Including waves in a hierarchical model using comparative survey data. How waves should be considered? A comparison of two models.

Pamela Isabel Inostroza Fernandez
Master in Statistics, KU Leuven, Belgium

ABSTRACT

Many types of research of comparative surveys are constantly investigated, for this, a hierarchical model is a standard method to analyse this data. In cases when waves (rounds) are included to check the relations over time, the data is not only comparative but longitudinal as well. Social Trust indicator is studied in this paper, looking for relevant factors that explain the variance across countries considering the time variation. Variables studied are gender, age, benevolence index, years of education of individuals and aggregated to a country level as well and gross domestic product index.

First, is studied how much variance can be divided into between countries and between rounds with a three-level model. As not much variance can be found among rounds of the survey for the Social Trust indicator, it is studied a two-level model using the time variation as a random slope as used in a growth curve modelling. Individual variables coefficient estimates behave the same in both models, considering coefficients and significance, variance explained differ slightly. In contrast, aggregated variables at the country level estimations differs in strength and significance comparing both models, also contextual effects vary considerably in strength.

Multilevel model considering waves as an intermediate level improves the total variance partition in country and waves variances, being explicated in more than 25% and 34% respectively after including predictors. On the other hand, only 17% is explained in the two-level model at the country level. Even though estimates for the random slope of benevolence index are consistent between both models, variances in the model differ, decreasing the residual variance in 6%, but increasing country level variance in 5% in the two-level model and only a 1% in the three-level model. Diagnostic of the model indicates that assumptions are correct for both models, predictions rank remain similar in both models showing changes only at the scale.

In conclusion, both models can estimate adequately individual effects, but time variation as a random slope will underestimate the relation of the higher-level variables. Consequently, misleading results will be obtained by no considering time variation as a nested characteristic of the participants, not allowing to extract the variation of each participant around waves. In this case, even it is tempting to use a longitudinal approach this should not be done in order to not risk the richness in information that a comparative longitudinal survey can offer.

RESEARCH QUESTION

It is a very usual task in research to analyse data from the European Social Survey, this data is freely available, this data has been available since a long time, therefore, a large number of rounds is available to be studied. The survey sampling design provides a hierarchical structure given by the country level. This study is focused on compare which methodology should be implemented in order to test which are the factors that affect the Social Trust scale considering all participating countries in the last 10 years rounds of the survey, and how this will make a difference in the results.

Due to reduced sample sizes of the groups, it is possible to pool the comparative survey data across time permitting an increase of statistical power, the resulting structure is called comparative longitudinal survey data (Fairbrother 2014). Thanks to the expansion of computational power, and improvements of statistical software, multilevel models have become a commonly used tool in social science. Different approaches have been studied to be used with this type of data, considering that the countries included are not random samples, the number of countries included is rather small and the small degrees of freedom at the country level limits the number of higher-level control variables that can be included (Schmidt-Catran, Fairbrother, and Andreß 2019). It is first stated that a three-level model should be used for this type of data but for this, waves (round level) should be explicative by itself.

In this study, it will be developed a three-level model considering waves (rounds) as a level itself and a two-level model considering waves as a random slope, in both models, the highest level is country and the lowest is the respondents. This paper traces the development of each model comparing differences, examining the direction and strength of the estimates. A two-level model is not fully recommended as participants in the study are not the same in each round, that makes this data not a longitudinal panel survey but a repeated cross-sectional survey. Either way, it would be interesting to compare in which instance these two models differ. The assumption of cross-sectional and longitudinal equivalence in many instances are not valid (Fairbrother 2014), for this, the intention of this research is not to make a generalization of the usage of the data, this is a particular case that should not be oversimplified.

This study is focused on the Social Trust scale included in the ESS, this construct is computed as a continuous indicator using Confirmatory factor analysis accordingly to (Breyer 2015). For modelling predictors, variables considered are extracted from the same survey, Age, Gender, and Years of education of participants as background characteristic. From the same survey, Benevolence subscale from the human value scale, will be used as an indicator of the personal characteristic of the participants, this index was calculated using Confirmatory factor analysis in order to obtain a continuous variable (Schwartz 2003). As higher-level variables in the model, the contextual effect is studied considering aggregate variable Country's years of education (average derived from participant level) and as country background characteristic the global variable, Gross Domestic Product.

DESCRIPTION OF THE DATASET

The European Survey (ESS) is a survey administered in over 35 countries to date and the objective is to monitor and interpret changing public attitudes, improve methodologies for cross-national survey measurement in Europe and to develop a series of European social indicators. Five last rounds of the survey will be used[1] (European Social Survey 2018), this corresponds from 5th round applied in 2010 to 9th in 2018, each round was applied by two years

[1] https://www.europeansocialsurvey.org/data/

difference. The survey was responded by persons aged 15 and over, resident within private households, regardless of their nationality, citizenship, language or legal status (European Social Survey 2018).

The hierarchical structure (Figure 1) of the data is given by 31 countries, with approximately 2000 observations by round, the majority of the participating countries has information from all five rounds, countries with less information, as they participate in just one round, are Greece (k=13), Croatia (k=14) and Serbia (k=31). Each number of participants by country/round can be seen in Table 6 in the appendix Figure 1.

