-group confirmatory factor analysis (MG-CFA) using AMOS 16 software. According to the literature this is a necessary step before latent mean comparison. First, I discuss the concept of measurement invariance, and then we present the most important features and issues related to the MG-CFA. The main part of the paper is dedicated to presenting a step by step procedure to test for MI using AMOS VB.NET. In order to facilitate understanding, all steps are logically described; also, data and syntaxes are provided. To illustrate the procedure, I analyze the "sexual permissiveness", as it measured in two large-scale surveys (EVS and WVS). I focus on the case of those 19 former communist countries included in these surveys, for which the available measurement covers at least three points in time, during the 1990-2009 period. To measure the "sexual permissiveness", four items are used: how justifiable are homosexuality, prostitution, abortion, and divorce. Scalar invariance of "sexual permissiveness" scale is proved for most countries, so means and relationships involving this value can be meaningfully interpreted. Finally, we discuss some alternatives for presenting the means of latent variables.

Sociológia 2010, Vol. 42 (No. 6: 639-669)

Key words: measurement invariance; multiple-group confirmatory factor analysis; sexual permissiveness scale; structural equation modelling; AMOS

Introduction

Comparative social survey research is exponentially growing, more and more data being available at longitudinal and cross-national levels. When unobservable concepts are involved, the analysis of this kind of data poses the issue of measurement invariance. In other words, if one wants to compare relationships or means of latent variables, one has to assure that those constructs (and the subsequent indicators) have same meaning across countries and over time. This paper illustrates how group differences related to latent

¹ This article was supported in part by CNCISIS-UEFISCSU grants PNII-IDEI 2174/2008 "Romanian Presidential Study 2009 – Romania, 2009", grant director Mircea Comșa, and ID56/2007. I am indebted to and wish to thank to Cosima Rughiniș, Bogdan Voicu and Mălina Voicu for their invaluable suggestions and comments.

² Correspondence address: Mircea Comșa, Department of Sociology and Social Work, "Babeș-Bolyai" University of Cluj-Napoca, B-dul 21 Dec. 1989 No. 128, Cluj-Napoca 400604, Romania (mcomsa@socasis.ubbcluj.ro).

variables (“sexual permissiveness” in our case) can be estimated. Those interested in comparative survey research could find some valuable suggestions and results in our analyses.

Before any empirical values comparisons, two methodological questions should be asked: “Are the values comparable across the full set of formerly communist countries?” and “Are the values comparable over time in each country separately?”. To answer to these questions, one needs to test for measurement equivalence or invariance of factors (latent variables), both across countries and longitudinally (Reise – Widaman – Pugh 1993; Riordan – Vanderberg 1994; Van de Vijver – Leung 1997; Widaman – Reise 1997; Steenkamp – Baumgartner 1998; Byrne – Watkins 2003; Boorsboom 2006; Vandenberg – Lance 2000; Davidov 2008; Schmitt – Kuljanin 2008).

Two major steps are necessary for establishing measurement invariance over time or across countries (Billiet 2003). First step, the data collection stage, requires comparable item translations into different languages, comparable response sets, comparable sampling procedures, and similar data collection techniques. The second step, part of the data analysis stage, consists in testing for measurement invariance of the theoretical constructs. In most cases, cross-national surveys pay considerable attention to the first set of conditions, and most of the research steps are standardized and closely supervised by specialized international teams (see for example the European Value Study – EVS or the European Social Survey – ESS). As concerns testing for measurement invariance, in recent years there has been a growing body of empirical research that relies on it (Vandenberg – Lance 2000; Vandenberg 2002; Schmitt – Kuljanin 2008). However, it received less attention, and this is the point where this paper tries to contribute to the existing literature.

There are several labels which stand for the concept that I present. “Functional equivalence”, “invariance of measures across groups”, or “measurement invariance” (MI) are used with the same meaning. In this paper I opt for the shorter MI. Irrespectively which label is used, the concept refer to “whether or not, under different conditions of observing and studying a phenomenon, measurement operations yield measures of the same attribute” (Horn – McArdle 1992: 117). One can say that a measure is invariant when “members of different populations who have the same standing on the construct being measured receive the same observed score on the test” (Schmitt – Kuljanin 2008).

Why is important to test for MI? Survey results (means, correlations, factor loadings, etc.) across different groups (same group across time or different groups at the same time) can be compared only if the measures used have the same meanings (Reise – Widaman – Pugh 1993; Riordan – Vanderberg 1994; Van de Vijver – Leung 1997; Widaman – Reise 1997; Steenkamp –

Baumgartner 1998; Vandenberg – Lance 2000; Byrne – Watkins 2003; Boorsboom 2006; Saris – Gallhofer 2007; Davidov 2008; Schmitt – Kuljanin 2008). In the absence of MI, comparisons are problematic and conclusions could be ambiguous or even biased. In other terms, if one finds that a statistic is different between two groups, one cannot say that the observed difference is a true difference in the underlying construct or that is due to systematic biases in the way people from different groups respond to certain items (Steenkamp – Baumgartner 1998). Some of the consequences of making inappropriate cross-cultural comparisons (when measurement invariance is not satisfactory) are illustrated in an empirical article by Chen (2008). Chen discusses about importing a scale developed in a certain reference group (let say that it offers a good measurement for the United States) to a focal group, on which the researcher wants to focus (in Chen's examples the focal group is China). Chen concludes that “when the reference group has higher [factor] loadings or intercepts, the mean is overestimated in that group but underestimated in the focal group, resulting in a pseudo group difference” (1005). With other words, there would be not such a big difference (or even any difference) between the two groups, if the scale would have offered a similar reliability in both groups.

The purpose of this paper is to illustrate the measurement invariance (MI) test using multiple-group confirmatory analysis (MGCFA) and AMOS 16 software. According to the literature testing for measurement invariance is a necessary step before latent mean comparison. First, I discuss the concept of measurement invariance, and then we present the most important features and issues related to the MGCFA. The main part of the paper is dedicated to presenting a step by step procedure to test for MI using AMOS VB.NET. In this part all steps are logically described and for each of them data and syntaxes used are provided. To illustrate the procedure, I analyze the “sexual permissiveness”, as it measured in two large-scale surveys (EVS and WVS). I focus on the case of those 19 former communist countries included in these surveys, for which the available measurement covers at least three points in time, during the 1990-2009 period. Finally, we discuss some alternatives for presenting the means of latent variables.

Literature review: the MGCFA approach

Testing for MI can be done by at least three techniques, namely: item response theory (De Beuckelaer 2005; De Jong – Steenkamp – Fox 2007), multiple group confirmatory factor analysis (Stark – Chernyshenko – Drasgow 2006; Davidov 2008) and testing for cognitive equivalence (Saris – Gallhofer 2007). Among available techniques, multiple group confirmatory factor analysis (MGCFA) appears to be the most used tool in the study of measurement invariance across multiple populations (Jöreskog 1971; Byrne – Shavelson –

Muthén 1989; Meredith 1993; Millsap – Everson 1993; Reise – Widaman – Pugh 1993; Widaman – Reise 1997; Steenkamp – Baumgartner 1998; Vandenberg – Lance 2000; Vandenberg 2002; Byrne 2004). Consequently, for our analysis we have preferred MGCFAs.

According to the measurement invariance literature, the comparison of latent means³ across cultures / over time and their meaningful interpretation requires that three levels of invariance are fulfilled (Horn – McArdle 1992; Meredith 1993; Steenkamp – Baumgartner 1998; Byrne 2004; Brown 2006; Saris – Gallhofer 2007: 334; Schmitt – Kuljanin 2008; Davidov 2008). The first condition, *configural invariance*, requires that the model holds for all the groups involved or, in other terms, that the items should exhibit the same configuration of salient and nonsalient factor loadings across different groups. The second condition, *metric invariance*, requires that, besides configural invariance, the slopes (factor loadings) are the same (invariant) in all the groups studied.⁴ The third condition, *scalar invariance*, requires that, besides metric invariance, the intercepts are the same across all groups (invariant).⁵ If scalar invariance is supported, one can confidently carry out mean comparisons. In SEM (structural equation modelling) literature, the three conditions mentioned above appear under different names, which could be confusing sometimes (Table 1)

Table 1: Measurement invariance related concepts names in different notations

Invariance Testing Notation	AMOS Notation	Bollen (1989) Notation
configural invariance	baseline model	form
metric (weak) invariance	equal measurement weights	equal factor loadings
scalar (strong) invariance	equal measurement intercepts	equal item intercepts

In conclusion, when comparing countries, and using factor scores, one should check first if the measurement model meets the three above conditions. This means that separated analyses should be carried out for each country, than for all countries together, and to show out that the scale items similarly load, in terms of significance, on the respective latent variable (configural invariance), they have similar factor loadings (metric invariance) and the intercepts for any item are the same in all considered countries (scalar invariance).