L3 ID: 0, k = 1 of 31 N2 5 , N1 8575	L3 ID: 1, k = 2 of 31 N2 3 , N1 4256	L3 ID: 2, k = 3 of 31 N2 5 , N1 6683	L3 ID: 3, k = 4 of 31 N2 3 , N1 2626	L3 ID: 4, k = 5 of 31 N2 5 , N1 8842
L3 ID: 5, k = 6 of 31 N2 5 , N1 12988	L3 ID: 6, k = 7 of 31 N2 3 , N1 4229	L3 ID: 7, k = 8 of 31 N2 5 , N1 9041	L3 ID: 8, k = 9 of 31 N2 4 , N1 6225	L3 ID: 9, k = 10 of 31 N2 5 , N1 8896
L3 ID: 10, k = 11 of 31 N2 5 , N1 8776	L3 ID: 11, k = 12 of 31 N2 5 , N1 9940	L3 ID: 12, k = 13 of 31 N2 1 , N1 2426	L3 ID: 13, k = 14 of 31 N2 1 , N1 1211	L3 ID: 14, k = 15 of 31 N2 5 , N1 6388
L3 ID: 15, k = 16 of 31 N2 5 , N1 10721	L3 ID: 16, k = 17 of 31 N2 4 , N1 6735	L3 ID: 17, k = 18 of 31 N2 4 , N1 5623	L3 ID: 18, k = 19 of 31 N2 5 , N1 7902	L3 ID: 19, k = 20 of 31 N2 5 , N1 7147
L3 ID: 20, k = 21 of 31 N2 5 , N1 6347	L3 ID: 21, k = 22 of 31 N2 4 , N1 5751	L3 ID: 22, k = 23 of 31 N2 3 , N1 5300	L3 ID: 23, k = 24 of 31 N2 4 , N1 5967	L3 ID: 24, k = 25 of 31 N2 5 , N1 5511
L3 ID: 25, k = 26 of 31 N2 2 , N1 2950	L3 ID: 26, k = 27 of 31 N2 2 , N1 2104	L3 ID: 27, k = 28 of 31 N2 2 , N1 1357	L3 ID: 28, k = 29 of 31 N2 3 , N1 5158	L3 ID: 29, k = 30 of 31 N2 3 , N1 5498
L3 ID: 30, k = 31 of 31 N2 1 , N1 1498				

Figure 1: Structure of the three-level model-implied hierarchy

In Figure 2 it is possible to see a visual representation of the relation between each predictor and the response variable, considering differences produced by each country and round they participated. All of them suggest a linear relation but that changes in the slope, some variables more than others.

Description of variables used in the model:

SOCIAL TRUST: (Continuous dependent variable) Generalized expectancy that the words or promises of others can be relied on.

GENDER: (Dichotomous independent variable) Sex of respondent (Female, Male).

AGE: (Continuous independent variable) Age of respondent.

BENEVOLENCE: (Continuous independent variable) Indicator for preservation, and enhancement of the welfare of people with whom one is in frequent personal contact.

YEARS OF EDUCATION: (Continuous independent variable) Full-time education completed by the respondent.

GDP: (Continuous independent variable) Gross Domestic Product[2] of the country by year

[2] https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

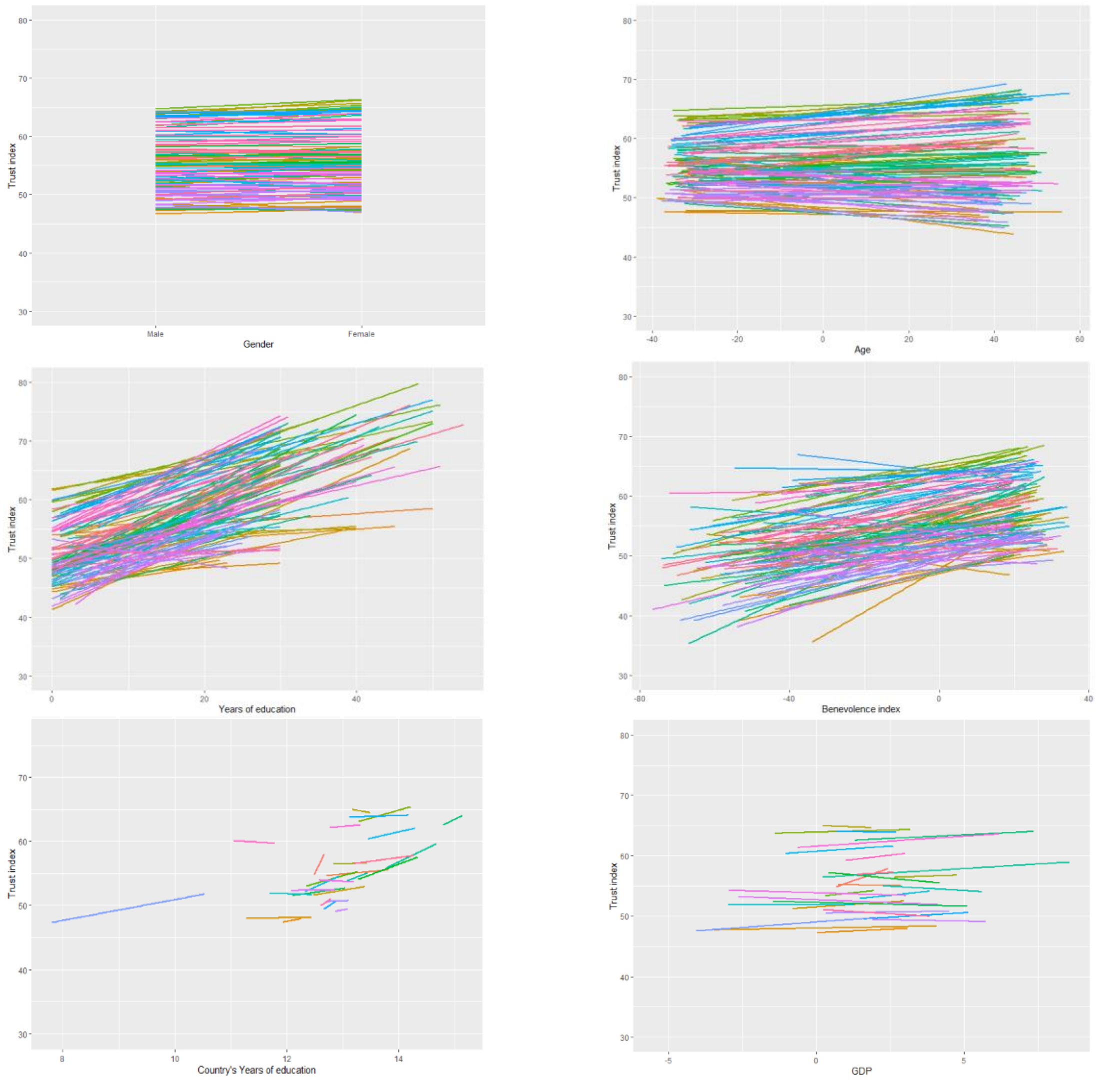

Figure 2: Relation among response variable and predictors variables

ANALYSIS

Multilevel models, also known as random effects, hierarchical, or mixed models, are regression models for the analysis of hierarchical data. Applications to two types of data are particularly common in the social sciences: **panel data (1)**, where measurement occasions are nested in persons or some other unit of analysis; and **hierarchical data (2)** where the primary units of analysis are nested in higher-level groups (Schmidt-Catran, Fairbrother, and Andreß 2019). Figure 3 explicitly indicates the structure of the model where ESS round level could be considered as level in the model or as a random slope.

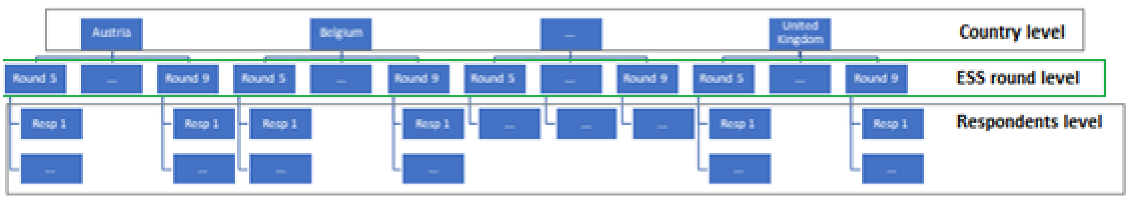

Figure 3: Structure of ESS survey

A multilevel model for continuous dependent variables is a generalization of the linear regression model, which includes a separate error component at each of its levels. In Equation 1 and 2, y indicates the response variable and x_1 to x_6 the predictors (independent) variables, β_1 to β_6 indicates their respective fixed effect.

In Equation 1, a **two-level panel data model** is expressed, indexes *i* indicates the individuals and *j* the countries. The intercept (β_{0ij}) include a fixed effect (β_0), random effects at country level (u_{0j}) and random residual at individual level (e_{0ij}). Same estimators are estimated for Benevolence index (β_{3ij}) and the time variable (β_{7ij}) because are included in the model as well as fixed and as random effects. As two levels are modelled, two variance-covariance matrices are calculated, each with three random slopes.

$$y_{ij} = \beta_{0ij}x_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_{3ij}x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij} + \beta_{7ij} time_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{3ij} = \beta_3 + u_{3j} + e_{3ij} \tag{1}$$

$$\beta_{7ij} = \beta_7 + u_{7j} + e_{7ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{3j} \\ u_{7j} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & & \\ \sigma_{u03} & \sigma^2_{u3} & \\ \sigma_{u07} & \sigma_{u37} & \sigma^2_{u7} \end{bmatrix} \qquad \begin{bmatrix} e_{0ij} \\ e_{3ij} \\ e_{7ij} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} & & \\ \sigma_{e03} & \sigma^2_{e3} & \\ \sigma_{e07} & \sigma_{e37} & \sigma^2_{e7} \end{bmatrix}$$

Equation 1: Two-level panel data model

In Equation 2, **a three-level model** is used where indexes *i* indicates the individuals, *j* the rounds (waves) and *k* the countries. Only one random slope besides intercept is included, Benevolence index, in this case, the random effects are considered at country level (v_{0k}) and round level (u_{0jk}) and a random residual at individual level (e_{0ijk}). Consequently, three variance-covariance matrices are computed each for two random slopes.

$$y_{ijk} = \beta_{0ijk}x_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_{3ijk}x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + \beta_6 x_{6ijk}$$

$$\beta_{0ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\beta_{3ijk} = \beta_3 + v_{3k} + u_{3jk} + e_{3ijk} \tag{2}$$

$$\begin{bmatrix} v_{0k} \\ v_{3k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} & \\ \sigma_{v03} & \sigma^2_{v3} \end{bmatrix} \qquad \begin{bmatrix} u_{0jk} \\ u_{3jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u03} & \sigma^2_{u3} \end{bmatrix} \qquad \begin{bmatrix} e_{0ijk} \\ e_{3ijk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \sigma^2_{e0} & \\ \sigma_{e03} & \sigma^2_{e3} \end{bmatrix}$$

Equation 2: Three-level model

In both models, random effects at each level are assumed to be normally distributed with a mean of zero and a constant variance and to be uncorrelated with each other and with the observed variables.

The share of the total unexplained variance attributable to the higher level, called intra-class correlation is calculated for each model using $ICC = \sigma_u/(\sigma_u + \sigma_e)$ for (1) and $ICC = \sigma_v/(\sigma_v + \sigma_u + \sigma_e)$ for (2). This coefficient indicates what proportion of the overall variance is at the country level, if this value were close to zero, there would be no intercountry differences to explain, and a multilevel model would not be necessary.