What if the test of equal slopes or intercepts shows that the parameters are not invariant for some items? In this case we have three options: to exclude groups (the non-invariant ones), to remove items (the non-invariant ones), or to

³ The comparison of relationships (correlations, regression coefficients) requires only configural and metric invariance.

⁴ $\lambda_1 = \lambda_2 = \dots = \lambda_g$, where λ stands for the factor loading for an indicator and g represents the group.

⁵ $\tau_1 = \tau_2 = \dots = \tau_g$, where τ stands for intercept of one indicator and g represents the group.

test for partial factorial invariance. If none of these options solve the problem, one has to conclude that the measured construct has different meaning / measurement unit / scale origin across groups and rely on a weaker form of invariance. Excluding groups does not really solve the problem; it only restrains the universe of interest. Removing items⁶ could be helpful only if the scale has more than three items and if the noninvariance does not affect all the items. In our case (four indicators per factor and research interest in all the selected countries), we preferred to test for partial invariance, a concept introduced by Byrne, Shavelson and Muthén (1989). They argued that full metric invariance was not necessary in order for further tests of invariance and substantive analyses (comparisons of factor means in our case) to be meaningful. Their study shows that such comparisons can be done if at least one item, other than the one fixed at unity to define the scale of each latent construct, is metrically invariant (see also Meredith 1993; Steenkamp – Baumgartner 1998; Rensvold – Cheung 1998). If partial metric invariance is supported, partial scalar invariance can be tested. This can be done by freeing the intercepts of those items that are not held to be metrically invariant across groups, while the intercepts of the other items are held invariant, at least initially. If the initial model of partial scalar invariance is rejected, MIs (modification indices) and EPCs (expected parameter change indices) can be used to locate intercepts that are not cross-groups invariant. In subsequent models, the invariance constraints on these intercepts are relaxed.

In conclusion, if the goal is to conduct comparisons of means across groups, we need metric and scalar invariance for at least two items per construct or per dimension if the construct is multidimensional (Byrne et al. 1989; Meredith 1993; Steenkamp – Baumgartner 1998). If this is not the case, comparing means of different groups is meaningless since the measurement scales are different across groups. Without (partial) scalar invariance, the presence or absence of mean differences might be due to real differences in the value of the latent construct, but also to additive bias and / or different scale metrics (Steenkamp – Baumgartner 1998). In this context, one limit of MGCFA should be mentioned. Saris and Gallhofer (2007: chapter 16) show that the source of the response differences could be the result of differences in measurement errors, in cognitive processes, or substantive differences between countries or a combination thereof. They argue that sometimes, testing for MI via MGCFA approach can lead “to a rejection of the model, while the cause of the problem is not that the indicators have different interpretations across countries, but that instead the respondents in the different countries employ the scales

⁶ Other possibilities are to treat the missing indicators for a country and wave as latent variables (McArdle 1994) or to allow different path diagrams for different groups (Byrne 2001; 2004; Arbuckle 2007). Because our models contain few measured variables, using these options could severely modify means estimations for countries and waves with missing indicators.

distinctively” (338). The conclusion is obvious: the standard approach to test for equivalence is too restrictive. The alternative proposed is the cognitive equivalence of measurement instrument (invariance after correcting for differences in the measurement process using data from MTMM⁷ experiments or predictions from SQP⁸).

Before comparing the means of a construct with (partial) scalar invariance across groups, one should consider several important issues. One of them is the choice of constraints for model identification across groups. In most studies, one of the two methods for identifying the factor model in single group studies is used: either fixing the variance of each factor to 1, or fixing a loading for each factor to 1.

However, in the case of MGCFA, setting the variance of a factor to one in each group is not recommended because “the resulting implicit standardization would alter the loadings in each group, and would do so in ways that differ across groups if the factor variances differ. As a result, factor pattern matrices that are truly invariant would not appear to be so because of these standardizations” (Yoon – Millsap 2007). Consequently, for model identification, the best method is to fix to 1, in each group, one of the loadings (the same indicator for all the groups). By doing this, we also set the measurement unit for the latent variable (Jöreskog – Sörbom 1996). The problem is that the use of a fixed loading for identification purposes can interfere with the process of finding the correct model of partial invariance (Cheung – Rensvold 1999). Remember that for model identification one of the indicators for each factor should be set to one.

The question is which of the indicators should be chosen. Sometimes the selection of loadings that should be held invariant is considered somewhat arbitrary (Steiger 2002). Due to the fact that different constraint choices can lead to quite different model fit results (Millsap 2001; Steiger 2002), the selection of which loading to constrain might be more important than would appear at first sight (French – Finch 2008). If the chosen reference item is actually one with highly different functioning across groups two problems could appear: first, the researcher may not detect this noninvariance because the unstandardized factor loadings of the marker indicator are fixed to one in all groups; second, the forced fixed loading can distort the pattern of invariant and noninvariant loadings and can lead to choice of the wrong model (Brown 2006: 302; Yoon – Millsap 2007). Because the reference variable is fixed to one across groups, it should be invariant. The problem is that invariance cannot be established without estimating a model, and model estimation requires an

⁷ Multitrait-Multimethod.

⁸ Survey Quality Prediction (<http://www.sqp.nl>).

invariant referent, which brings the process back to the original invariant referent assumption (French – Finch 2008).

To solve this circular situation, several methods were proposed. Cheung and Rensvold (1999) and Rensvold and Cheung (1998, 2001) have proposed the factor-ratio test and the stepwise partitioning procedure in combination. In this line of reasoning, if the number of indicators is small (our case), the easiest approach is to re-run the MGCFA with different marker indicators (Vandenberg 2002; Brown 2006: 302). Consequently, for our analyses we have preferred this approach.

Another possible problem one could face is to miss some of the potential items of interest for some groups. One could solve this problem in two ways: (1) by excluding the countries with missing items for a wave (thus reducing the number of countries) or (2) by measuring the value of interest using only the available items (thus reducing the number of indicators). If one has only three items per latent variable, one cannot further reduce the number of items (for a simple, one-factorial model). This is due to the fact that for model specification we need at least three indicators.⁹ Besides this methodological constraint, measuring a latent variable by few indicators poses the problem of content validity.

Sometimes, missing data can be another problem. The possible solutions to this issue are: listwise deletion, pairwise deletion, simple imputation (mean or conditional mean), hot deck imputation, expectation maximization imputation, full information maximum likelihood (FIML) imputation, and multiple imputation (Graham – Hofer 2000; Schafer – Graham 2002; Tufiş 2008). According to the missing data literature, FIML and multiple imputations are to be preferred. If, after listwise deletion, we lose less than 5 % of cases, listwise deletion may be a better option. A short list of reasons are: listwise deletion is easier to apply; pairwise is using different samples and it has more chances to produce a not positive definite matrix; FIML is available in AMOS, but only with raw data.

Before proceeding to the next step some further comments about model comparison are necessary. The successive steps presented before (configural, metric, and scalar) produce three different (nested¹⁰ within one another) models. Other things equal, a model with less parameters and a better fit is to be preferred. In order to see if the fit of models is different we have to compare

⁹ A necessary condition, but not sufficient, for a simple (one unmeasured variable) model identification is to have at least as many correlations as unknown parameters, consequently at least three measured variables (Saris – Gallhofer 2007: 202). If only two indicators per unmeasured variable are available, combining two or more simple models and adding some constraints could be an alternative in order to achieve identification.

¹⁰ A model A is nested within model B when model A is a special case of model B. In our case, scalar model is nested in the metric model, and metric model is nested in the configural model. In other words, the models are the same, but configural model is less restrictive than metric model (the factor loadings are set to be equal), and metric model is less restrictive than scalar model (the intercepts are set to be equal).

their fit.¹¹ Classical approaches for model comparison are based on chi-square difference test (Bentler 1990). Because of its sensitivity to sample size (Brown 2006), differences in chi-square between two models with large sample size (1000 or more) are almost every time significant. Instead of χ^2 difference test ($\Delta\chi^2$ test) some researchers (Cheung – Rensvold 2002; Chen 2007) have proposed the use of goodness-of-fit indexes (Δ goodness-of-fit indexes) for assessing multi-sample model invariance. Based on simulation studies, they also proposed some cut-off values for differences. In general, an increase of RMSEA by 0.02 and a decrease of CFI by 0.02 would indicate non-invariance of factor loadings / intercepts. But, because Δ goodness-of-fit indexes were designed to assess model fit in terms of covariance structure (Chen – Sousa – West 2005), it is not clear how they perform when mean structure invariance (as in our case) is the research focus (Fan – Sivo 2009). Based on a Monte Carlo simulation experiment, Fan and Sivo (2009) argued that, in general, Δ goodness-of-fit indexes (like Δ RMSEA, Δ CFI) are so sensitive to model size (number of factor and indicators) that they are not useful in mean structure invariance analysis.

After establishing (partial) scalar invariance across countries and time, we can test for latent factor means differences via multiple-group confirmatory factor analysis (MGCFA) with mean structures. (MG)CFA provides a stronger framework than traditional methods that do not account for response bias and random or non-random measurement errors (e.g. ANOVA, MANOVA or ordinary least squares approaches such as correlation / multiple regression assume variables in the analysis are free of measurement error) (Brown 2006: 3; Thompson – Green 2006). For these reasons, MGCFA is a better choice.¹²

In order to be able to compare means across groups, the average has to be constrained to zero in one group to get the model identified (Sörbom 1974; Jöreskog – Sörbom 1996; Little et al. 2006; Arbuckle 2007; Davidov 2008). Such a comparison does not allow for the estimate the absolute mean in each group, but rather the mean difference in the latent variables between the groups (Sörbom 1974; Arbuckle 2007; Davidov 2008). Consequently, if one wants to compare means across all groups, one should compute as many models as groups minus one and choose for each model a different reference group. Alternatively, one can compute confidence intervals for each mean using the standard errors. This method makes possible the comparison between all pairs of means. In the Box 1, the steps mentioned before are shortly presented:

¹¹ The fit refers to the ability of a model to reproduce the data (i.e., usually the variance-covariance matrix). There are a lot of fit measures (three categories: absolute, relative, parsimonious), many of them being available in specialized software.

¹² Also, using MGCFA approach, one can account for method effects. This is when additional covariance among indicators is introduced by the measurement approach (Brown 2006: 3).

Box 1: Testing for measurement invariance: a summary of recommended steps

Step 0: Testing for the configural (CFA) model for each group separately. Ideally, the model should be the same for all groups or contains minor variations. The fit indices should be above the rules of thumb. The number and pattern of factor loadings should be the same and significant (substantive and statistical; usually above 0.40; Brown, 2006: 130).

Step 1: Testing for the configural model for all groups together (multigroup model). If the fit indices are above the rules of thumb the model has a similar factor structure, but it is not necessarily equivalent across groups.

Step 2: Testing for metric invariance. Constrain the factor loadings to be equal across groups and check the fit of the metric model. If the fit difference between configural and metric models is above the rules of thumb look for the source of misfit and free some factor loadings (partial metric invariance) or eliminate some groups from analysis.

Step 3: Testing for scalar invariance. Constrain the intercepts to be equal across groups and check the fit of the scalar model. If the fit difference between metric and scalar models is above the rules of thumb look for the source of misfit and free some intercepts (partial scalar invariance) or eliminate some groups from analysis.

Step 4: Chose a reference group and set its mean to zero. Compute and compare the mean of latent variable across groups.

A final statistical consideration is related to the assumptions of MGCFA. Most of our indicators are ordinal (four, five or ten point-scales) and some of them violate the assumption of normality. Lubke and Muthén (2004) criticize the use of MGCFA for Likert items (as they may distort the factor structure differently across groups). On the other hand, in simulation studies, De Beuckelaer (2005) argued that using Likert scales and skewed data does not significantly affect the probability of incorrect conclusions in MGCFA.

A commented example

Based on the general framework proposed by multiple authors (Vandenberg – Lance 2000; Byrne 2001, 2004; Brown 2006; Schmitt – Kuljanin 2008) and some research examples (Byrne 2008; Davidov 2008) we have followed a set of steps for testing for scalar measurement invariance, and subsequent means comparisons. As a general approach, despite the fact that the number of indicators for our model is relatively small, we have preferred a “step-up” strategy (Brown 2006: 269) instead of a “step-down” strategy (Horn – McArdle 1992)¹³.

¹³ A “step-up” strategy starts with the lowest level of invariance (configural) and test progressively for a higher one (metric, than scalar, etc.). A “step-down” strategy is moving from highest level of invariance to a lower one (if the highest level is not attained).

In order to illustrate the invariance measurement test using the MGCFA approach we employ both a longitudinal and a cross national analysis, aiming to estimate values change and differences in values among Eastern European populations after the fall of the communist regimes. The value of interest here is “sexual permissiveness”. The research questions are: (1) “Are there cross-national differences in sexual permissiveness?” and (2) “Is the trend of sexual permissiveness different across countries?”.

Sexual permissiveness cannot be observed and measured directly. It is an unobservable attribute or a latent variable. All we can do is to measure the manifest indicators of the underlying construct of sexual permissiveness. In other words, we can measure a set of attitudes that are a reflection of what we understand by sexual permissiveness. From many attitudes that can stand for this construct, one is the acceptance of some behaviour that transcends the borders of “normality”: homosexual behaviour, prostitution, abortion, divorce, extramarital affairs, having casual sex, etc. In this case, the sexual permissiveness (latent variable) is supposed to determine (“cause”) the response to the acceptance of homosexual behaviour, prostitution, etc. (observed variables). We should add that there is also some unique variance in each of these indicators that is not accounted for by the latent factor of sexual permissiveness. This unique variance is also known as measurement error, error variance, or indicator unreliability (Harrington 2009: 35). Also, it is possible that some of the indicators have correlated unique variances.

In our analyses sexual permissiveness was tapped by four indicators that measures how justifiable are the following behaviours: homosexuality (V202), prostitution (V203), abortion (V204), and divorce (V205).¹⁴ A ten-point scale was used, where 1 means “never justifiable” and 10 means “always justifiable”. The factor we proposed has a (Inglehart – Baker 2000) and is presented in a similar form in other analyses either as a simple index (Norris – Inglehart 2003; Esmer 2003; Scott – Braun 2006: 69) or as a factor score (principal component analysis) (Hagenaars – Halman – Moors 2003). Norris and Inglehart (2003) use the label “sexual liberalization scale” including here homosexuality, abortion and divorce; Esmer (2003) talk about “sexual tolerance scale” (homosexuality, abortion, prostitution); Scott and Braun (2006) use the label “sexual values index” measuring it as an average of the valid values of four items: attitudes towards homosexuality, abortion, divorce, and adultery; Hagenaars et al. (2003) talk about “personal sexual permissiveness” index including here the acceptability of homosexuality, abortion, divorce, euthanasia, suicide, and having casual sex. In all cases the scale is one-dimensional. The source of these indexes seems to be the work of Inglehart about materialist-postmaterialist

¹⁴ Variable names are from the four wave WVS/EVS official integrated dataset (<http://www.wvsevssdb.com/wvs/WVSDData.jsp>).

values (Inglehart 1997; Inglehart – Baker 2000). Even they compare the countries and cultures regarding their level of sexual liberalism, none of the above mentioned scholars is using a CFA model to test for adequacy of their measure. Also, model adequacy to data (fit indices), measurement invariance of construct and measurement errors issues are ignored. Our approach permits to account for all these aspects and offers to the researchers in the values field the possibility to compare sexual permissiveness longitudinally and across countries.

We have used data are from the WVS and EVS studies.¹⁵ We have restrained our value comparison to 19 formerly communist countries: for each of these at least three EVS or WVS waves are available from a total of five (2: 1990 – 1993, 3: 1994 – 1998, 4:1999 – 2004, 5: 2005 – 2007 and 6: 2008 – 2009).¹⁶ These waves cover twenty years (1990 – 2009). All analyses are illustrated using AMOS 16 software and SPSS 13. As estimator we used Maximum Likelihood (ML) for all models.

Step one: Computing the variance-covariance matrices

Before computing the matrices in SPSS, in order to equalize the influence of sample volume on statistical power across groups, the database was weighted with the provided weights and the number of cases was fixed to 1500 for each group (a combination of country and wave). Then, for each group, we computed the variance-covariance matrices containing all the sexual permissiveness measures (Box 2). We preferred this method instead of using raw data because AMOS (Arbuckle 2007) ignores the weights, and also, because we can use modification indices¹⁷ (they report about possible model misspecifications). Since all the indicators for most of the groups have no more than 5% missing values and around 10% if they are considered together, we preferred to work with data matrices (listwise deletion¹⁸) for all the analyses.¹⁹

¹⁵ The four wave WVS/EVS official integrated dataset.

(<http://www.wvsevsvdb.com/wvs/WVSDData.jsp>) and the first release of the fourth EVS wave 2008 (<http://www.europeanvaluesstudy.eu/evs/surveys/survey-2008.html>).

¹⁶ WORLD VALUES SURVEY 1981 – 2008 OFFICIAL AGGREGATE v.20090901, 2009. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid.

¹⁷ AMOS does not compute modification indices if your data has missing values.

¹⁸ In the case of pairwise deletion the risk to obtain a matrix not positive definite is bigger. If the matrix to be analyzed is found to be not positive definite, AMOS will give you an error message and no solution.

¹⁹ Alternatively, one can deal with the missing values problem by using the raw data and the full information maximum likelihood procedure (FIML in AMOS).