As the models studied include not only fixed effects but random effects as well, estimation is relevant. Maximum Likelihood Estimators (MLE) provide estimates that maximize the so-called likelihood function. That is, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameter. The literature states that two models with nested fixed effects can be compared using likelihood-based methods if they are fit using FIML (not REML). In contrast, models with the same fixed effect structure but different random effects should be compared by fitting with REML. (Jones and VS Subramanian 2019) For this reason, estimation for fixed effects will be obtained using Full Information Maximum Likelihood (FIML), which is known to produce biased estimates of the variance terms, the advantages of this approach are that it is more efficient in computational time and it is possible to use the associated Deviance in significance testing in a Likelihood ratio test. For estimating random effects, Restricted (or Residual) Maximum Likelihood (REML) will be used, which involves modifying the maximum likelihood estimation procedures to take account of the sampling variation in the fixed effects when producing the variance estimates.

In multilevel models it is important to make a special difference in order to estimate correctly compositional effects, two specifications are possible. Within-between specification conduct an indirect estimation of the contextual effect by deducting from the within effect the between effect ($\beta_c = \beta_w - \beta_b$). This is obtained by including the within variable centred to the group mean (β_w) and the between variable as the group mean of the variable (β_b). The other specification is called Mundlak's formulation, that produces a direct estimation of the contextual effect in the between level by including in the model the within variable not centred (β_w) and in the between level the group mean of the variable (β_c).

In both cases the within-effect (β_w) is the expected difference in the response for two participants that belong to the same group but differ by one on the individual predictor; the between-effect (β_b) is the expected difference in the mean response of two countries which differ by one in the mean of the group variable. The contextual effect (β_c) is the expected difference in the response for a participant who have the same individual value on the level 1 predictor but who belong to countries that differ by one in their group mean; this is the potential differential effect on the response from belonging to groups or contexts with different means.

In this study the direct calculation (Mundlak's formulation) will be used to estimate compositional effects. The analysis was conducted using MLwiN software and following directions from (Jones and VS Subramanian 2019).

RESULTS

Model 0: Random intercept models

To validate that a multilevel model should be applied, the null model should indicate that enough variance exists in the nested levels of the data, in this case, between countries *and between rounds*.

Results in Table 1 indicates that the total variance of the response variable Social Trust can be separated into between countries and within participants, this means, Intraclass correlation is 22.7% for the two-level model. When rounds are considered as an intermediate level, 22.4% of the total variance is at country level and 0.8% at the round level.

Table 1: Estimates for two and three-level M0 models

	Two-level M0			Three-level M0		
Fixed Part						
Intercept	51.712	(1.55)	***	51.742	(1.54)	***
Random Part						
Level: Country						
Var (Intercept)	73.896	(18.78)	***	72.417	(18.61)	***
Level: ESS round						

	Two-level M0			Three-level M0		
Var (Intercept)				2.477	(0.4)	***
Level: Participants						
Var (Intercept)	250.224	(0.82)	***	248.37	(0.82)	***
Units						
Level: Country	31			31		
Level: ESS Round				117		
Level: Participants	184921			184921		
-2*loglikelihood:	1546209.791			1545074.35		

For the two-level model, when time is included as a random slope ($M0_1$ in Table 2), 21.4% of the total variance is accounted at the country level. Slope corresponding to ESS round (recoded as time) is significant being considered as fixed and random effect in both levels. Differences between country level can be explained in 3% by differences in the ESS rounds (1 - (71.71/73.9)). For understanding the low percentage of explained variance for the rounds, categorical time slopes were studied $M0_2$; $M0_3$, with this it is possible to see that coefficients for first three rounds included in the model (5th, 6th and 7th rounds) are not significantly different from 0, just starting from round 8th, waves will make a difference in the Social Trust index.

Table 2: Estimates of time slopes in a Two-level model

	Two-level $M0_1$ with random time slope			Two-level $M0_2$ with Categorical fixed time slopes			Two-level $M0_3$ with categorical random time slopes		
Fixed Part									
Intercept	50.463	(1.55)	***	50.973	(1.56)	***	51.018	(1.59)	***
ESS round (time)	0.684	(0.14)	***						
ESS round (6)				-0.015	(0.11)		0.12	(0.37)	
ESS round (7)				0.309	(0.12)	**	0.105	(0.38)	
ESS round (8)				1.265	(0.12)	***	1.16	(0.42)	***
ESS round (9)				2.851	(0.13)	***	2.785	(0.47)	***
Random Part									
Level: Country									
Var (Intercept)	74.107	(18.93)		74.827	(19.02)		77.623	(19.88)	
Var (time)	0.528	(0.15)							
Var (essround_6)							3.149	(0.98)	
Var (essround_7)							2.938	(1)	
Var (essround_8)							3.904	(1.27)	
Var (essround_9)							4.912	(1.6)	
Level: Participant									
Var (Intercept)	258.794	(1.69)		249.315	(0.82)		248.37	(0.82)	
Var (time)	4.01	(0.5)							
-2*loglikelihood:	1545255.363			1545537.293			1544882.961		

Even though by pooling multiple waves of comparative survey data, the statistical power can be increased and it is possible to test hypotheses via within-specifications, this would not be possible if the variables of interest do not change over time. Similarly, using levels becomes less useful if the variables of interest change only marginally. In that case, the available variance to identify the effect in that level is small and the estimates will be imprecise. The following results will be focused on a comparison of a two-level model with continuous random time slope and a three-level model. With this, it would be possible to identify differences for both models considering if the low amount of variance captured by the waves in the response variable validates the not-use of a three-level model.

Model 1: Random intercepts with level 1 predictors centred on the group mean

Variables at the lowest level (participants) selected were centred to the group mean, i.e. the mean of the country/round group was subtracted. Estimates for Age of participants, Gender (Male as reference), Years of education and Benevolence index are significant[3] as fixed effects. Estimates are not different between both models.