Box 2: The SPSS syntax used for computing the variance-covariance matrix for Slovakia, wave 2²⁰

```
USE ALL.  
COMPUTE filter_$(S024A = 7032).  
FILTER BY filter_$.  
EXECUTE.  
CORRELATIONS  
/VARIABLES= V202 V203 V204 V205  
/PRINT=TWOTAIL SIG  
/STATISTICS DESCRIPTIVES  
/MISSING=LISTWISE  
/matrix out ("C:\COV\cov_Slovakia (2).sav").
```

Step two: Estimating a general model

The aim of this stage is to identify and validate a measurement model for sexual permissiveness, according with the theoretical model on which this analysis relies. For this step, in order to find easier the general model, all cases (from different countries and waves) were treated as a single group. First, we tested the simplest model: one factor, four indicators and four measurement errors (all other possible relations were set to 0). The initial measurement model looks like the final one (Figure 1) without the correlation between the two measurement errors (the double arrow between e202 and e203).

For model identification we have set to 1 the loading of V202 indicator (marker indicator: justifiable homosexuality) (Yoon – Millsap 2007).²¹ By doing this, we have also set the measurement unit for the latent variable (Jöreskog – Sörbom 1996). The problem is that, as discussed above, the use of a fixed loading for identification purposes can interfere with the process of finding the correct model of partial invariance (Cheung – Rensvold 1999). To solve this problem we chose to re-run the CFA with different marker indicators (Vandenberg 2002; Brown 2006: 302) and select the best model according to goodness-of-fit indexes. Even so, the measures of goodness-of-fit related to the initial model revealed a poor fit (Table 2). Consequently we investigate for the source of poor fit by looking at Modification Indices (report about model misspecifications) (Arbuckle 2007). The biggest modification index (18,930) corresponds to the covariance between e202 and e203 measurement errors. The second and third largest modification indices appear for the two regression weights associated to V202 and V203 variables (6,147 for V202 to V203, respectively 6,068 for V203 to V202). Theoretically, this makes sense: apart

²⁰ The syntax could be easily replicated for each group by using Excel software (Annexes: Box 5).

²¹ The loadings of measurement errors are set to 1 by AMOS Graphics, but in VB.NET this has to be done by analyst (all you have to do is to put "(1)" before a measurement error).

from their dependency on sexual permissiveness, prostitution and homosexuality are relatively close related to religious permissiveness also; prostitution and homosexuality are less tolerated than abortion and divorce; besides that, frequently, media presents these practices all together. Therefore we have theoretically reasons to modify the initial model by including the correlation between the two measurement errors. According to the rules of thumb (Hu – Bentler 1999; Marsh – Hau – Wen 2004) the goodness-of-fit indexes for the respecified (final) model (Figure 1, Table 2) indicate a close fit to the data. Also, the standardized regression weights are at least satisfactory (.476 for V202, .444 for V203, .839 for V204, and .791 for V205).

Figure 1: **The final measurement model of sexual permissiveness – AMOS Graphics**²²

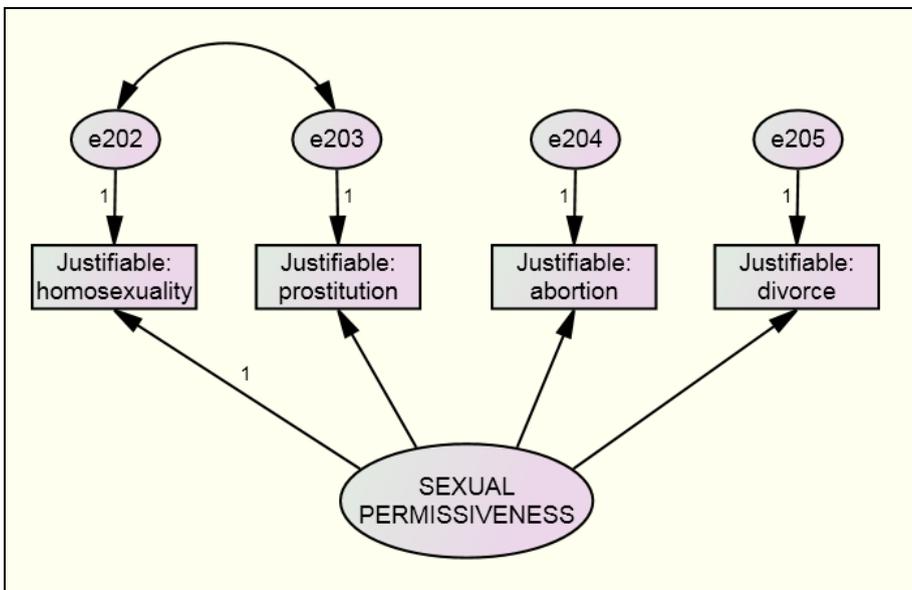


Table 2: **Baseline models assessment (goodness-of-fit statistics)**

Model	CMIN/DF	TLI	CFI	RMSEA	PCLOSE
Initial	10581	.508	.836	.315	.000
Final	86	.996	.999	.028	1.000

* sample size = 106680

²² The initial model is identical with the final model except the double arrow between e202 and e203 (auto-correlated residuals). See also Annexes: Box 6.

Before passing to the next step one more caution is necessary. According to the literature (Byrne 2001: 175-176, 2006; Davidov 2008), before conducting a multi-group comparison, one should pay attention to single-group analyses. This step is required in order to check if the baseline model fits adequately in each group. Consequently, the final model was tested separately for each group (combination of countries and waves). The analyses revealed a close fit to the data, and that all items have a relatively high factor loading on their construct. Before proceeding to the next step, we investigate also the standardized factor loadings (regression weights), variances and squared multiple correlations for each group, looking for out of range values (Heywood cases). In the case of Serbia wave 5 we found one standardized regression weight larger than 1 (1.8 for V204), one negative variance (-23 for V204) and one squared multiple correlation above the maximum value possible which is 1 (3.5 for V204). As a result of these findings we eliminate this group from all subsequent analyses.

Table 3: Goodness-of-fit statistics for a MGCFA of sexual permissiveness (all groups)

Invariance type	CMIN/DF	TLI	CFI	RMSEA	PCLOSE
Configural	5.447	.983	.997	.007	1.000
Metric full	19.999	.928	.952	.014	1.000
Scalar full	55.331	.793	.761	.023	1.000
Scalar partial i1 *	36.114	.866	.868	.019	1.000
Scalar partial i12 **	25.591	.906	.923	.016	1.000

* V202 intercept relaxed.

** V202 and V203 intercepts relaxed.

Step three: Testing for configural invariance

In this stage we have considered all groups simultaneously and tested for configural invariance. For doing this, when the number of groups is big, AMOS syntax (VB.NET) is a better alternative to AMOS Graphics. In the last one, defining the groups takes more time than in syntax if you follow the steps mentioned in the annex (Box 7). The corresponding AMOS syntax is presented in Box 3. The names associated to model parameters (regression weights and intercepts) are not necessary for this step but it is easier have them this way in the light of the next steps (namely, testing for metric and scalar invariance). Because the name of one parameter differs among groups this does not make any difference for the configural model (it does not constrains the parameters). As one can see in the syntax, in the model section, the configural model has no constrains. The fit indices calculated for the configural model revealed a very

close fit to the data (TLI = .983, CFI = .997, RMSEA = .007, Pclose = 1.000)²³ (Hu – Bentler 1999; Marsh, Hau – Wen 2004). In addition, all factor loadings are highly significant (p=.000), even if in some cases two of them (V202 and V203) are relatively small (below .400). All the findings suggest that the configural invariance is most probably supported for each group (combination of country and wave). This finding allows us to proceed to the next step (while keeping in mind that configural invariance is a pre-condition for metric invariance).

Box 3: Testing for configural and metric invariance using a AMOS syntax²⁴

```

Module MainModule
  Public Sub Main()
    Dim Sem As AmosEngine
    Sem = New AmosEngine
    Sem.Initialize("E:\MI\Models\2.2 - configural and metric") ' you need to
modify the path
    Sem.TextOutput ()
    AnalysisProperties(Sem)
    ModelSpecification(Sem)
    Sem.FitAllModels()
    Sem.Dispose()
  End Sub

  Sub ModelSpecification(Sem As AmosEngine)
    Sem.GenerateDefaultCovariances(False)

' put your line codes for groups here
' don't forget to delete the models of groups with no data
Sem.BeginGroup("E:\MI\Matrices\cov_Albania (3).sav" , "cov_Albania (3)")
Sem.GroupName("Albania (3)")
Sem.AStructure("V202 = (i1_Albania_3 + (1) SEXUAL_PERM + (1) e202")
Sem.AStructure("V203 = (i2_Albania_3 + (w2_Albania_3) SEXUAL_PERM + (1)
e203")
Sem.AStructure("V204 = (i3_Albania_3 + (w3_Albania_3) SEXUAL_PERM + (1)
e204")
Sem.AStructure("V205 = (i4_Albania_3 + (w4_Albania_3) SEXUAL_PERM + (1)
e205")

```

²³ Because of relative large samples size (around 1500 for each country), we chose to not apply the chi-square difference test (Cheung and Rensvold, 2002; Davidov, 2008). Following the recommendation of the same authors we preferred to use the CFI (Gamma hat and McDonald's Noncentrality Index are not available in AMOS) and RMSEA.

²⁴ The group part of the syntax (grey part) could be easily replicated for each group by using Excel software (Annexes: Box 7).

```

Sem.AStructure("e203 <--> e202 (c1_1)")
Sem.AStructure("SEXUAL_PERM (v1_1)")
Sem.AStructure("e202 (v1_1)")
Sem.AStructure("e203 (v2_1)")
Sem.AStructure("e204 (v3_1)")
Sem.AStructure("e205 (v4_1)")
Sem.GroupName("Ukraine (6)")
Sem.AStructure("V202 = (i1_Ukraine_6 + (1) SEXUAL_PERM + (1) e202)")
Sem.AStructure("V203 = (i2_Ukraine_6 + (w2_Ukraine_6) SEXUAL_PERM + (1) e203)")
Sem.AStructure("V204 = (i3_Ukraine_6 + (w3_Ukraine_6) SEXUAL_PERM + (1) e204)")
Sem.AStructure("V205 = (i4_Ukraine_6 + (w4_Ukraine_6) SEXUAL_PERM + (1) e205)")
Sem.AStructure("e203 <--> e202 (c1_79)")
Sem.AStructure("SEXUAL_PERM (v1_79)")
Sem.AStructure("e202 (v1_79)")
Sem.AStructure("e203 (v2_79)")
Sem.AStructure("e204 (v3_79)")
Sem.AStructure("e205 (v4_79)")

' define and give a name to your models
' don't forget to delete parameters of groups with no data
    Sem.Model("Configural", "")
    Sem.Model("Metric full", "w2_Albania_3 = ... w2_Ukraine_6;
w3_Albania_3 = ... w3_Ukraine_6;
    w4_Albania_3 = ... w4_Ukraine_6")

End Sub

Sub AnalysisProperties(Sem As AmosEngine)
    Sem.Iterations(50)
    Sem.InputUnbiasedMoments
    Sem.FitMLMoments
    Sem.Standardized
    Sem.Smc
    Sem.ResidualMoments
    Sem.FactorScoreWeights
    Sem.Mods(4)
    Sem.Seed(1)
End Sub
End Module

```

Step four: Testing for metric invariance

In this stage we have tested the model for metric invariance by imposing constraints on particular parameters (Jöreskog – Sörbom 1996; Bentler 2004). Testing for metric invariance means that the factor loadings between the indicators and their corresponding factor (sexual permissiveness) were constrained to be the same in each country and wave. These constraints appear in the model section (Box 3), under the name “Metric full”. All the fit indices suggest that the metric invariance model should not be rejected (TLI = .928, CFI = .952, RMSEA = .014, Pclose = 1.000 (Table 3).

Even the fit indexes for metric model looks fine, they are worse than those associated to the configural model (CMIN/DIF drops 14.5 and CFI .045). Usually, for model comparison and assessing multi-sample model invariance, instead of χ^2 difference test ($\Delta\chi^2$ test), the goodness-of-fit index difference (Δ goodness-of-fit index) is preferred (Cheung – Rensvold 2002; Chen 2007). According to the rules of thumb, Δ RMSEA above .01 and Δ CFI above .02 are signs of fit decreasing. In our case, Δ RMSEA was smaller than .01, but the Δ CFI and Δ TLI were larger than .02. This is a sign for lack of invariance in the case of some regression weights. The source of this lack of invariance can be multiple: one (the same) regression weight for all groups, more regression weights but only for some groups, or a combination of the above.

The solution for the first problem is to relax some constraints and to test for partial metric invariance. So, we tried to improve the fit by freeing one by one the regression weights. The best fit was attained in case of V203 variable (CFI increased by .013 and TLI by .001) but the differences are still above the recommended criteria.

For the second source of misfit one possible solution²⁵ is to remove from analyses the groups with large modification indices. After inspecting the modification indices, we observed that in the case of some groups the fit was not so good. Removing those groups (Albania 4, Bosnia Herzegovina 6, Czech Republic all, Germany East all, Slovakia 6 and Slovenia 4-5.) improved the fit (+.017 for CFI and +.025 for TLI), but not close enough to the fit of configural model.

The two approaches presented before produced a better solution when used together (Table 4). This time, the Δ goodness-of-fit indexes for the metric model are smaller than the maximum recommended.

²⁵ For models with more indicators and factors one could add new parameters (correlations between errors, cross-loadings) for some groups. In our case (one factor and four indicators) this strategy could not be applied, in some cases, and it did not substantially improve the fit, in other cases.

Table 4: Goodness-of-fit statistics for a MGCFA of sexual permissiveness (some groups excluded *)

Invariance type	CMIN/DF	TLI	CFI	RMSEA	PCLOSE
Configural	5.318	.983	.997	.007	1.000
Metric full	12.788	.953	.969	.012	1.000
Metric partial **	10.884	.981	.981	.011	1.000
Scalar partial ***	18.078	.932	.955	.015	1.000

* Albania 4, Bosnia Herzegovina 6, Czech Republic all, Germany East all, Slovakia 6 and Slovenia 4-5.

** V203 regression weight relaxed.

*** V202 and V203 intercepts relaxed.

Step five: Testing for scalar invariance

Scalar invariance means that, besides metric invariance, the intercepts of the items are equal across groups. Only scalar invariance allows for the comparison of factors means (which is our research objective). The TLI and CFI fit indices for the model with all groups (full scalar model, Table 3) indicate that one should reject it, despite the fact that RMSEA and Pclose indicate a good fit. In order to improve the fit we have relaxed the equality constraint for one intercept (“scalar partial i1” model), than for two intercepts (“scalar partial i12” model). Now fit indices suggest that we should not reject the “scalar partial i12” model. Because the Δ goodness-of-fit indexes (TLI and CFI) are relatively large, we preferred to modify the model by relaxing one weight and removing some groups (see step four). The new scalar partial model (Table 4) has a better fit and the Δ CFI is close to the maximum of .02. Thus, we can conclude that the value scales meet the partial scalar invariance test. Consequently, we may compare the means of sexual permissiveness across these groups (Byrne, 2008; Davidov, 2008).

Step six: Computing and comparing the latent factor means

After establishing partial scalar invariance across groups, we have tested for latent factor means differences via MGCFA with mean structures (Box 4). The model specification for the SEM model is now “Sem.ModelMeansAndIntercepts”. Also, a “Sem.Mstructure” command line was added for each group except the reference one (Romania 2 in this case).

Box 4: Testing for scalar invariance using a AMOS syntax²⁶

```
Module MainModule
  Public Sub Main()
    Dim Sem As AmosEngine
    Sem = New AmosEngine
    Sem.Initialize("E:\MI\Models\2.3 - scalar") ' you need to modify the path
    Sem.TextOutput
    AnalysisProperties(Sem)
    ModelSpecification(Sem)
    Sem.FitAllModels()
    Sem.Dispose()
  End Sub

  Sub ModelSpecification(Sem As AmosEngine)
    Sem.ModelMeansAndIntercepts
    Sem.GenerateDefaultCovariances(False)
    ' put your line codes for groups here
    ' don't forget to delete the models of groups with no data
    ' don't forget to delete the "Sem.Mstructure" line for the reference group
    Sem.BeginGroup("E:\MI\Matrices\cov_Albania (3).sav" , "cov_Albania (3)")
    Sem.GroupName("Albania (3)")
    Sem.AStructure("V202 = (i1_Albania_3 + (1) SEXUAL_PERM + (1) e202)")
    Sem.AStructure("V203 = (i2_Albania_3 + (w2_Albania_3) SEXUAL_PERM + (1) e203)")
    Sem.AStructure("V204 = (i3_Albania_3 + (w3_Albania_3) SEXUAL_PERM + (1) e204)")
    Sem.AStructure("V205 = (i4_Albania_3 + (w4_Albania_3) SEXUAL_PERM + (1) e205)")
    Sem.AStructure("e203 <--> e202 (c1_1)")
    Sem.AStructure("SEXUAL_PERM (v1_1)")
    Sem.AStructure("e202 (v1_1)")
    Sem.AStructure("e203 (v2_1)")
    Sem.AStructure("e204 (v3_1)")
    Sem.AStructure("e205 (v4_1)")
    Sem.Mstructure("SEXUAL_PERM (m_Albania_3)")
    Sem.BeginGroup("E:\MI\Matrices\cov_Ukraine (6).sav" , "cov_Ukraine (6)")
    Sem.GroupName("Ukraine (6)")
    Sem.AStructure("V202 = (i1_Ukraine_6 + (1) SEXUAL_PERM + (1) e202)")
    Sem.AStructure("V203 = (i2_Ukraine_6 + (w2_Ukraine_6) SEXUAL_PERM + (1) e203)")
    Sem.AStructure("V204 = (i3_Ukraine_6 + (w3_Ukraine_6) SEXUAL_PERM + (1) e204")
```