In the two-level and three-level model country and round levels variances (when applied) are not reduced, this is called a compositional effect which means that there is substantial variance left that is due to country-level effects. If most of the variance between countries could be explained by compositional effects, we would have to conclude that any differences between countries are not related only to characteristics of the individuals that constitute the populations of these countries (Schmidt-Catran, Fairbrother, and Andreß 2019).

Table 3: Estimates for level 1 predictors for model with two and three levels

	Two-level M1			Three-level M1		
Fixed Part						
Intercept	50.341	(1.53)	***	51.617	(1.54)	***
Age	0.022	(0.002)	***	0.021	(0.002)	***
Gender (Female)	0.228	(0.07)	***	0.236	(0.07)	***
Years of education	0.614	(0.01)	***	0.613	(0.01)	***
Benevolence index	0.151	(0.003)	***	0.151	(0.003)	***
ESS round (time)	0.685	(0.14)	***			
Random Part						
Level: Country						
Var (Intercept)	71.792	(18.34)	***	72.499	(18.64)	***
Var(time)	0.508	(0.15)	***			
Level: ESS Round						
Var (Intercept)				2.486	(0.4)	***
Level: Participant						
Var (Intercept)	249.687	(1.63)	***	239.62	(0.79)	***
Var(time)	4.052	(0.49)	***			
-2*loglikelihood:	1538639.203			1538446.819		

Model 2: Random intercepts with level 1 predictors centred on the group mean and level 2 predictors, compositional effect.

By adding the predictors at the highest level, Country's years of education and Social spending as GDP index, fixed effects are significant. Country-level unexplained variance from previous model (M1) is reduced in 17.4% in the two-level model and 25.1% in the three-level model, meanwhile round (waves) level variance is reduced by 33.9% in the three-level model.

Country's years of education by rounds was included as country-level effect, which after controlling for compositional effects can be interpreted as contextual effect. As suggested for cross sectional data, Mundlak's formulation was used, with this, contextual effect is calculated directly in Table 4, being small for the two-level model 0.67 against 1.69 in the three-level model. Intercept coefficient is also being affected by the inclusion of compositional effects, going from 50 to 34 and 52 to 22 in M2, two-level and three-level model, respectively. The remaining estimates continue similar to the previous model.

[3] Significance codes: *** < 0.01; ** < 0.05

Table 4: Estimates for predictors at level 1 and level 2 with two and three level models

	Two-level M2			Three-level M2		
Fixed Part						
Intercept	**34.248**	**(3.11)**	***	**21.969**	**(4.603)**	***
Age (group centred)	0.022	(0.002)	***	0.021	(0.002)	***
Female (not centred)	0.227	(0.07)	***	0.237	(0.07)	***
Years of education (not centred)	0.614	(0.01)	***	0.613	(0.01)	***
Benevolence index (group centred)	0.151	(0.003)	***	0.151	(0.003)	***
Country's years of education (group centred)	**0.665**	**(0.22)**	***	**1.689**	**(0.35)**	***
Gross Domestic Product index (group centred)	**0.129**	**(0.03)**	***	**0.174**	**(0.07)**	**
ESS round (time)	0.447	(0.14)	***			
Random Part						
Level: Country						
Var (Intercept)	59.33	(15.17)	***	54.329	(13.92)	***
Var(time)	0.455	(0.13)	***			
Level: ESS Round						
Var (Intercept)				1.643	(0.28)	***
Level: Participant						
Var (Intercept)	249.552	(1.63)	***	239.62	(0.79)	***
Var(time)	4.018	(0.49)	***			
-2*loglikelihood:	1538589.688			1538402.171		

Model 3: Random intercept and random slopes with level 1 predictor centred on the group mean and level 2 predictors with contextual effect.

As evidenced in Figure 2, the relation of the independent variables with the response variable suggest a linear relation but that differs in slope, the relation of Benevolence index is more clear to vary over time and countries, for this reason is included in the model as a random slope. The incorporation of this random slope allows each country regression to have a different slope by round(wave), this means that the explanatory variable would have different effect for each group. Correlation between random effect of Benevolence index and Intercept in the two-level model was fixed to 0, as the relation was not significant. Variances for this effect is significant at all levels. Fixed estimates coefficients remain significant.

Table 5: Estimates for random effects, for two and three level model

	Two-level M3			Three-level M3		
Fixed Part						
Intercept	**34.408**	**(3.12)**	***	**22.730**	**(4.564)**	***
Age (group centred)	0.021	(0.002)	***	0.021	(0.002)	***
Female (not centred)	0.202	(0.07)	***	0.209	(0.072)	***
Years of education (not centred)	0.615	(0.010)	***	0.614	(0.010)	***
Benevolence index (group centred)	0.151	(0.011)	***	0.149	(0.010)	***
Country's years of education (group centred)	**0.650**	**(0.222)**	***	**1.628**	**(0.343)**	***
Gross Domestic Product index (group centred)	**0.119**	**(0.027)**	***	**0.175**	**(0.073)**	**
ESS round (time)	0.468	(0.14)	***			
Random Part						
Level: Country						
Var (Intercept)	62.575	(15.99)	***	55.045	(14.13)	***
Var (time)	0.466	(0.136)	***			
Var (Benevolence index)	0.003	(0.001)	***	0.002	(0.001)	***
Level: ESS Round						
Var (Intercept)				1.625	(0.272)	***
Var (Benevolence index)				0.004	(0.001)	***
Level: Participant						
Var (Intercept)	232.848	(1.683)	***	223.488	(0.917)	***
Var (time)	3.823	(0.479)	***			
Var (Benevolence index)	0.119	(0.005)	***	0.116	(0.005)	***
-2*loglikelihood:	1537435.968			1537131.105		

Likelihood ratio test, calculated by subtracting one deviance (loglikelihood) from another that is equivalent to divide one likelihood from another, indicates that there is not a statistically improvement in goodness of fit (p-value=0). This means either model would be sufficient depending on the objective of the task, although if the goal is to study contextual by compositional effects, considering a three-level model would be recommended as this estimate correctly the slope considering the time variant.

Residual Diagnostic

Caterpillar plots are shown in Figure 4, highest level residuals (country effects) indicates that in the two-level model, Serbia has the lowest residual for intercept estimates meanwhile in the three-level model Slovakia is the lowest. Denmark is the highest in both models at the country level. For the Benevolence index, in the two-level model, Greece has the highest residual and Slovakia the lowest, in the three-level model France is the lowest and Finland the highest, in this model standard errors are consistently bigger for every country than in the two-level model. For time estimates, in the two-level model some countries' standard error of residuals are higher than the rest, this may be due to few rounds where they participate, the highest residual for this estimate is Austria and the lowest is Spain.

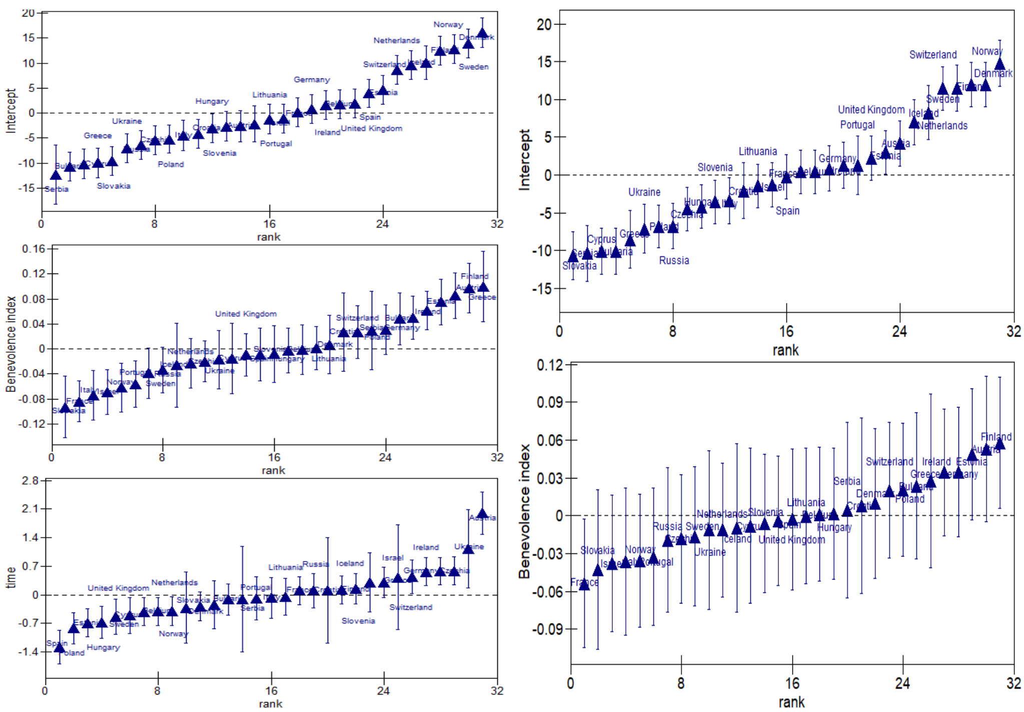

Figure 4: Residuals with 1.96 CI, random slopes, for two-level model (left) and three-level model (right).

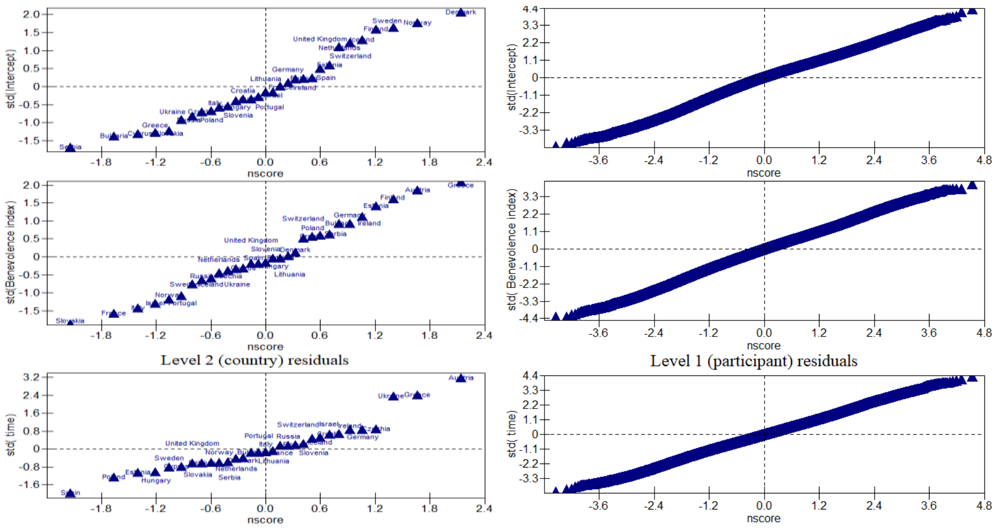

Figure 5: Normality of residuals for two-level model

Multilevel models assume that the distribution of residuals at each level come from a multivariate Gaussian distribution. Graphically is possible to identify univariate normality, Figure 5 and Figure 6, show the distribution of the residuals at each level in two and three-level models respectively. At country level, tails of the intercept distribution are not completely lineal in both models, this may be due to outliers or model misspecification, at the round level and participant level can be considered as normally distributed for all random slopes.

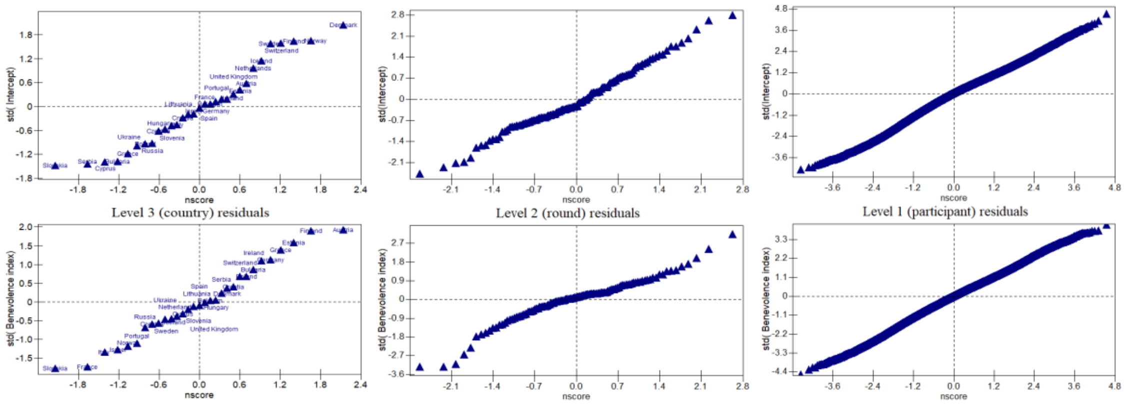

Figure 6: Normality of residuals for three-level model

Plots for heteroscedasticity, residuals versus predicted values does not show a specific structure, and the values are inside the critical region to be considered non-heteroscedastic in all level of the models studied, as shown in Figure 7. In both models, Portugal has the lowest prediction value followed by Greece and Iceland has the highest prediction value, followed by Ireland.

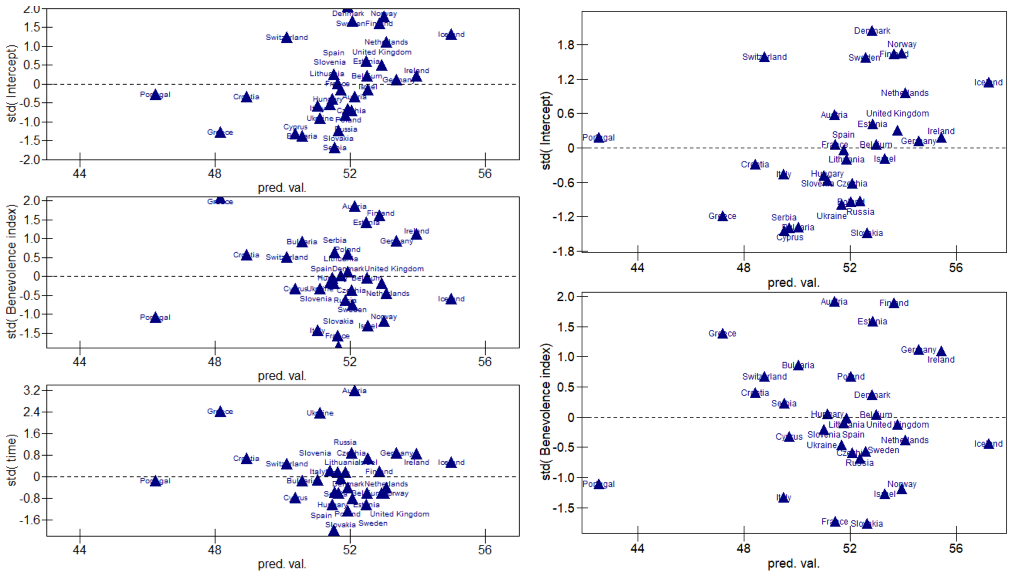

Figure 7: Residuals (standardized) vs predicted values at highest level for two-level model (left) and three-level model (right).

In conclusion, both models behave similarly considering estimates for fixed and random effects in the individual level. The compositional effect is influenced by the numbers of levels included in the model; this also apply to the higher-level estimations. Considering prediction, even though estimation differs, the order of the predictions is the same in both models. Even though there is not much significant difference among rounds particularly before 7th round, it is not recommended to avoid the use of rounds as a level in a multilevel model as still some variance is captured and explained.

REFERENCES

Breyer, B. 2015. 'Social Trust Scale (ESS)'. *ZIS - The Collection Items and Scales for the Social Sciences*. https://doi.org/10.6102/ZIS235.

European Social Survey. 2018. 'ESS-9 2018 Documentation Report. Edition 1.2'. Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC.

Fairbrother, Malcolm. 2014. 'Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets'. *Political Science Research and Methods* 2 (1): 119–40. https://doi.org/10.1017/psrm.2013.24.

Jones, Kelvyn, and VS Subramanian. 2019. *Developing Multilevel Models for Analysing Contextuality, Heterogeneity and Change Using MLwiN 3, Volume 1 (Updated March 2019); Volume 2 Is Also on RGate*. Centre for Multilevel Modelling, University of Bristol. https://doi.org/10.13140/2.1.2069.4727.

Schmidt-Catran, Alexander W., Malcolm Fairbrother, and Hans-Jürgen Andreß. 2019. 'Multilevel Models for the Analysis of Comparative Survey Data: Common Problems and Some Solutions'. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 71 (S1): 99–128. https://doi.org/10.1007/s11577-019-00607-9.

Schwartz, Shalom. 2003. 'A Proposal for Measuring Value Orientations across Nations'. *Questionnaire Package of ESS*, 259–90.

APPENDIX

Table 6: Sample size of 5 rounds in the ESS by country

	5th round (2010)	6th round (2012)	7th round (2014)	8th round (2016)	9th round (2018)
Austria	-	-	1795	2010	2499
Belgium	1704	1869	1769	1766	1767
Bulgaria	2434	2260	-	-	2198
Croatia	1649	-	-	-	-
Cyprus	1083	1116	-	-	781
Czechia	2386	2009	2148	2269	2398
Denmark	1576	1650	1502	0	0
Estonia	1793	2380	2051	2019	1904
Finland	1878	2197	2087	1925	1755
France	1728	1968	1917	2070	2010
Germany	3031	2958	3045	2852	2358
Greece	2715	-	-	-	-
Hungary	1561	2014	1698	1614	1698
Iceland	-	752	-	880	-
Ireland	2576	2628	2390	2757	2216
Israel	2294	2508	2562	2557	-
Italy	-	960	-	2626	2745
Lithuania	1677	2109	2250	2122	-
Netherlands	1829	1845	1919	1681	1673
Norway	1548	1624	1436	1545	1406
Poland	1751	1898	1615	1694	1500
Portugal	2150	2151	1265	1270	-
Russia	2595	2484	-	2430	-
Serbia	-	-	-	-	2043
Slovakia	1856	1847	-	-	-
Slovenia	1403	1257	1224	1307	1318
Spain	1885	1889	1925	1958	-
Sweden	1497	1847	1791	1551	-
Switzerland	1506	1493	1532	1525	1542
Ukraine	1931	2178	-	-	-
United Kingdom	2422	2286	2264	1959	2204

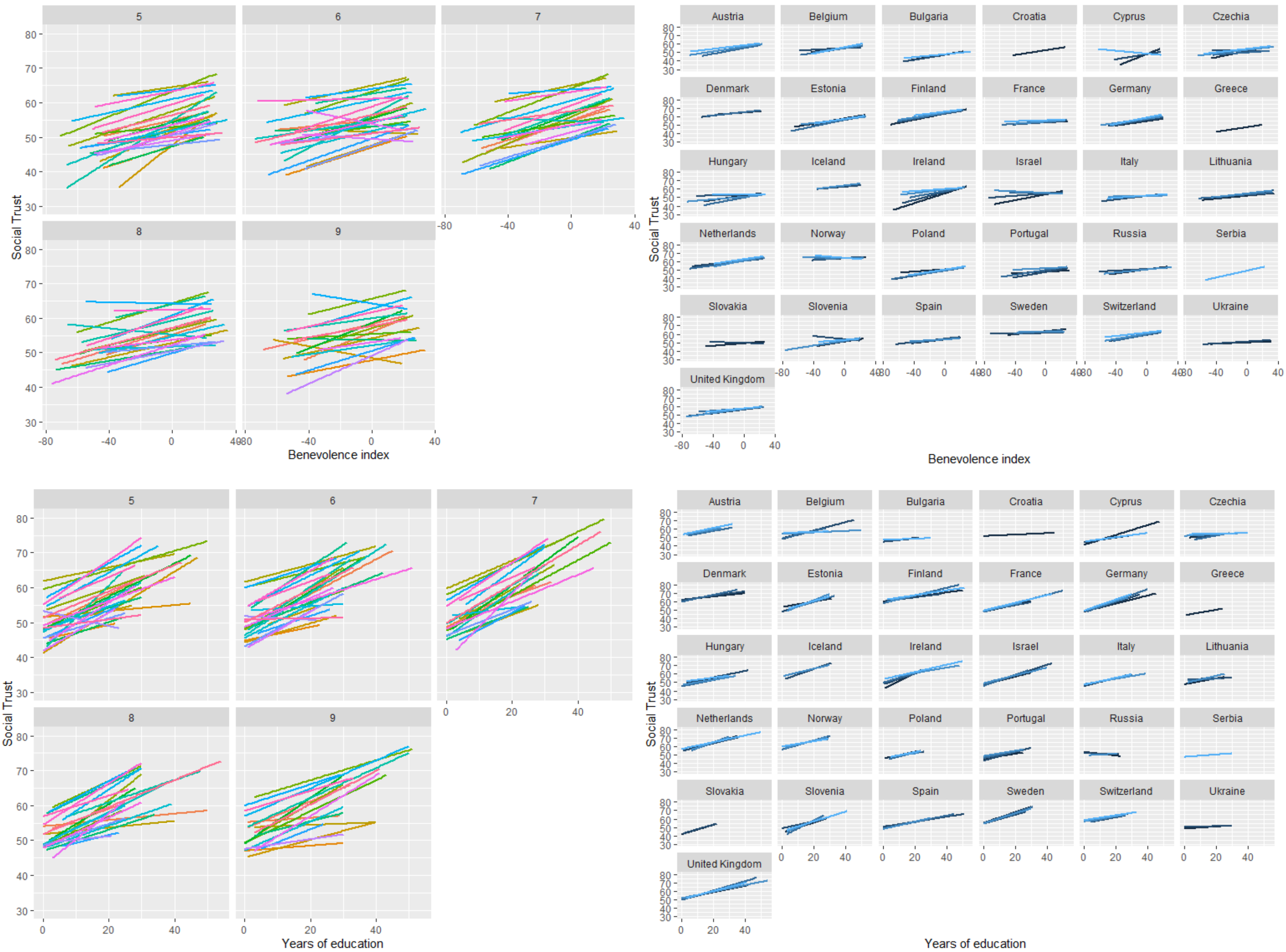

Figure 8: Relation years of education and benevolence index by round and country