²⁶ The group part of the syntax (grey part) can be easily replicated for each group by using Excel software (Annexes: Box 7).

```

Sem.AStructure("V205 = (i4_Ukraine_6 + (w4_Ukraine_6) SEXUAL_PERM + (1)
e205")
Sem.AStructure("e203 <--> e202 (c1_79)")
Sem.AStructure("SEXUAL_PERM (v1_79)")
Sem.AStructure("e202 (v1_79)")
Sem.AStructure("e203 (v2_79)")
Sem.AStructure("e204 (v3_79)")
Sem.AStructure("e205 (v4_79)")
Sem.Mstructure("SEXUAL_PERM (m_Ukraine_6)")
' define and give a name to your models
' don't forget to delete parameters of groups with no data
      Sem.Model("Scalar full", "w2_Albania_3 = ... w2_Ukraine_6; w3_Albania_3
= ... w3_Ukraine_6;
          w4_Albania_3 = ... w4_Ukraine_6; i1_Albania_3 = ... =
i1_Ukraine_6; i2_Albania_3 = ... = i2_Ukraine_6;
          i3_Albania_3 = ... = i3_Ukraine_6; i4_Albania_3 = ... =
i4_Ukraine_6")
      Sem.Model("Scalar partial i1", "w2_Albania_3 = ... w2_Ukraine_6;
w3_Albania_3 = ... w3_Ukraine_6;
          w4_Albania_3 = ... w4_Ukraine_6; i2_Albania_3 = ... =
i2_Ukraine_6;
          i3_Albania_3 = ... = i3_Ukraine_6; i4_Albania_3 = ... =
i4_Ukraine_6")
      Sem.Model("Scalar partial i12", "w2_Albania_3 = ... w2_Ukraine_6;
w3_Albania_3 = ... w3_Ukraine_6;
          w4_Albania_3 = ... w4_Ukraine_6; i3_Albania_3 = ... =
i3_Ukraine_6; i4_Albania_3 = ... = i4_Ukraine_6")
      End Sub

Sub AnalysisProperties(Sem As AmosEngine)
  Sem.Iterations(50)
  Sem.InputUnbiasedMoments
  Sem.FitMLMoments
  Sem.Standardized
  Sem.Smc
  Sem.ResidualMoments
  Sem.FactorScoreWeights
  Sem.Mods(4)
  Sem.Seed(1)
End Sub
End Module

```

The AMOS output file includes at “estimates-scalars” section the means of sexual permissiveness, together with significance test. We should keep in mind that the reference group is Romania wave 2, so the p-values only inform us

about the statistical significance of the difference between the mean of Romania 2 and another group. If we want to test for other differences we can change the reference group or, better, we can compute the confidence intervals for means using the standard errors²⁷ provided by AMOS in the same table. Using these results, one can explore the sexual permissiveness trend within countries and make comparisons between countries at the same time. The means can be represented in a table together with confidence intervals, p-values or arrows (in order to suggest the sense and the magnitude of the difference one may use up / down arrows and one to three arrows). Also, a graphic trend line can be added (Table 5). For the purpose of visual comparisons, the minimum and maximum values for the Y axis should be the same, but with different signs.

Alternatively, the means could be graphically represented using a clustered column chart (Graphic 1A) or a High-Low chart from SPSS (the high and the low values correspond to the confidence intervals) (Graphic 1B).

Our results show that there are differences in sexual permissiveness cross nationally. No matter the wave, in some of the countries (Bosnia Herzegovina, Lithuania, Moldavia, Poland, Romania, and Ukraine) the sexual permissiveness is lower than the reference value. In rest of the countries, the average of sexual permissiveness is oscillating around the reference value, and only in Slovenia is constantly above the reference value. We should add that most of the differences are below one, so, in substantive terms, this is a small difference on a ten-point scale.

Related to the second research question (“Is the trend of sexual permissiveness different across countries?”) the answer is yes. One category of countries is relatively stable: below reference value (Poland), above the reference (Slovenia) or close to the reference (Bosnia Herzegovina, Montenegro, Russian Federation, and Slovakia). In some countries (Moldavia, Romania, and Ukraine) the sexual permissiveness is decreasing and in others is increasing close to the reference (Lithuania). And finally, some countries show mixed tendencies: the sexual permissiveness is increasing, then it is decreasing closer to (Belarus, Bulgaria, Estonia, and Hungary) or lower (Albania, Latvia, and Serbia) than the reference value. If we are using the data from all countries the conclusions are quite similar. In the missing countries (Germany East and Czech Republic), the sexual permissiveness is generally above the reference value (only Germany wave 2 is close to the reference). The trend is up above the reference, then it is decreasing but still above the reference.

²⁷ For a 95% confidence interval, use the formula: mean \pm 2SE.

Table 5: Sexual permissiveness latent means across countries and waves

Sexual permissiveness	Wave 2 ~1990	Wave 3 ~1995	Wave 4 ~1998	Wave 5 ~2005	Wave 6 ~2008	Trend (1990-1995-1998-2005-2008)
Albania		0,311 ↗↗			-0,524 ↘↘	
Belarus	<u>-0,062</u> ≡	<u>-0,018</u> ≡	0,225 ↗		-0,136 ≡	
Bosnia Herzegovina		-0,156 ↘	-0,178 ↘			
Bulgaria	<u>-0,033</u> ≡	0,286 ↗		0,364 ↗↗	<u>-0,054</u> ≡	
Estonia	<u>0,090</u> ≡	0,221 ↗			-0,103 ≡	
Hungary	<u>0,018</u> ≡	0,328 ↗↗			<u>0,075</u> ≡	
Latvia	0,156 ↗	0,313 ↗↗			-0,237 ↘	
Lithuania	-0,472 ↘↘↘	-0,333 ↘↘	-0,323 ↘↘		<u>-0,027</u> ≡	
Moldavia		-0,395 ↘↘	-0,414 ↘↘	-0,547 ↘↘↘	-0,902 ↘↘↘	
Montenegro		-0,104 ≡	-0,203 ↘		-0,202 ↘	
Poland	-0,536 ↘↘↘	-0,404 ↘↘		-0,523 ↘↘↘	-0,473 ↘↘↘	
Romania	base	<u>0,033</u> ≡	-0,326 ↘↘	-0,509 ↘↘↘	-0,336 ↘↘	
Russian Federation	-0,135 ≡	<u>0,034</u> ≡	<u>-0,009</u> ≡	-0,156 ↘	-0,091 ≡	
Serbia		0,145 ≡	<u>0,002</u> ≡		-0,266 ↘	
Slovakia	0,102 ≡	0,158 ↗				

Slovenia	0,578 <u>↗↗↗</u>	0,357 <u>↗↗</u>		0,438 <u>↗↗</u>		
Ukraine		-0,169 <u>↘</u>	-0,334 <u>↘↘</u>	-0,271 <u>↘</u>	-0,467 <u>↘↘↘</u>	

* *Values underlined are not significant (0.05). The rest of values are significant for $p < 0.01$.*

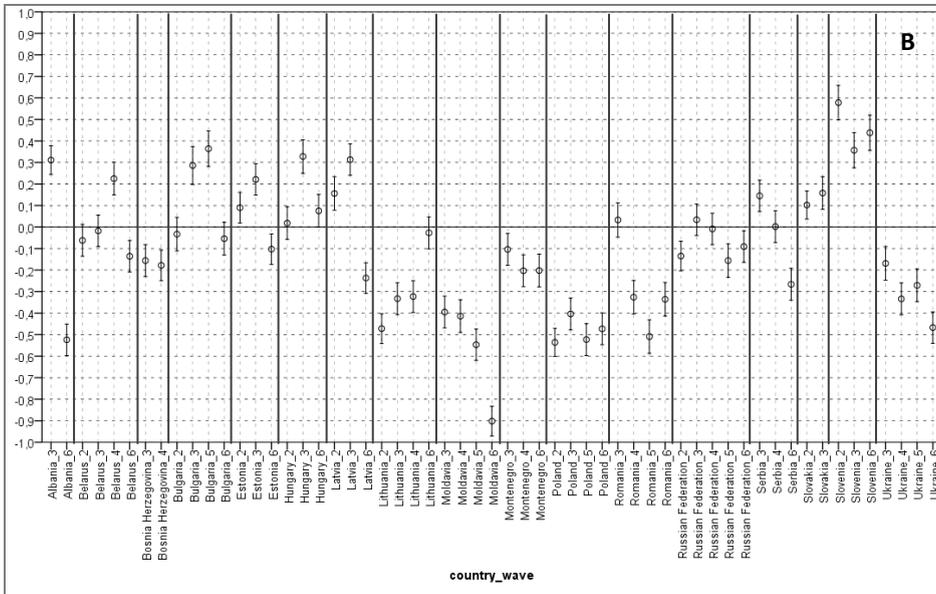
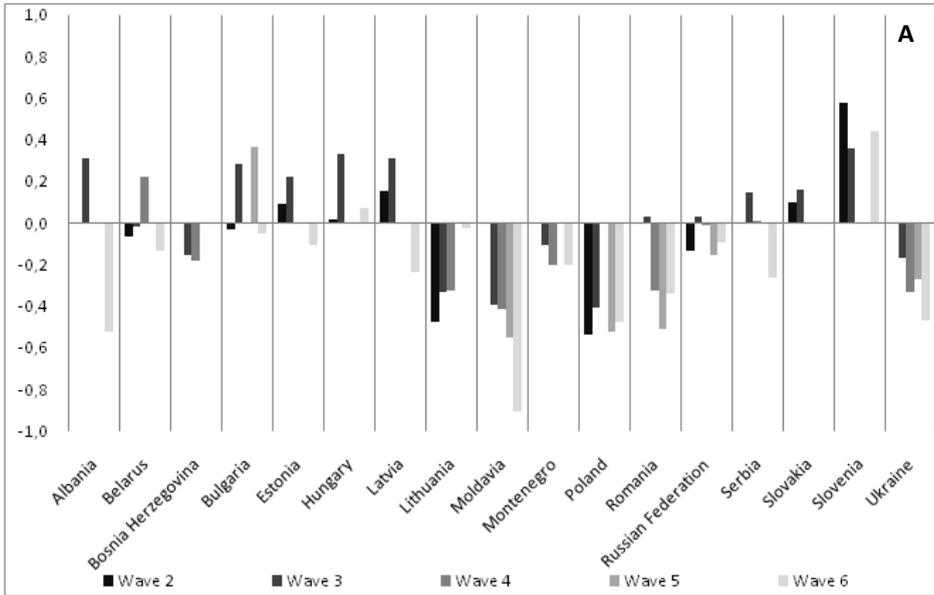
Legend:

< -12SE <u>↘↘↘</u>	< -8SE <u>↘↘</u>	< -4SE <u>↘</u>	<-4SE; 4SE> <u>≡</u>	> 4SE <u>↗</u>	> 8SE <u>↗↗</u>	> 12SE <u>↗↗↗</u>
-----------------------	---------------------	--------------------	----------------------------	-------------------	--------------------	----------------------

Conclusions

In this paper I have tried to argue about the usefulness of testing for invariance measurement, and I provided an example of how to do it using Amos 16 software. The technique, apparently complex and difficult to use, prove to be easy to implement using contemporary software packages. It became a must when analyzing trends and assessing cross-county differences in large-scale datasets. As example I have used the latent orientation towards sexual permissiveness, and I have discussed the measurement and how it varied between 1990 and 2008 in post-communist societies. The scale that the value surveys use proved to be adequate to almost all Eastern-European countries. Based on our findings, researchers can use sexual permissiveness scale for longitudinal and across communist countries comparisons. Scalar invariance of sexual permissiveness scale assures them that relationships with other variables and means can be meaningfully interpreted. Also, the results reveal interested differences across societies. Similar analyses could be carried out for most of the latent variables whose effects are measured in international surveys.

Graphic 1: Sexual permissiveness latent means across countries and waves



Annexes

Box 5: How to compute easily the variance-covariance matrix for more groups in SPSS

Step 0: Use "1.1 - variance-covariance matrix.xlsx" file.²⁸
Step 1: Do an excel table with the names and numeric values for all your groups. Pay attention to the columns with yellow background.
Step 2: Fill sheet 1 to 10 according to your number of groups. Change the text in columns with yellow background from sheet 5 and 10 according to your need.
Step 3: Copy first to columns of sheet 1 to 10 in order in sheet "joined". Use "paste special" command and chose only "values".
Step 4: Copy sheet "joined" to sheet "final". Sort the sheet "final" by first column, ascending.
Step 5: Copy the second column from sheet "final" into a SPSS syntax file.
Step 6: Open your SPSS data file and run the syntax.

Box 6: The final measurement model of sexual permissiveness – AMOS VB.NET²⁹

```
Module MainModule
  Public Sub Main()
    Dim Sem As AmosEngine
    Sem = New AmosEngine
    Sem.Initialize("E:\MI\Models\2.1.2 - configural all final") ' you need to
modify the path
    Sem.TextOutput ()
    AnalysisProperties(Sem)
    ModelSpecification(Sem)
    Sem.FitAllModels()
    Sem.Dispose()
  End Sub

  Sub ModelSpecification(Sem As AmosEngine)
    Sem.GenerateDefaultCovariances(False)

    Sem.BeginGroup("E:\MI\Matrices\cov_All.sav" , "cov_All")
    Sem.GroupName("All")
    Sem.AStructure("V202 = (1) SEXUAL_PERMISSIVENESS + (1) e202")
    Sem.AStructure("V203 = SEXUAL_PERMISSIVENESS + (1) e203")
    Sem.AStructure("V204 = SEXUAL_PERMISSIVENESS + (1) e204")
    Sem.AStructure("V205 = SEXUAL_PERMISSIVENESS + (1) e205")
```

²⁸ Variable names and codes are from the first release of the fourth EVS wave 2008 (<http://www.europeanvaluesstudy.eu/evs/surveys/survey-2008.html>).

²⁹ The initial model is identical with the final model except the "Sem.AStructure("e203 <-> e202 (c1_1)")" command line (auto-correlated residuals e202 and e203).

```

Sem.AStructure("e203 <--> e202")
Sem.Model("Configural all final", "")
End Sub

Sub AnalysisProperties(Sem As AmosEngine)
Sem.Iterations(50)
Sem.InputUnbiasedMoments
Sem.FitMLMoments
Sem.Standardized
Sem.Smc
Sem.ResidualMoments
Sem.FactorScoreWeights
Sem.Mods(4)
Sem.Seed(1)
End Sub
End Module

```

Box 7: How to do easily a measurement invariance test in AMOS

Step 0: Use “2.0 - amos syntax.xlsx” file.³⁰

Step 1: Write your model using AMOS notation (see the sheet “model”).

Step 2: Do a excel table with the names of the groups (sheet “model”).

Step 3: For sheet “1” to “10” change the text in columns with yellow background according to your data.

Step 4: If everything is ok in “joined” sheet you will find your AMOS syntax, unordered. Use “copy” and “paste special” (chose only “values”) functions and copy the syntax in “metric” and “scalar” sheets.

Step 5: In sheet “metric” delete the last line command for each group (Sem.Mstructure), and then sort the data ascending by first column.

Step 6: Copy the second column from sheet “metric” into the AMOS syntax file (“2.2.1 - configural and metric.vb” file).

Step 7: Check the matrices for missing data (if there are no data in some of the variance-covariance matrices). Eliminate de model lines for those groups.

Step 8: Impose the constraints required by the test of metric invariance (you can use “constr” sheet to do this). Run the syntax.

Step 9: In sheet “scalar” sort the data ascending by first column.

Step 10: Copy the second column from sheet “scalar” into an AMOS syntax file (“2.3.1 - scalar.vb” file).

Step 11: Check the matrices for missing data (if there are no data in some of the variance-covariance matrices). Eliminate de model lines for those groups.

Step 12: Impose the constraints required by the test of scalar invariance (you can use “constr” sheet to do this). Chose a reference group (remove the "Sem.Mstructure" line code for that group). Run the syntax.

³⁰ Variable names and codes are from the four wave WVS/EVS official integrated dataset (<http://www.wvsevsvdb.com/wvs/WVSDData.jsp>).

Mircea Comşa is associate professor at Faculty of Sociology and Social Work, „Babeş-Bolyai” University of Cluj-Napoca. His main fields of interest include the electoral sociology, survey methodology, and quantitative data analysis. His recent books are: “The European Parliament Elections. Romania 2009” (editor, Polirom, 2010), “Work related attitudes in Romania” (co-author, Soros Foundation Romania, 2008), “Romania higher education system” (co-author, Soros Foundation Romania, 2007), “Lifestyles in Romania after 1989” (Cluj University Press, 2006), “Social life in urban Romania” (co-author, Polirom, 2006), and “General elections 2004. A Sociological Perspective” (editor, Eikon, 2005).

REFERENCES

- ARBUCKLE, L. J., 2007: Amos16.0 User’s Guide. SPSS Inc.
- BENTLER, P. M., 1990: Comparative fit indexes in structural models. *Psychological Bulletin*, 107: 238-246.
- BENTLER, P. M., 2004: Maximal Reliability for Unit-weighted Composites. Department of Statistics, UCLA. Paper 2004100701. <http://repositories.cdlib.org/uclastat/papers/2004100701>.
- BILLIET, J., 2003: Cross-cultural equivalence with structural equation modelling. In: Harkness J., Van de Vijver F., Mohler P. (eds) *Cross-cultural survey methods*, New York: Wiley. pp. 247-265.
- BOLLEN, A. K., 1989: *Structural Equations with Latent Variables*. Wiley-Interscience.
- BOORSBOOM, D., 2006: Relevance of invariance testing: When does measurement invariance matter? *Medical Care*, 44 (11-3): 176-181.
- BROWN, A. T., 2006: *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- BYRNE, M. B. – WATKINS, D., 2003: The issue of measurement invariance revisited. *Journal of Cross-cultural Psychology*, 34: 155-175.
- BYRNE, M. B. – SHAVELSON, R. J. – MUTHÉN, B. O., 1989: Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105: 456-466.
- BYRNE, M. B., 2001: *Structural equation modelling with Amos. Basic concepts, application, and programming*. London: Lawrence Erlbaum Associates.
- BYRNE, M. B., 2004: Testing for Multigroup Invariance Using Amos Graphics: A Road Less Traveled. *Structural Equation Modeling*, 11 (2): 272-300.
- BYRNE, M. B., 2008: Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4): 872-882.
- CHEN, F. F. – SOUSA, K. H. – WEST, S. G., 2005: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, 12: 471-492.
- CHEN, F. F., 2007: Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, 14: 464-504.

- CHEN, F. F., 2008: What Happens If We Compare Chopsticks With Forks? The Impact of Making Inappropriate Comparisons in Cross-Cultural Research. *Journal of Personality and Social Psychology*, 95 (5): 1005-1018.
- CHEUNG, G. W. – RENSVOOLD, R. B., 1999: Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. *Journal of Management*, 25(1): 1-27.
- CHEUNG, G. W. – RENSVOOLD, R. B., 2002: Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9: 233-255.
- DAVIDOV, E., 2008: A Cross-Country and Cross-Time Comparison of the Human Values Measurements with the Second Round of the European Social Survey. *Survey Research Methods*, 2(1): 33-46.
- De BEUCKELAER, A., 2005: *Measurement invariance issues in international management research*. (Unpublished doctoral dissertation, Limburgs Universitair Centrum, Limburg, the Netherlands)
- De JONG, M. G. – STEENKAMP, J.-B. E. M. – FOX, J.-P., 2007: Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, 34: 260-278.
- ESMER, Y., 2003: Is There an Islamic Civilization?. In Inglehart, R. (eds) *Human values and social change. Findings from the values surveys*. Leiden Boston: Brill. pp. 35-68.
- FAN, X. – SIVO, A. S., 2009: Using Goodness-of-Fit Indexes in Assessing Mean Structure Invariance. *Structural Equation Modeling*, 16(1) 54-69.
- FRENCH, B. F. – FINCH, W. H., 2008: Multigroup Confirmatory Factor Analysis: Locating the Invariant Referent Sets. *Structural Equation Modeling*, 15(1): 96-113.
- GRAHAM, W. J. – HOFER, M. S., 2000: Multiple Imputation in Multivariate Research. In: Little, T. D., Schnabel, K. U. and Baumert, J. (eds) *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples*. Lawrence Erlbaum Associates, Inc.
- HAGENAARS, J. – HALMAN, L. – MOORS, G., 2003: Exploring Europe's basic values map. In: Arts, Wil. Hagenaars, Jacques and Halman, Loek. (eds) *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study*, Leiden: Brill. pp. 23-58.
- HARRINGTON, D., 2009: *Confirmatory Factor Analysis*. Oxford: Oxford University Press.
- HORN, J. L. – McARDLE, J. J., 1992: A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18: 117-144.
- HU, L. – BENTLER, P. M., 1999: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6: 1-55.
- INGLEHART, R. – BAKER, E. W., 2000: Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, 65(1) 19-51.
- INGLEHART, R., 1997: *Modernization and Postmodernization: Cultural, Economic and Political Change in 43 Societies*. Princeton: Princeton University Press.
- JÖRESKOG, K. G. – SÖRBOM, D., 1996: LISREL 8: Structural equation modelling with the SIMPLIS command language. SSI, 4th printing.

- JÖRESKOG, K. G., 1971: Simultaneous factor analysis in several populations. *Psychometrika*, 36: 409-426.
- LITTLE, T. D. – SLEGGERS, D. W. – CARD, N. A., 2006: A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, 13(1): 59-72.
- LUBKE, G. H. – MUTHÉN, B. O., 2004: Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11: 514-534.
- MARSH, H. W. – HAU, K. T. – WEN, Z. L., 2004: In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralising Hu and Bentler (1999) findings. *Structural Equation Modeling*, 11: 320-341.
- McARDLE, J. J., 1994: Structural Factor Analysis Experiments with Incomplete Data. *Multivariate Behavioral Research*, 29(4): 409-454.
- MEREDITH, W., 1993: Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58: 525-543.
- MILLSAP, R. E. – EVERSON, H. T., 1993: Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17: 297-334.
- MILLSAP, R. E., 2001: When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling*, 8(1): 1-17.
- NORRIS, P. – INGLEHART, R., 2003: Islamic culture and democracy: testing the 'clash of civilizations' thesis. In Inglehart, R. (eds) *Human values and social change. Findings from the values surveys*. Leiden • Boston: Brill. pp. 5-34.
- REISE, S. P. – WIDAMAN, K. F. – PUGH, R. H., 1993: Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*, 114: 552-566.
- RENSVOLD, R. B. – CHEUNG, G. W., 1998: Testing measurement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement*, 58: 1017-1034.
- RENSVOLD, R. B. – CHEUNG, G. W., 2001: Testing for metric invariance using structural equation models: Solving the standardization problem. In Schriesheim, C. A. and Neider, L. L. (eds), *Equivalence in measurement Research in management*, Vol. I. Greenwich, CT: Information Age. pp. 21-50.
- RIORDAN, C. M. – VANDENBERG, R. J., 1994: A central question in cross-cultural management research: Do employees of different cultures interpret work-related constructs and measures in an equivalent manner? *Journal of Management*, 20(3): 643-671.
- SARIS, W. E. – GALLHOFER, I. N., 2007: *Design, evaluation and analysis of questionnaires for survey research*. New York: Wiley.
- SCHAFFER, J. L. – GRAHAM, J. W., 2002: Missing data: Our view of the state of the art. *Psychological Methods*, 7: 147-177.
- SCHMITT, N. – KULJANIN, G., 2008: Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18: 210-222.

- SCOTT, J. – BRAUN, M., 2006: Individualization of family values? In: Ester, Peter, Braun, Michael and Mohler, Peter. *Globalization, Value Change and Generations*. Boston: Brill, pp.61-87.
- SÖRBOM, D., 1974: A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27: 229-239.
- STARK, S. – CHERNYSHENKO, O. S. – GRASGOW, F., 2006: Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91: 1292-1306.
- STEENKAMP, J.-B. E. M. – BAUMGARTNER, H., 1998: Assessing measurement invariance in cross national consumer research. *Journal of Consumer Research*, 25: 78-90.
- STEIGER, J. H., 2002: When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modelling. *Psychological Methods*, 7: 210-227.
- THOMPSON, M. S. – GREEN, S. B., 2006: Evaluating between-group differences in latent means. In: G. R. Hancock and R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course*. Greenwich: CT: Information Age, pp. 119-169.
- TUFIŞ, C., 2008: Multiple Imputation as a Solution to the Missing Data Problem in Social Sciences. *Calitatea vieţii*, 1-2: 199-212.
- Van de VIJVER, F. J. R. – LEUNG, K., 1997: *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
- VANDENBERG, R. J., 2002: Towards a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5: 139-158.
- VANDENBERG, R. J. – LANCE, C. E., 2000: A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3: 4-69.
- WIDAMAN, K. F. – REISE, S. P., 1997: Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In: Bryant, K. J., Windle, M. and West, S. G. (eds), *The science of prevention*. Washington, DC: American Psychological Association. pp. 281-324.
- YOON, M. – MILLSAP, E. R., 2007: Detecting Violations of Factorial Invariance Using Data-Based Specification Searches: A Monte Carlo Study. *Structural Equation Modeling*, 14(3): 435-463.
- European Values Study Group and World Values Survey Association. European and World Values Surveys Four-Wave Integrated Data File, 1981 – 2004, v. 20060423, 2006. Aggregate File Producers: Análisis Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain/Tilburg University, Tilburg, The Netherlands. Data Files Suppliers: Analisis Sociologicos Economicos y Politicos (ASEP) and JD Systems (JDS), Madrid, Spain/Tillburg University, Tillburg, The Netherlands/ Zentralarchiv fur Empirische Sozialforschung (ZA), Cologne, Germany.) Aggregate File Distributors: Análisis Sociológicos Económicos y Políticos (ASEP) and JD Systems (JDS), Madrid, Spain/Tillburg University, Tilburg, The Netherlands/Zentralarchiv fur Empirische Sozialforschung (ZA) Cologne, Germany.

EVS Foundation/Tilburg University: European Values Study 2008, 4th wave, Integrated Dataset. GESIS Cologne, Germany, ZA4800 Data File Version 1.0.0 (2010-06-30) DOI:10.4232/1.10059.

World Values Survey 2005 Official Data File v.20081015, 2008. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid.