VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

STRUCTURAL EQUATION MODELLING OF SOCIAL CAPITAL

Master's Thesis

Author: Agnė Klimavičienė
VU email address: agne.kavaliauskaite@mif.stud.vu.lt
Supervisor: Prof. Vydas Čekanavičius

Vilnius

2024

Abstract

Social capital holds a fundamental role in shaping societies and influencing various outcomes, encompassing economic development, political stability, and individual well-being. This master's thesis aims to construct a comprehensive measurement instrument of global social capital, leveraging extensive datasets from the World Values Survey and World Development Indicators. The study establishes a robust measurement framework, differentiating between cognitive and structural components, discerning the emotional and behavioral elements shaping social relationships. The findings unveil that higher levels of education and engagement in collective work or study environments are associated with increased social capital. While social capital generally escalates with age, there appears to be a potential decline in structural social capital due to reduced energy or time available for group activities and political engagement. Notably, control of corruption emerges as a pivotal country-specific factor influencing social capital. The findings of this research complement existing studies by revealing nuanced influences on social capital. Also, the comprehensive analysis of global data sets this study apart, as many previous analyses tend to focus on regional research.

Keywords: Social Capital, SEM, MIMIC, WVS, WDI.

1 Introduction

Social capital, an extensively studied concept, holds a range of interpretations across disciplines. It is understood as the inherent capacity of individuals to collaborate and participate in collective civic activities. Social capital is also perceived as an accumulated asset, deriving from a sense of belonging, shared norms, and mutual trust among individuals or within a community. This multifaceted concept garners interest from disciplines spanning sociology, economics, political science, and public health, each exploring its influence on societal dynamics, economic prosperity, governance, and community well-being.

Social capital is a crucial factor in understanding societal interactions and dynamics. Despite its widespread adaptation, challenges have surfaced due to its broad usage and lack of precision. This issue often arises from oversimplification, relying on singular indicators or limited proxies, overlooking its inherently complex nature. Further complexity arises from the absence of a universally agreed-upon definition, leading to diverse interpretations across disciplines.

This research aims to approach various aspects of social capital systematically. Beginning with a short literature review, the conceptualizations, dimensions, and determinants of social capital observed in previous studies are explored. This serves as a foundation to explore the multifaceted aspects of social capital, particularly its cognitive and structural components, which encompass both emotional and behavioral elements. To capture these dimensions, indicators are constructed utilizing data from the World Values Survey and World Development Indicators datasets. These global datasets provide a comprehensive understanding of social capital that extends across diverse populations and contexts worldwide.

Once a comprehensive measurement instrument is constructed, the study dives into examining patterns of social capital, both within and between countries. By incorporating socio-demographic factors (age, gender, education, employment, religion, and marital status) and contextual indicators like Corruption Control, GINI Index, Poverty Gap, Unemployment, and GDP per capita, the research enriches the understanding of how these elements contribute to social capital. The study seeks to provide meaningful insights into the factors that contribute to social capital, aiming to enhance the understanding of how these elements impact the shaping of societies on a global scale.

2 Short Literature Review

Social capital is a complex and multifaceted concept that has been studied extensively in social science research. As defined by Robert Putnam (1995) [1] social capital refers to "the networks, norms, and trust that facilitate coordination and cooperation among individuals and groups." However, the definition of social capital extends beyond this basic understanding. According to Norman Uphoff (2000) [2], social capital is "an accumulation of various types of social, psychological, cognitive, institutional, and related assets that increase the amount (or probability) of mutually beneficial cooperative behavior". This definition highlights the multifaceted nature of social capital, as the diverse assets that promote cooperative behavior are shaped by the specific social contexts and populations they are embedded in. Consequently, measuring and understanding social capital requires a nuanced approach that recognizes the contextual and dynamic nature of its forms and distribution.

Onyx and Bullen (2000) [3] conducted a study aiming to find the key components of social capital. In a study conducted in five Australian communities, a questionnaire containing 68 potential social capital items was created. The data collected was subjected to a hierarchical factor analysis, which identified a general underlying factor and eight specific factors that accounted for 49% of the variance. Those factors included: Participation in the Local Community, Social Agency, or Proactivity in a Social Context, Feelings of Trust and Safety, Neighborhood Connections, Family and Friends Connections, Tolerance of Diversity, and Value of Life and Work Connections. The study's findings align with Robert Putnam's [1] theory that social capital is composed of three key components: social networks, trust, and civic engagement. These components emerged as significant determinants in the analysis, providing empirical support for Putnam's conceptualization of social capital.

Wim van Oorschot et al. (2006) [4] also used the same three dimensions of networks, trust, and civism to analyze the multidimensional nature of social capital in Europe. The study found that overall social capital levels do not differ significantly among European countries and regions, except for higher levels in Scandinavia. The accumulation of social capital was also observed, as people with more human and economic capital had higher levels of social capital. The study also revealed gender differences in the structure of social capital, with women having higher levels of trustworthiness and family orientation but less political engagement and involvement in voluntary associations. Politically, the study found that leftist individuals tend to trust others more and have higher levels of network capital, while religious individuals tend to have higher levels of social capital. Furthermore, the study showed that personal characteristics such as education and income have a greater impact on social capital than geographic location within Europe.

Lu and Zhang (2019) [5] explored the multifaceted nature of social capital, differentiating between cognitive and structural components. Cognitive social capital, shaped by trust and reciprocity, was assessed subjectively. Structural social capital relied on objective social indicators. Employing Structural Equation Modeling (SEM), authors estimated parameters for the proposed model. The research revealed that cognitive social capital significantly influences self-rated health, while structural social capital affects it indirectly.

Fidrmuc and Gërxhani (2005) [6] analyzed the determinants of social capital in a sample of 27 European countries, including EU member states and countries that were candidates for membership. The study found differences in social capital between the two groups of countries. The research revealed that the individual factors influencing social capital were largely similar in both candidate and EU member countries, meaning that the disparity in social capital can be attributed to aggregate factors specific to each country. Lower economic development, institutional quality and corruption in candidate countries resulted in lower levels of social capital.

A similar analysis was done by Kaasa and Parts (2008) [7]. In this research, 12 indicators of 5 dimensions (formal networks, informal networks, general trust, institutional trust, and norms) were used and all dimensions were analyzed separately. Moreover, countries were grouped into three clusters to make it easier to compare those that underwent the transition from communism and those that did not. Age, education, and religiosity were found to be the most influential determinants of social capital, while town size and number of children had the least impact. The study also revealed that the effects of determinants vary significantly within country subgroups, with religiosity having an effect on informal networks in northern countries but not in other country groups. The study found that the origins of social capital vary significantly between countries that have communist backgrounds and no communist backgrounds. These findings suggest that it is essential to recognize that both individual-level and macro-level determinants play important and complementary roles in shaping social capital.

Christoforou (2011) [8] conducted a study that focuses on group membership as a form of social capital and examines its determinants across European countries. Binary logistic regression models are applied to the data to explore the impact of individual and aggregate factors on group membership. The study confirmed that both micro (income, education, gender, age, marital status, employment) and macro level factors (GDP per capita, income inequality, social trust, trust in public institutions, corruption, unemployment, and violation of political and civil rights) play a role in determining social capital.

Ferragina (2013) [9] proposed a model that takes into account structural socio-economic factors and historical legacies in different regions of Europe to measure the effect of income inequality, economic development, labor market participation, and national divergence on social capital. The study found that income inequality, labor market participation, and national divergence explain most of the variance in social capital, with income inequality being the most significant predictor for formal social networks and social trust. Bundi and Freitag (2020) [10] investigated the impact of economic hardship on social capital in 27 European countries. The results showed that personal resources such as education and money are essential for civic engagement. Economic hardship negatively affects citizens' likelihood of engaging in solidarity.

Social capital is an intricate concept that refers to the resources that arise from social connections and networks. There is no single measure that can capture all aspects of social capital, and different disciplines and research approaches may use different methods to measure it.

3 Data and Methods

3.1 Data

In the study the data from the World Values Survey (WVS) wave 7 [11] is used. The WVS is a global research initiative that investigates individuals' values, beliefs, and behaviors. Survey data collection was started in 2018 and completed in 2022. The surveys employed rigorous sampling techniques, utilizing random probability representative samples of the adult population and primarily conducted face-to-face interviews. This analysis focuses on data featuring insights from 153,716 respondents across 90 countries and territories. The dataset offers a comprehensive and diverse sample, providing opportunities to explore and analyze the dynamics of social capital.

For the more comprehensive analysis, the World Development Indicators (WDI) [12] data were incorporated. WDI constitutes the principal collection of development indicators curated by the World Bank, meticulously assembled from internationally recognized official sources. Specific indicators — Control of Corruption, GINI index, Poverty Gap, and Unemployment — were extracted and integrated into the existing dataset by country. To ensure coherence and consistency, these indicators were sourced from the year 2018, aligning with the initiation of the WVS project, thereby avoiding any temporal discrepancies.

3.2 Sample Characteristics

Table 1 represents the distribution of respondents across different continents in the dataset. Europe and Asia collectively constitute the majority of observations, comprising 75.76% of the sample, with Europe being the most represented continent. South America, North America, Africa, and Oceania exhibit smaller yet distinctive presences, showcasing varying sample sizes and country distributions across continents.

Continent	**Count**	**Percentage (%)**	**Countries**
Europe	66525	45.74%	37
Asia	43656	30.02%	28
South America	11613	7.99%	9
North America	11600	7.98%	6
Africa	9456	6.50%	8
Oceania	2578	1.77%	2

Table 1: Sample characteristics by continent

The characteristics of the respondents are presented in Table 2. Females constitute a slightly larger portion of the dataset, comprising 53.94%, compared to males, who account for 46.06%. The distribution of age categories in the dataset reveals a diverse representation of respondents. The education distribution in the dataset reveals that the majority fall into the middle education category (38.74%), followed by upper education (33.55%), and lower education (27.72%). The data illustrates a diverse range of employment statuses among the surveyed population, with approximately 38.11% in full-time employment, 17.86% retired or pensioned, and 11.23% self-employed. Notably, 62.96% of respondents identify as religious, while 28.35% do not consider themselves religious, and 8.69% explicitly identify as atheists. This comprehensive overview portrays a global dataset with diverse demographic and regional representations, underscoring the varied characteristics and geographical distribution of respondents.

	Count	**Percentage (%)**
Gender		
Female	78450	53.94%
Male	66978	46.06%
Age		
0-20	4834	3.32%
21-30	26400	18.15%
31-40	27927	19.20%
41-50	26048	17.91%
51-60	24220	16.65%
61-70	20818	14.31%
71-80	11285	7.76%
80+	3896	2.68%
Education		
Lower	40308	27.72%
Middle	56335	38.74%
Upper	48785	33.55%
Employment		
Full time	55423	38.11%
Retired/pensioned	25974	17.86%
Self employed	16331	11.23%
Housewife (not otherwise employed)	15054	10.35%
Unemployed	11536	7.93%
Part time	10977	7.55%
Student	7731	5.32%
Other	2402	1.65%
Religion		
A religious person	91562	62.96%
Not a religious person	41222	28.35%
A convinced atheist	12644	8.69%

Table 2: Sample characteristics by gender, age, education, employment, religion

3.3 Social Capital Measurement

The measurement model of social capital was based on an assumption that it consist of cognitive and structural components. Structural social capital refers to observable actions and connections, such as associational links or networks, while cognitive social capital refers to feelings, encompassing their values and perceptions [13]. Cognitive social capital was measured by subjective indicators - general trust, institutional trust and belonging. Structural social capital was measured by participation and political action, which is more objective indicators. Descriptive statistics of chosen social capital indicators are shown in table 3.

General trust was evaluated through a Likert sum scale, based on responses to six survey questions. Participants were asked to express their level of trust in people from various groups, using a scale where 1 represented "do not trust at all" 2 denoted "do not trust very much" 3 indicated "trust somewhat" and 4 signified "trust completely" The survey covered questions related to trust in family, neighborhood, people known personally, individuals met for the first time, people of another religion, and people of another nationality. Cronbach's Alpha, a measure of internal consistency, was calculated to assess the reliability of the general trust sum scale. Producing a value of 0.79, the measurement suggests an acceptable level of internal consistency.

Institutional trust was assessed using a Likert sum scale, with participants providing responses to ten survey questions. The questions prompted respondents to evaluate their confidence in various entities, including churches, the press, labor unions, parliament, civil services, political parties, major companies, the environmental protection movement, the justice system and courts, and the United Nations. Participants rated their confidence levels on a scale where 1 represented no confidence and 4 represented a great deal of confidence. The scale has an alpha reliability of 0.85, indicating of strong internal consistency.

Belonging was assessed using a Likert sum scale, derived from responses to five survey questions. Participants were requested to express their sense of proximity to the world, their continent, their country, county or region and their local area (village, town, or city). Using a scale ranging from 1 to 4, where 4 signifies a profound emotional closeness and 1 denotes a lack of closeness entirely, respondents assessed their emotional connection. The reliability of the belonging sum scale, as measured by Cronbach's Alpha, was found to be 0.79, indicating acceptable internal consistency.

Participation was evaluated by respondents indicating membership in various voluntary organizations. Participants were asked to specify their affiliation with organizations related to religion, education, arts, music, cultural activities, labor unions, political parties, environment, ecology, animal rights, professional associations, sports, recreation, consumer groups, humanitarian or charitable organizations, self-help or mutual aid groups and others. To quantify the participation, the sum of positive responses across these questions was calculated, representing the number of fields of voluntary organizations in which an individual is involved.

Political action was evaluated using a Likert sum scale, where participants indicated their engagement in specific activities, rating whether they have actually done any of these things, might do it, or would never do it. The assessed actions included signing a petition, joining in boycotts, attending lawful peaceful demonstrations, and joining unofficial strikes. The cumulative responses to these four questions were aggregated to form the political action scale, which demonstrated an alpha reliability of 0.80, indicating acceptable reliability.

	n	Min value	Max value	Mean	Standard deviation
General Trust	145428	0	24	16.22	3.42
Institutional Trust	145428	0	40	22.49	6.21
Belonging	145428	0	20	14.55	3.42
Participation	145428	0	11	1.62	2.33
Political Action	145428	0	12	6.31	2.41

Table 3: Descriptive statistics of Social Capital indicators

3.4 Methods

To estimate the parameters for the proposed model Structural equation modeling was employed. SEM enables the examination of complex relationships by incorporating measurement models for latent variables and structural models for relationships among these latent constructs [14]. This makes SEM suitable for exploring multifaceted concept, such as social capital, where cognitive and structural components intertwine.

Structural Equation Modeling is a statistical technique employed by researchers to understand complex relationships among multiple variables. SEM is a combination of multiple regression and factor analysis. Unlike multiple regression or analysis of variance, SEM allows for the simultaneous estimation of relationships among observed and unobserved variables. It enables researchers to model theoretical constructs, which might not be directly measurable, by using multiple observed indicators. SEM considers measurement error within observed variables, enhancing the precision of estimating relationships among latent constructs. This approach helps researchers delve into intricate causal pathways, networks of relationships, and interactions among variables within a single model, providing a more comprehensive understanding of the underlying phenomena in various scientific fields. SEM encompasses two main approaches: covariance-based SEM (CB-SEM), which evaluates how well a proposed model reproduces the observed data's covariance matrix, and partial least squares SEM (PLS-SEM), known for its focus on explaining variance in dependent variables and its utility in predictive modeling [15].

SEM involves two types of models: the measurement model addresses the relationship between latent variables and their indicators, and the structural model defines how different constructs within a model interrelate. Let y and x be vectors of the observed variables. The measurement model of SEM then is defined by [16]:

$$y = \Lambda_y \eta + \epsilon$$

$$x = \Lambda_x \xi + \delta$$

here

Λ_y is the (p x m) coefficients matrix relating y to η, and m is the number of elements of η,

Λ_x is the (p x m) coefficients matrix relating x to ξ, and m is the number of elements of ξ,

η is the (m x 1) vector of latent endogenous random variables,

ξ is the (n x 1) vector of latent exogenous random variables,

ϵ is the (p x 1) vector of measurement errors for y,

δ is the (q x 1) vector of measurement errors for x.

It is assumed that η, ξ and δ are random vectors with zero means; ϵ is uncorrelated with η, ξ and δ; and δ is uncorrelated with ξ, η and ϵ. All observed variables are measured in deviations from their mean.

Another component of SEM, structural model, is represented in the following manner:

$$\eta = B\eta + \Gamma\xi + \zeta$$

here
B is the (m x m) coefficients matrix for endogenous latent variables,
Γ is the (m x n) coefficients matrix for exogenous latent variables,
ζ is the (m x 1) vector of latent errors in equations.

Different approaches can be used for SEM estimation. In this research maximum likelihood (ML) method is used. ML uses derivatives to minimize the function:

$$F_{ML} = log|\Sigma(\Theta)| + tr(S\Sigma^{-1}(\Theta)) - log|S| - (p + q)$$

where $p + q$ is the number of variables, S - observed covariance matrix and Σ is the covariance matrix implied by the model [17].

The assumptions underlying Structural Equation Modeling encompass several critical considerations. Firstly, there is an expectation of multivariate normality within observed variables, essential for maximum likelihood estimation. However, this assumption might be challenged, especially when dealing with ordinal or discrete scales, indicated by skewness and kurtosis values. Latent variables' normal distribution, a key presumption, tends to be violated in practice. Handling missing data is critical, assuming no missing data in any variable or, if present, applying imputation techniques during ML estimation [18]. Ensuring the absence of outliers is essential as they negatively impact the model's significance. Moreover, multicollinearity among independent variables should be avoided, and adequate sample size should be ensured. SEM generally assumes no correlation between error terms, yet researchers might explicitly introduce correlations based on their conceptual models [19].

Multilevel Structural Equation Modeling integrates multilevel modeling and structural equation modeling, particularly in hierarchical data structures. Unlike conventional SEM, where latent variables and indicators are independent across units, multilevel SEM accounts for within-cluster dependence, common in nested unit settings like clusters within clusters. It addresses this by allowing latent or observed variables to vary at different levels, explaining correlations among lower-level units nested within higher-level units. In multilevel regression models, random effects signify unobserved heterogeneity across levels, while in SEM, latent variables reflect constructs underlying observed items or fallible measurements. Multilevel SEM is pivotal for valid statistical inference in hierarchically nested observations, especially when exploring relationships between variables measured by fallible instruments across nested clusters. This approach offers versatility in modeling, accommodating unbalanced data and missing values, while enabling investigations that traditional approaches may not validly address [20]. The model specification for two-level dataset:

$$y_W = \Lambda_W \eta_W + \epsilon_W$$

$$\mu_B = \mu + \Lambda_B \eta_B + \epsilon_B$$

here
μ_B represent the random intercepts of the variables y_W that vary across groups. The first equation addresses variations within groups, while the second equation accounts for variations between groups and group level means. After integrating these equations, the following is obtained:

$$Y_{ij} = \mu + \Lambda_W \eta_W + \Lambda_B \eta_B + \epsilon_B + \epsilon_W$$

here
μ denotes the vector of group level means,
Λ_W is the coefficients matrix at the within level,
Λ_B is the coefficients matrix at the between level,
ϵ_W is the residual errors at the within level,
ϵ_B is the residual errors at the between level.

The Multiple Indicator Multiple Causes (MIMIC) model is statistical technique used in structural equation modeling to examine relationships between latent variables and their indicators or observed variables. The model serves as a valuable alternative for testing measurement invariance and understanding population heterogeneity. Comprising a measurement model and a structural model, MIMIC examines relationships between latent variables and their observable indicators, alongside the direct impact of covariates representing group membership on factor means or item indicators [21]. The model specification for MIMIC model is:

$$\eta = B\eta + \Gamma\xi + \zeta$$

$$y = \Lambda_y \eta + \epsilon$$

$$x = \xi$$

The Multilevel Multiple Indicator Multiple Causes (ML-MIMIC) model is a specialized framework within structural equation modeling designed to assess relationships between observed variables and latent factors across multiple levels of analysis. It extends the MIMIC model to two or more hierarchical levels, allowing an examination of how these relationships vary within and between these levels. A critical assumption in ML-MIMIC models is the across-level invariance of factor loadings, ensuring that the same latent variable is effectively measured at both levels. Equal factor loadings across levels facilitate the comparison of latent factor variances, enabling the estimation of the proportion of variance attributable to between-patient differences, often quantified by the intraclass correlation (ICC). Overall, the ML-MIMIC model provides a robust framework to explore how latent constructs and observed variables interact across multiple levels of analysis, offering insights into within-group and between-group variations in these relationships [22].

Python package semopy and R package lavaan were used for the analysis. The semopy package is free and open-source package which surpasses the widely used R package lavaan in terms of performance, stability in the optimization process, and accuracy in parameter estimates [23]. However, this package does not offer the capability to create multilevel structural equation models. Consequently, the R package lavaan was utilized for this purpose. The lavaan package serves as a free, open-source tool offering high-quality capabilities for latent variable modeling. It allows for the estimation of diverse multivariate statistical models, spanning path analysis, confirmatory factor analysis, structural equation modeling, and growth curve models [24]. Both packages uses covariance-based SEM approach. CB-SEM fits the sample covariance matrix to align with the model's implied covariance matrix and then estimates parameters through maximum likelihood techniques.

The evaluation of model fit incorporated several indices: the chi-square test statistic, root mean square error of approximation (RMSEA), goodness of fit index (GFI), normed fit index (NFI) and standardised root mean square residual (SRMR). The determination of model fit was guided by Hooper, Coughlan, and Mullen's guidelines for structural equation modelling [25]. Consensus is lacking regarding an acceptable chi-square statistic ratio, with suggestions spanning from 2 to 5. For RMSEA, lower limits trending toward 0 and upper limits below 0.08 characterize well-fitting models. Traditionally, a GFI value above 0.90 signals acceptable fit, yet simulation studies advise 0.95 for low factor loadings and sample sizes. NFI values, ranging from 0 to 1, typically exceed 0.90 for a good fit, although recent recommendations advocate for a stricter criterion of NFI $\geq$ 0.95. SRMS values range from 0 to 1, where well-fitting models typically achieve values below 0.05. However, values up to 0.08 are considered acceptable.

4 Analysis and results

4.1 Measurement Model

Initially, the assumptions of SEM were assessed to ensure the validity of analytical approach. After testing multivariate normality, it became evident that variables distribution significantly deviates from a normal distribution. With a Henze-Zirkler (Hz) test statistic of approximately 19.44 and a corresponding p-value of 0.0, the data does not meet the assumption of normality.

Multicolinearity was assessed using Variance Inflation Factor (VIF). The VIF values for all variables were around 1, indicating an absence of significant multicollinearity concerns among the predictors. Upon identifying missing values within certain variables, observations containing these not-applicabe values were excluded, resulting in a dataset comprising 145,428 complete observations. Outliers were detected using the Interquartile Range (IQR), where values falling below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR were considered outliers. These outliers were subsequently replaced with mode values, given that all variables have discrete values, using the mean or median could potentially result in fractional values. Despite the deviation from multivariate normality the analysis proceeded, supported by the absence of significant multicollinearity concerns among the predictors, no missing data or outliers and a sufficient number of observations in dataset.

Table 4 displays the relationships between indicators of social capital. The correlation matrix illustrates the associations between various social capital indicators. Notably, Institutional Trust is moderately associated with General Trust, Belonging, and Political Action. General Trust shows similar associations with Belonging. Political Action has modest correlations with Institutional Trust and Participation, while Participation displays moderate associations with General Trust and Political Action.

	Institutional Trust	General Trust	Belonging	Political Action	Participation
Institutional Trust	1.0***				
General Trust	0.262***	1.0***			
Belonging	0.101***	0.129***	1.0***		
Political Action	0.065***	0.221***	0.022***	1.0***	
Participation	0.106***	0.137***	0.034***	0.237***	1.0***

Table 4: Correlation Matrix of social capital indicators. *** denotes $p < 0.001$

The created structural equation model scheme for social capital is shown in the Figure 1. It is measured by two latent variables: Cognitive Social Capital and Structural Social Capital. Cognitive Social Capital is influenced significantly by General Trust, Institutional Trust, and Belonging, with estimates of 2.411, 2.190, and 0.563 respectively. Structural Social Capital is influenced by Participation and Political Action, with estimates of 0.554 and 1.382 respectively. All estimates are statistically significant. Additionally, there is a notable bidirectional relationship observed between Cognitive Social Capital and Structural Social Capital, with an estimated coefficient of 0.479, indicating mutual influence between these constructs. The variances of the latent variables also demonstrate their individual significance within the model, highlighting the varying impacts of each factor on the overall social capital constructs. Goodness of fit results together with standardized estimates are shown in Table 5 and reveal a well fitting model. RMSEA = 0.042 (<0.05), GFI = 0.97 (>0.95), NFI = 0.97 (>0.95).

Goodness of fit statistics			
RMSEA	0.042		
GFI	0.97		
NFI	0.97		
Factor		Cognitive Social Capital	Structural Social Capital
Cognitive Social Capital	Institutional Trust	2.190***	
	General Trust	2.411***	
	Belonging	0.563***	
Structural Social Capital	Political Action		1,382***
	Participation		0.554***

Table 5: Goodness-of-Fit Statistics and Standardized Estimates in a Structural Equation Model

*Note: Reference categories in parentheses; significance: *=0.1, **=0.05, ***=0.01*

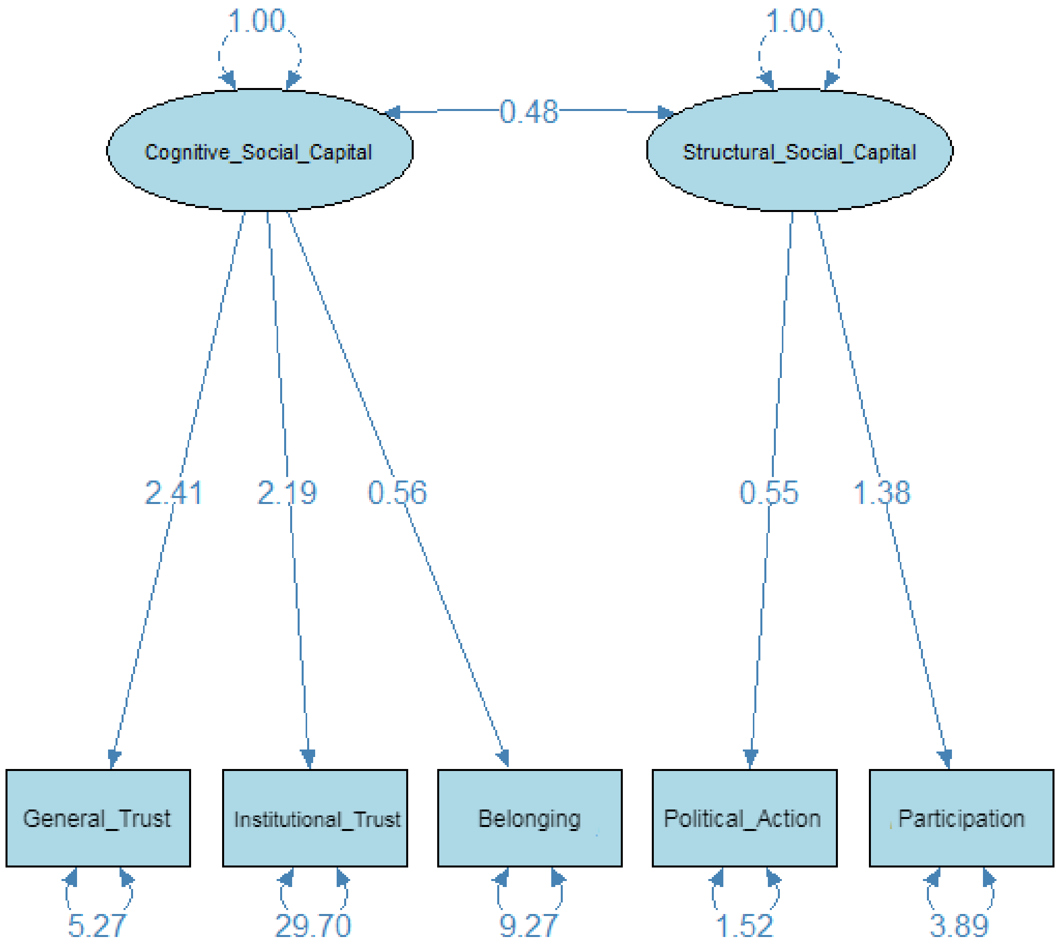

Figure 1: Structural equation model

4.2 Multilevel SEM model

It is important to not only to evaluate individual social capital indicators, but also to see how social capital operates within specific national contexts and how it varies across different countries.

After establishing the robustness of the factor structure at the individual level, examination of item variability employing Intraclass Correlation Coefficients (ICC) was performed. The ICC is the unexplained proportion of the variance by any predictors in the model, that can be attributed to the group level variable [21]. ICC is bounded between 0 and 1, where 0 represents no reliability within clusters and 1 indicates high dependency among clusters.

The results showed in table 6 indicate the amount of variance in each latent variable that can be attributed to differences between countries. Since ICC values range from 0.148 to 0.210, this indicates that between-level structure within the model is feasible.

Institutional Trust	General Trust	Belonging	Political Action	Participation
0.196	0.210	0.148	0.193	0.181

Table 6: Intraclass correlation results

Table 7 displays the results from MLSEM model. Overall, the model demonstrates a strong fit, indicated by an RMSEA of 0.02 and high GFI and NFI values, both at 0.99. The SRMR within the covariance matrix of 0.008, suggests a good fit. However, the SRMR between covariance matrices is relatively higher, meaning a higher level of discrepancy in the covariance matrices between the countries.

Institutional Trust, General Trust, and Belonging exhibit substantial positive impacts on Cognitive Social Capital within countries. However, the impact of Belonging appears minimal and not significant at the between level. Standardized parameter estimates of Cognitive Social Capital are lower at between level. Regarding Structural Social Capital, both Political Action and Participation exhibit similar standardized estimates at both within and between levels, showcasing substantial and statistically significant positive influences.

Goodness of fit statistics					
RMSEA	0.02				
GFI	0.99				
NFI	0.99				
SRMR (within)	0.008				
SRMR (between)	0.083				
Factor		Within level		Between level	
		Cognitive	Structural	Cognitive	Structural
Cognitive Social Capital	Institutional Trust	2.848***		0.856***	
	General Trust	1.533***		1.148***	
	Belonging	0.971***		0.001	
Structural Social Capital	Political Action		0.925***		0.942***
	Participation		0.445***		0.377***

Table 7: Standardized factor loadings

*Note: Reference categories in parentheses; significance: *=0.1, **=0.05, ***=0.01*

4.3 MIMIC Model with individual covariates

The subsequent phase of analysis involved employing Multiple Indicators - Multiple Causes models (MIMIC). These models, within structural equation modeling, serve as multivariate techniques aiming to determine how latent variables are influenced by relevant covariates [21]. Unlike a conventional SEM model that solely examines relationships between latent variables, the MIMIC model extends this framework by incorporating observed variables, allowing for a more comprehensive analysis.

A MIMIC model was employed to explore how different individual factors influence cognitive and structural social capital. Covariates such as age, gender, education, employment, religion and marital status were included in the analysis. Visual representation of the model is shown in Figure 2. The model results are presented in the Table 8. The goodness-of-fit outcomes indicate a strong fit within the model. The RMSEA value of 0.02 falls below the recommended threshold of 0.05, signifying a good fit. Additionally, the GFI value of 0.99 surpasses the acceptable level of 0.95, while the NFI value of 0.98 also exceeds the 0.95 threshold, both indicating robust model fitting.

Age exhibits a positive influence on both cognitive and structural social capital dimensions. Similarly, being male demonstrates an increase in both aspects of social capital. However, lower or middle education levels exhibit a negative impact on both cognitive and structural social capital.

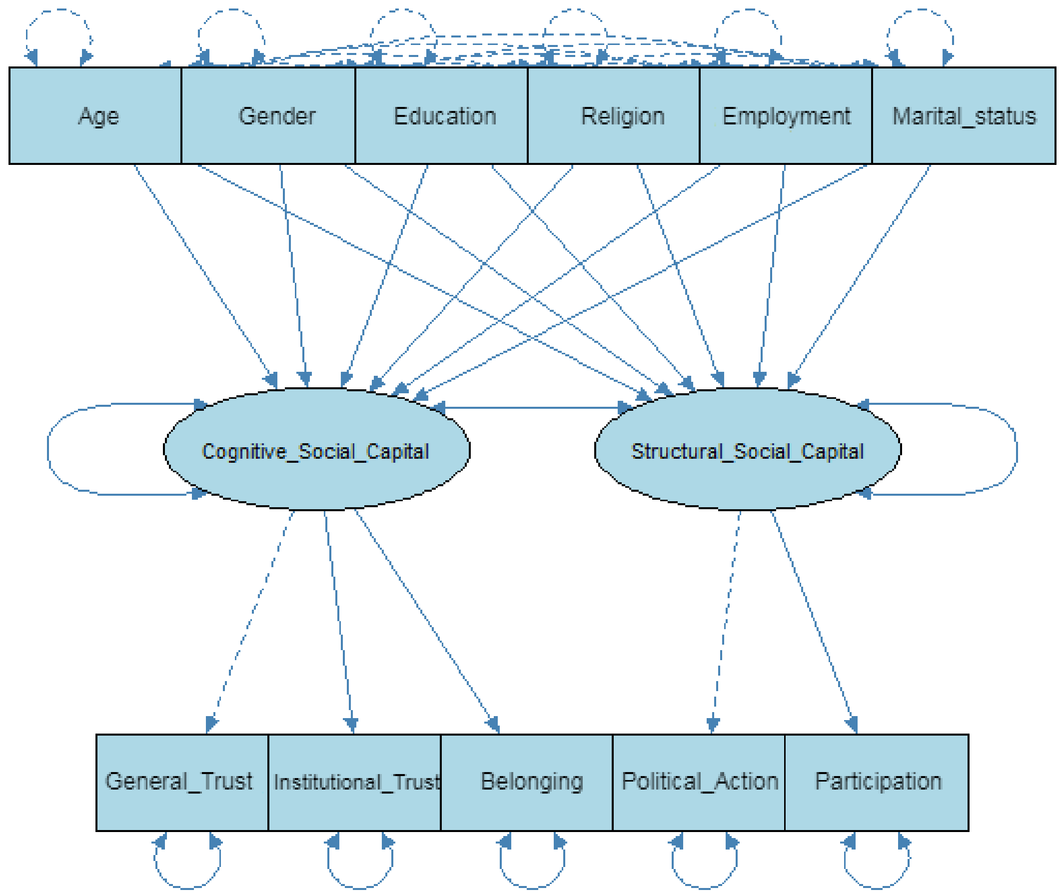

Figure 2: MIMIC model with individual covariates

Regarding employment status, individuals who are self-employed, housewives, or unemployed experience a negative effect on both dimensions of social capital. Retired individuals tend to possess higher cognitive social capital but showcase lower structural social capital, potentially indicating reduced participation in group activities and political engagement among retired populations. Interestingly, full-time or part-time working individuals, along with students, demonstrate higher levels of both social capital dimensions. This might suggest that engagement in work and educational pursuits contributes to higher social interactions, consequently enhancing the accumulation of social capital. However, being a full-time worker does not significantly impact structural social capital. Full-time workers might have less availability for such activities as structural social capital involves participating in group activities and engaging in political action.

Both religiously affiliated and non-religious individuals show negative effects on both cognitive and structural social capital dimensions. Religiously affiliated individuals might have limited engagement in social activities outside their religious circles, affecting their overall social capital. Non-religious individuals may lack community involvement, which can diminish their social networks and, consequently, their social capital.

In terms of marital status, only individuals living together in a married arrangement exhibit higher structural social capital. Other marital statuses demonstrate a negative impact on both cognitive and structural social capital. This could be due to being single, divorced, or widowed, might lead to fewer opportunities for social interactions or shared experiences, resulting in lower social capital. Notably, being married does not significantly impact cognitive social capital, while being separated does not significantly affect structural social capital.

Goodness of fit statistics			
RMSEA	0.02		
GFI	0.99		
NFI	0.98		
		Cognitive Social Capital	Structural Social Capital
Age		0.02***	0.01***
Gender (Female)	Male	0.04***	0.2***
Education (Higher)	Lower	-0.7***	-1.20***
	Middle	-0.5***	-0.67***
Employment (Other)	Full Time	0.24***	0.05
	Part Time	0.29***	0.14***
	Self employed	-0.06	-0.12**
	Retired	0.27***	-0.25***
	Housewife	-0.32***	-0.7***
	Student	0.51***	0.23***
	Unemployed	-0.21***	-0.31***
Religion (atheist)	Religious person	-0.16***	-0.81***
	Not religious person	-0.07***	-0.43***
Marital Status (single)	Married	-0.03	-0.40***
	Living as married	-0.43***	0.41***
	Divorced	-0.29***	-0.16***
	Separated	-0.64***	0.02
	Widowed	-0.49***	-0.98***

Table 8: Individual covariate effect results

*Note: Reference categories in parentheses; significance: *=0.1, **=0.05, ***=0.01*

4.4 MIMIC Model with country covariates

Figure 3 presents visual representation for an extended MIMIC model, incorporating information about the countries where participants reside, building upon initial SEM model. Control of Corruption, GINI index, poverty gap, unemployment and GDP per capita are added into analysis.

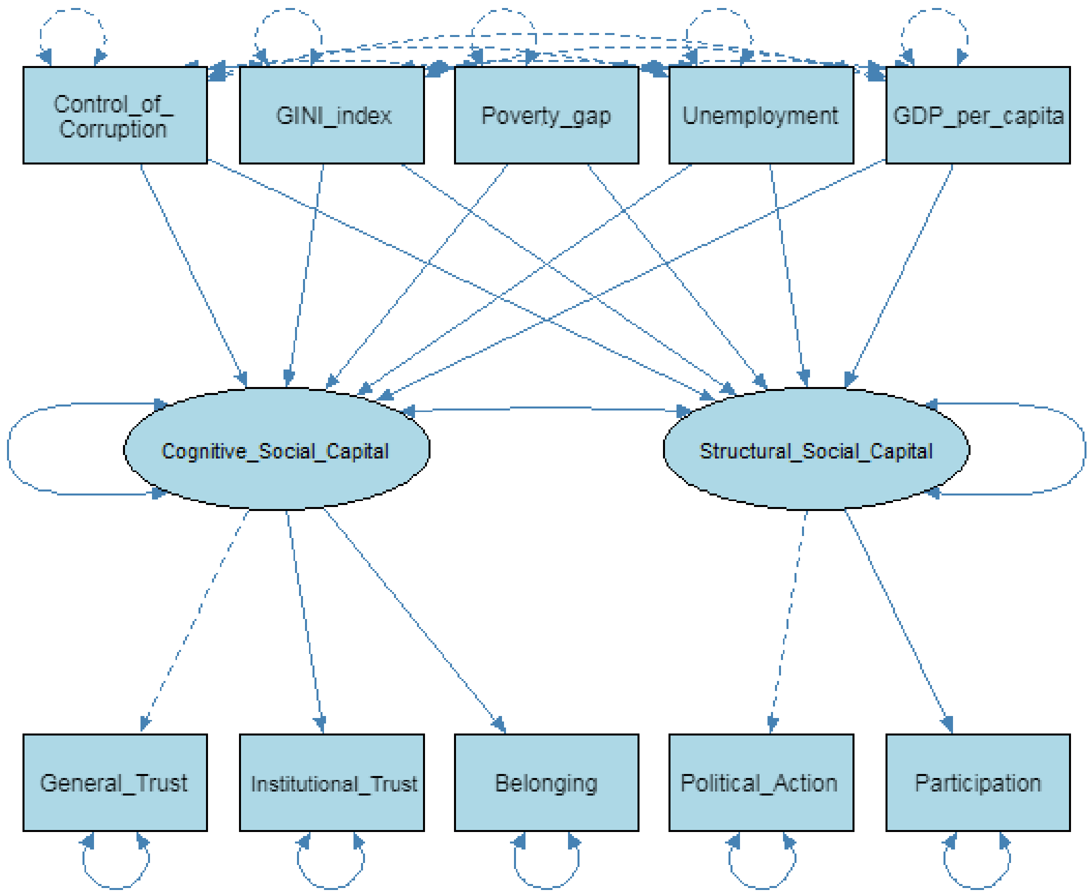

Figure 3: MIMIC model with country covariates

Table 9 presents outcomes obtained from MIMIC model with country-related covariates. Analyzing the coefficients, Control of Corruption exhibits a significantly positive association with both dimensions of social capital. Conversely, the GINI Index showcases a significant negative association with Cognitive Social Capital, while displaying a slightly positive but negligible association with Structural Social Capital. The Poverty Gap indicates a substantial positive effect on both dimensions of social capital, but relatively smaller in comparison. Unemployment demonstrates a positive impact on both forms of social capital, more pronounced for Cognitive Social Capital. However, GDP per capita does not exhibit a substantial impact on either Cognitive or Structural Social Capital. These findings underline the nuanced relationships between socio-economic indicators and different dimensions of social capital, emphasizing the varying impacts these factors exert on the formation and structure of social capital within a given context.

Goodness of fit statistics		
RMSEA	0.05	
GFI	0.97	
NFI	0.97	
	Cognitive Social Capital	Structural Social Capital
Control of Corruption	0.20***	0.24***
GINI Index	-0.05***	0.05***
Poverty Gap	0.02***	0.01***
Unemployment	0.03***	0.01***
GDP per capita	0.00***	0.00***

Table 9: Country covariate effects results

*Note: significance: *=0.1, **=0.05, ***=0.01*

4.5 Multilevel MIMIC model

Understanding the dynamics of social capital requires a nuanced examination of its influences across individual and country-level contexts. The Multilevel MIMIC model presents a robust approach to dissecting these intricate associations, allowing for a comprehensive analysis of how various socio-demographic factors operate within distinct individual contexts and the broader implications across diverse country-level settings.

The table 10 displays the results from multilevel MIMIC model. The model shows good overall fit with the observed data, indicated by RMSEA at 0.02, GFI and NFI both at 0.99. Within-cluster fit, as indicated by SRMR at 0.011, is strong, suggesting minimal discrepancy. However, between-cluster fit, represented by SRMR at 0.331, shows higher divergence, indicating more variability in covariances between different clusters or country levels.

At the within-country level, several socio-demographic factors exhibit significant associations. Education emerges as a prominent predictor, showcasing substantial positive effects for Middle and Higher education levels on both dimensions of social capital. On the other hand, being Male exhibits a significant negative impact on both Cognitive and Structural Social Capital. Employment statuses such as Retired, Housewife, Student, and Unemployed demonstrate detrimental effects, particularly on Cognitive Social Capital. Moreover, the religious affiliation of being not religious displays contrasting effects on both dimensions, negatively impacting Cognitive but positively affecting Structural Social Capital. Marital statuses, including Living as married and Divorced, exhibit negative impacts on Cognitive and Structural Social Capital, while being Married or Single shows differing effects on each dimension.

At the between-country level, the model reveals associations between socio-economic indicators like Corruption Control, GINI Index, Poverty Gap, Unemployment, and GDP per capita with both Cognitive and Structural Social Capital. However, in contrast to the simple MIMIC model, none of these factors appear significant for Cognitive Social Capital. In the context of Structural Social Capital, a higher GINI index and an increased unemployment rate demonstrate a significant positive association, suggesting that these factors contribute to the enhancement of structural social capital.

The Multilevel MIMIC model sheds light on the multifaceted nature of social capital dynamics. This comprehensive approach underscores the complex interplay of factors at different levels and highlights their differential contributions to the formation and enhancement of social capital.

Goodness of fit statistics			
RMSEA	0.02		
GFI	0.99		
NFI	0.99		
SRMR (within)	0.011		
SRMR (between)	0.331		
		Cognitive Social Capital	Structural Social Capital
Age		0.005***	0.001
Gender (Female)	Male	-0.091***	-0.208***
Education (Higher)	Lower	0.123***	0.470***
	Middle	0.408***	1.140***
Employment (Other)	Full Time	-0.007	0.067***
	Part Time	-0.021	0.105***
	Self employed	0.011	-0.306***
	Retired	-0.117***	-0.25***
	Housewife	-0.266***	-0.433***
	Student	-0.176***	-0.312***
	Unemployed	-0.248***	-0.294***
Religion (atheist)	Religious person	-0.374***	-0.003
	Not religious person	-0.484***	0.313***
Marital Status (single)	Married	-0.135***	0.131***
	Living as married	-0.254***	0.091***
	Divorced	-0.269***	-0.025
	Separated	-0.142***	-0.258***
	Widowed	-0.105***	0.085***
Control of Corruption		0.428	0.404
GINI Index		-0.103	0.101***
Poverty Gap		0.029	-0.03
Unemployment		0.061	0.108***
GDP per capita		0.000	0.000***

Table 10: Individual and country covariate effect results

*Note: Reference categories in parentheses; significance: *=0.1, **=0.05, ***=0.01*

5 Conclusions

Social capital has been recognized as a crucial factor in shaping societies and influencing various outcomes, ranging from economic development to individual well-being. Studies highlighted the multifaceted nature of social capital and its determinants at both individual and macro levels. The measurement model of social capital was based on an assumption that it consists of Cognitive (General Trust, Institutional Trust, and Belonging) and Structural (Participation and Political Action) components. Such model provides a clear distinction between the emotional and behavioral aspects of social relationships.

After conducting the multilevel SEM analysis, it was observed that the impact of Belonging appears to have minimal significance at the between-countries level compared to other determinants of social capital. Conversely, other factors influencing social capital demonstrated consistent statistically significant positive influences at both within and between country levels.

Individual and place attributes play different roles in explaining social capital dimensions. Higher age increase both cognitive and structural social capital dimensions. Male were observed to have higher social capital than women. Students, as well as working individuals, demonstrate higher levels of both social capital dimensions. This implies that active participation in work and educational activities increases social interactions, thereby social capital. Exception is full-time workers, since they lack of significant impact on structural social capital. This might be explained as having less free time for participation in group activities and taking political action. Retired individuals show higher cognitive but lower structural social capital, hinting reduced participation in collective activities and political engagement. Regarding the impact of marital status, only people living in a married arrangement exhibit higher structural social capital, while being single, divorced, or widowed, result in lower social capital, likely due to limited social interactions and shared experiences.

After analyzing country attributes, Control of Corruption emerges as the most influential factor, displaying a significantly positive association with both dimensions of social capital compared to other indicators. When corruption is well managed within a society, people tend to have greater trust in institutions, authorities, and fellow citizens. This enhance cooperation, civic engagement, and a sense of community. Conversely, GINI Index, which represents inequality, showcases a significant negative association with Cognitive Social Capital, while displaying a slightly positive but negligible association with Structural Social Capital. This could be understood as inequality make people have less trust and feelings of belonging, but it encourages to join voluntary groups and take political action to change it. Other indicators, such as GDP per capita, Poverty Gap and Unemployment exhibit very small positive impact on either Cognitive or Structural Social Capital.

At the cross-country level, only a higher inequality and increased unemployment demonstrate a significant positive association with Structural Social Capital, indicating their potential contribution to taking political action and joining group activities.

Research's findings complement existing studies by revealing nuanced influences on social capital. Notably, the impact of age, gender, and employment status on social capital dimensions resonates with the observations in previous researches. Additionally, exploration of country-level attributes such as corruption control and inequality aligns with studies highlighting their associations with social capital components. However, the distinct separation of structural and cognitive social capital components in this study offers a deeper understanding of the factors influencing individuals' sentiments versus their actions related to social capital. The comprehensive analysis of global data sets this study apart, as many previous analyses tend to focus on regional research.

Despite the substantial findings and their alignment with theoretical expectations, it is essential to acknowledge the study's limitations. The multivariate normality assumption was not met within the dataset. Despite the departure from normality, other SEM assumptions were satisfied, and the goodness of fit indices (GFI, NFI, RMSEA) supported the suitability of the chosen SEM approach, affirming the reliability of the findings. In future researches, the consideration of alternative methodologies like Partial Least Squares Structural Equation Modeling (PLS-SEM), which does not require normality assumptions, could be considered to potentially enhance and complement the current findings.

References

[1] Robert David Putnam. "Bowling alone: America's declining social capital". In: *Journal of democracy* 6.1 (1995), pp. 65–78. URL: https://doi.org/10.1353/jod.1995.0002.

[2] Norman Uphoff. *Understanding social capital: learning from the analysis and experience of participation. Social capital: A multifaceted perspective.* 2000, p. 216.

[3] Jenny Onyx and Paul Bullen. "Measuring Social Capital in Five Communities". In: *The Journal Of Applied Behavioral Science* 36.1 (2000), pp. 23–42. URL: https://doi.org/10.1177/0021886300361002.

[4] Wim van Oorschot, Wil Arts, and John Gelissen. "Social Capital in Europe: Measurement and Social and Regional Distribution of a Multifaceted Phenomenon". In: *Acta Sociologica* 49.2 (2006), pp. 149–167. URL: https://doi.org/10.1177/0001699306064770.

[5] Nan Lu and Jingyue Zhang. "Social Capital and Self-Rated Health among Older Adults Living in Urban China: A Mediation Model". In: *Sustainability* 11.20 (2019). ISSN: 2071-1050. DOI: 10.3390/su11205566. URL: https://www.mdpi.com/2071-1050/11/20/5566.

[6] Jan Fidrmuc and Klarita Gërxhani. "Formation of social capital in Central and Eastern Europe: Understanding the gap vis-à-vis developed countries". In: *CEPR Discussion Paper* 5068 (2005). URL: https://ssrn.com/abstract=776545.

[7] Anneli Kaasa and Eve Parts. "Individual-Level Determinants of Social Capital in Europe: Differences between Country Groups". In: *Acta Sociologica* 51.2 (2008), pp. 145–168. URL: http://www.jstor.org/stable/20460044.

[8] Asimina Christoforou. "Social Capital Across European Countries: Individual and Aggregate Determinants of Group Membership". In: *American Journal of Economics and Sociology* 70.3 (2011), pp. 699–728. URL: http://www.jstor.org/stable/41329206.

[9] Emanuele Ferragina. "The socio-economic determinants of social capital and the mediating effect of history: Making Democracy Work revisited". In: *International Journal of Comparative Sociology* 54.1 (2013), pp. 48–73. URL: https://doi.org/10.1177/0020715213481788.

[10] Pirmin Bundi and Markus Freitag. "Economic hardship and social capital in Europe: A comparative analysis of 27 democracies". In: *European Journal of Political Research* 59.2 (2020), pp. 290–311. URL: https://doi.org/10.1111/1475-6765.12359.

[11] Christian Haerpfer et al. *World Values Survey: Round Seven - Country-Pooled Datafile Version 5.0. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat.* 2022. URL: https://doi.org/10.14281/18241.20.

[12] *World Development Indicators*. World Bank Open Data. The World Bank, accessed 2023-12-06. URL: https://databank.worldbank.org/source/world-development-indicators#.

[13] Trudy Harpham. "The measurement of community social capital through surveys". In: *Social capital and health.* Ed. by I. Kawachi, S. Subramanian, and D Kim. New York: Springer, 2008, pp. 51–62. DOI: https://doi.org/10.1007/978-0-387-71311-3_3.

[14] Catherine M. Stein et al. "Structural Equation Modeling". In: *Statistical Human Genetics: Methods and Protocols*. Ed. by Robert C. Elston. New York, NY: Springer New York, 2017, pp. 557–580. ISBN: 978-1-4939-7274-6. DOI: 10.1007/978-1-4939-7274-6_28. URL: https://doi.org/10.1007/978-1-4939-7274-6_28.

[15] Joseph F. Hair et al. "An Introduction to Structural Equation Modeling". In: *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*. Cham: Springer International Publishing, 2021, pp. 1–29. ISBN: 978-3-030-80519-7. DOI: 10.1007/978-3-030-80519-7_1. URL: https://doi.org/10.1007/978-3-030-80519-7_1.

[16] Kenneth A Bollen. *Structural Equations with Latent Variables*. New York: Wiley, 1989, pp. 319–323.

[17] John M. Ferron and Melinda R. Hess. "Estimation in SEM: A Concrete Example". In: *Journal of Educational and Behavioral Statistics* 32.1 (2007), pp. 110–120. ISSN: 10769986, 19351054. URL: http://www.jstor.org/stable/20172071 (visited on 01/06/2024).

[18] Sunil Kumar. "Structure Equation Modeling Basic Assumptions and Concepts: A Novices Guide". In: *Asian Journal of Management Sciences* 03.07 (2015), pp. 25–28.

[19] Lawrence Boboye Ajayi and Adeyinka Taoheed Adebayo. "Structural Equation Model (SEM)". In: *American Journal of Humanities and Social Sciences Research (AJHSSR)* 5.7 (2023). Research Paper Open Access, pp. 11–19. ISSN: 2378-703X.

[20] Sophia Rabe-Hesketh, Anders Skrondal, and Xiaohui Zheng. "10 - Multilevel Structural Equation Modeling". In: *Handbook of Latent Variable and Related Models*. Ed. by Sik-Yum Lee. Handbook of Computing and Statistics with Applications. Amsterdam: North-Holland, 2007, pp. 209–227. DOI: https://doi.org/10.1016/B978-044452044-9/50013-6. URL: https://www.sciencedirect.com/science/article/pii/B9780444520449500136.

[21] Fernando Mata and Ravi Pendakur. "Social Capital Formation and Diversity in Canadian Cities: Impacts of Individual and Place-Related Characteristics". In: *Journal of Ethnic and Migration Studies* 38 (Dec. 2012). DOI: 10.1080/1369183X.2012.711030.

[22] Rob Kessels and Mirjam Moerbeek. "A comparison of the multilevel MIMIC model to the multilevel regression and mixed ANOVA model for the estimation and testing of a cross-level interaction effect: A simulation study". In: *Biometrical Journal* 65.5 (2023), p. 2200112. DOI: https://doi.org/10.1002/bimj.202200112.

[23] Anna A. Igolkina and Georgy Meshcheryakov. "semopy: A Python Package for Structural Equation Modeling". In: *Structural Equation Modeling: A Multidisciplinary Journal* 27.6 (2020), pp. 952–963. DOI: 10.1080/10705511.2019.1704289.

[24] Yves Rosseel. "lavaan: An R Package for Structural Equation Modeling". In: *Journal of Statistical Software* 48.2 (2012), pp. 1–36. DOI: 10.18637/jss.v048.i02. URL: https://www.jstatsoft.org/index.php/jss/article/view/v048i02.

[25] Daire Hooper, Joseph Coughlan, and Michael R. Mullen. "Structural Equation Modelling: Guidelines for Determining Model Fit". In: *The Electronic Journal of Business Research Methods* 6.1 (2008), pp. 53–60. URL: www.ejbrm.com.

6 Appendix A

Python programme code:

```
import pandas as pd
import numpy as np
import boto
import semopy
import pingouin as pg
import seaborn as sns
from scipy.stats import spearmanr
import matplotlib.pyplot as plt
import dython
from dython.nominal import associations
from dython.nominal import identify_nominal_columns
get_ipython().run_line_magic('matplotlib', 'inline')
import matplotlib
matplotlib.rcParams['figure.figsize'] = (12,6)
from semopy.multigroup import multigroup
pd.options.mode.chained_assignment = None
import graphviz
semopy.graphviz_dot = 'C:/Program Files/Graphviz/bin/dot.exe'
from sklearn.preprocessing import LabelEncoder
from pingouin import multivariate_normality
from statsmodels.stats.outliers_influence import variance_inflation_factor
from statsmodels.tools.tools import add_constant

dff = pd.read_csv("C:/Users/Hp/Desktop/wdi.csv", delimiter=";")
df = pd.read_csv("C:/Users/Hp/Documents/ZA7505_v4-0-0.csv", delimiter=",")

columns_to_select = [
    "uniqid","year","cntry", "cntry_AN",
    "X001", "X002", "X003", "X007", "X025R", "V004RM", "V004RF", "X028", "E012",
    "X047E_EVS5", "F034", "A170", "E012", "F025", "F050", "A001", "A002",
    "A027", "A029", "A030", "A032", "A034", "A035", "A038", "A039", "A040",
    "A041", "A042", "A065", "A066", "A067", "A068", "A071", "A072", "A074",
    "A078", "A079", "A080_01", "A080_02", "A165", "D001_B", "G007_18_B",
    "G007_33_B", "G007_34_B", "G007_35_B", "G007_36_B", "E025", "E026",
    "E027", "E028", "E263", "E264", "E069_01", "E069_04", "E069_05", "E069_07",
    "E069_08", "E069_12", "E069_13", "E069_14", "E069_17", "E069_20", "F114A",
    "F115", "F116", "F117", "F118", "F119", "F120", "F121", "F122", "F123",
    "F132", "E290", "F144_02", "G006", "G062", "G063", "G255", "G256", "G257"
]
data = df[columns_to_select]

#testing internal consistency between columns
column_pa = data[['E025','E026','E027','E028']]
```

```
alpha = pg.cronbach_alpha(column_pa)
print("Cronbach's Alpha for Political Action:", alpha)

column_p = data[["A065", "A066", "A067", "A068", "A071", "A072", "A074", "A078",
                "A079", "A080_01", "A080_02"]]
alpha = pg.cronbach_alpha(column_p)
print("Cronbach's Alpha for Participation:", alpha)

column_it = data[["E069_01", "E069_04", "E069_05", "E069_07", "E069_08", "E069_12",
                "E069_13", "E069_14", "E069_17", "E069_20"]]
alpha = pg.cronbach_alpha(column_it)
print("Cronbach's Alpha for Institutional Trust:", alpha)

column_t = data[["D001_B", "G007_18_B", "G007_33_B", "G007_34_B", "G007_35_B",
                "G007_36_B"]]
alpha = pg.cronbach_alpha(column_t)
print("Cronbach's Alpha for General Trust:", alpha)

column_m = data[["F114A", "F115", "F116", "F117", "E290"]]
alpha = pg.cronbach_alpha(column_m)
print("Cronbach's Alpha for Morality:", alpha)

column_b = data[["G062", "G063", "G255", "G256", "G257"]]
alpha = pg.cronbach_alpha(column_b)
print("Cronbach's Alpha for Belonging:", alpha)

#item parceling
data["Political_Action"] = data[['E025','E026','E027','E028']].sum(axis=1, skipna=True)
data["Participation"] = data[["A065", "A066", "A067", "A068", "A071", "A072", "A074", "A078",
                            "A079", "A080_01", "A080_02"]].sum(axis=1, skipna=True)
data["Institutional_Trust"] = data[["E069_01","E069_04","E069_05","E069_07","E069_08","E069_12",
                            "E069_13", "E069_14", "E069_17", "E069_20"]].sum(axis=1, skipna=True)
data["General_Trust"] = data[["D001_B", "G007_18_B", "G007_33_B", "G007_34_B", "G007_35_B",
                        "G007_36_B"]].sum(axis=1, skipna=True)
data["Morality"] = data[["F114A", "F115", "F116", "F117", "E290"]].sum(axis=1, skipna=True)
data["Belonging"] = data[["G062", "G063", "G255", "G256", "G257"]].sum(axis=1, skipna=True)

data.rename(columns={'X001': 'Gender', 'X002': 'Birth_year', 'X003': 'Age',
            'X007': 'Marital_status', 'X025R': 'Education', 'X028': 'Employment',
            'X047E_EVS5': 'Income', 'F034':'Religion'}, inplace=True)
selected_columns = ['uniqid', 'year', 'cntry_AN', 'country', 'continent', 'Gender', 'Age',
                    'Marital_status', 'Education','Employment', 'Religion', 'General_Trust',
                    'Political_Action','Participation', 'Institutional_Trust', 'Morality',
                    'Belonging']
data = data[selected_columns]
data.isnull().sum()
data.isna().sum()
```

```
cleaned_data = data.dropna()
cleaned_data.isna().sum()

#merging WVS and WDI datasets
merged_data = pd.merge(cleaned_data, dff, on='country', how='inner')
data = merged_data

#sample characteristics
column_names = ['Gender', 'Education', 'Marital_status', 'Employment', 'Religion']
total_observations = len(data)
for column in column_names:
    distinct_values_count = data[column].value_counts()
    percentage_values = (distinct_values_count / total_observations) * 100
    percentage_values_rounded = percentage_values.round(2)
    result_df = pd.DataFrame({'Count': distinct_values_count,
                              'Percentage': percentage_values_rounded})
    print(f"Column: {column}")
    print(result_df)
    print()

bin_edges = [0, 20, 30, 40, 50, 60, 70, 80, 100]
bin_labels = ['0-20', '21-30', '31-40', '41-50', '51-60', '61-70', '71-80', '80+']
data['Age_Category'] = pd.cut(data['Age'], bins=bin_edges, labels=bin_labels, right=False)
age_category_counts = data['Age_Category'].value_counts()
percentage_values = (age_category_counts / len(data['Age_Category'])) * 100
result_df = pd.DataFrame({'Count': age_category_counts, 'Percentage': percentage_values.round(2)})
print("Age Categories:")
print(result_df)

column = 'continent'
distinct_values_count = data[column].value_counts()
total_observations = len(data)
percentage_values = (distinct_values_count / total_observations) * 100
percentage_values_rounded = percentage_values.round(2)
countries_per_continent = data.groupby(column)['country'].nunique()

result_df = pd.DataFrame({'Count': distinct_values_count, 'Percentage': percentage_values_rounded,
            'Countries': countries_per_continent})
print(f"Column: {column}")
print(result_df)

print("Summary Statistics:")
print(data.describe())

#outliers are removed using mode values this way discrete values remain
def return_outliers(data):
```

```
    outliers = []
    Q1 = data.quantile(0.25)
    Q3 = data.quantile(0.75)
    IQR = Q3-Q1
    Lower_bound = Q1-1.5*IQR
    Upper_bound = Q3+1.5*IQR
    for i in data:
        if i>Upper_bound or i<Lower_bound:
            outliers.append(i)
    return outliers

outlier_values = return_outliers(data['General_Trust'])
data['General_Trust']=data['General_Trust'].apply(lambda x: data['General_Trust'].mode().iloc[0]
                        if x in outlier_values else x)
outlier_values_i = return_outliers(data['Institutional_Trust'])
data['Institutional_Trust']=data['Institutional_Trust'].apply(lambda x: data['Institutional_Trust']
                            .mode().iloc[0] if x in outlier_values_i else x)
outlier_values_b = return_outliers(data['Belonging'])
data['Belonging']= data['Belonging'].apply(lambda x: data['Belonging'].mode().iloc[0]
                    if x in outlier_values_b else x)
outlier_values_p = return_outliers(data['Participation'])
data['Participation']=data['Participation'].apply(lambda x: data['Participation'].mode().iloc[0]
                        if x in outlier_values_p else x)
outlier_values_pa = return_outliers(data['Political_Action'])
data['Political_Action']=data['Political_Action'].apply(lambda x: data['Morality'].mode().iloc[0]
                        if x in outlier_values_pa else x)
outlier_values_m = return_outliers(data['Morality'])
data['Morality']=data['Morality'].apply(lambda x: data['Morality'].mode().iloc[0]
                    if x in outlier_values_m else x)

#testing SEM assumptions
selected_columns = ['General_Trust','Political_Action','Participation', 'Institutional_Trust',
                    'Belonging']
data1 = cleaned_data[selected_columns]
sampled_data = data1.sample(frac=0.1, random_state=42)
# Multivariate Normality Test
multivariate_normality(sampled_data, alpha=.05)

#Multicollinearity
data1 = add_constant(data1)
# Calculate VIF for each variable
vif = pd.DataFrame()
vif["VIF Factor"] = [variance_inflation_factor(data1.values, i) for i in range(data1.shape[1])]
vif["Variable"] = data1.columns
print(vif)

#correlation matrix
```

```
columns_of_interest = ['Institutional_Trust', 'General_Trust', 'Belonging', 'Political_Action',
                    'Participation']
corr_matrix, p_values = spearmanr(data[columns_of_interest], nan_policy='omit')
rounded_corr_matrix = pd.DataFrame(corr_matrix, columns=columns_of_interest,
                    index=columns_of_interest).round(3)
symbols = (pd.DataFrame(p_values, columns=columns_of_interest, index=columns_of_interest)
           .applymap(lambda x: '' if x>0.05 else '*' if x>0.01 else '**' if x>0.001 else '***')
           .astype(str))
correlation_matrix_with_symbols = rounded_corr_matrix.astype(str) + symbols
print("Correlation Matrix:")
print(correlation_matrix_with_symbols)

#Measurement model
model_s = """
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Structural_Social_Capital ~~ Cognitive_Social_Capital
        """
models = semopy.Model(model_s)
res_opt = models.fit(data
estimatess = models.inspect()
res_opt
models.inspect(std_est=True)
statss = semopy.calc_stats(models)
pd.options.display.float_format = '{:.6f}'.format
print(statss.T)
estimatess
g = semopy.semplot(models, "model.png", plot_exos=True, plot_covs=True)

#MIMIC model with individual covariates
categorical_columns = ['Gender', 'Education', 'Employment', 'Religion', 'Marital_status']
data_with_dummies = pd.get_dummies(data, columns=categorical_columns)
model_spec = """
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~ Age + Gender_1 + Education_1 + Education_2 + Religion_1 +
    Religion_2 + Employment_1 + Employment_2 +Employment_3 + Employment_4 + Employment_5 +
    Employment_6 + Employment_7 + Marital_status_1 + Marital_status_2 + Marital_status_3 +
    Marital_status_4 + Marital_status_5
    Structural_Social_Capital ~ Age + Gender_1 + Education_1 + Education_2 +Religion_1 +
    Religion_2 + Employment_1 + Employment_2 + Employment_3 + Employment_4 + Employment_5 +
    Employment_6 + Employment_7 + Marital_status_1 + Marital_status_2 + Marital_status_3 +
    Marital_status_4 + Marital_status_5
    Structural_Social_Capital ~~ Cognitive_Social_Capital
        """
model = semopy.Model(model_spec)
res_opt = model.fit(data_with_dummies, obj="WLS")
```

```
estimates = model.inspect()
stats = semopy.calc_stats(model)
pd.options.display.float_format = '{:.6f}'.format
print(stats.T)
g = semopy.semplot(model, "model2.png")

# MIMIC model with country covariates
cleaned_data_country = data.dropna()
categorical_columns = ['Gender', 'Education', 'Employment', 'Religion', 'Marital_status']

model_sp = """
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~ Control_of_Corruption + GINI_index + Poverty_gap +
                            Unemployment + GDP_per_capita
    Structural_Social_Capital ~ Control_of_Corruption + GINI_index + Poverty_gap +
                            Unemployment + GDP_per_capita
    Structural_Social_Capital ~~ Cognitive_Social_Capital
        """
model = semopy.Model(model_sp)
res_opt = model.fit(cleaned_data_country)
estimates = model.inspect()
estimates

stats = semopy.calc_stats(model)
pd.options.display.float_format = '{:.6f}'.format
print(stats.T)
```

R programme code:

```
library(lavaan)
data  <- read.csv("C:/Users/Hp/Desktop/data.csv", sep=",", header=TRUE)

#Multilevel SEM model
model_multi <- '
    level: 1
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~~ Structural_Social_Capital
    level: 2
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~~ Structural_Social_Capital
'
fit <- sem(model = model_multi, data = data, std.lv = TRUE)
summary(fit, fit.measures=TRUE)
```

```
#Multilevel MIMIC model
model_multi2 <- '
    level: 1
    Cognitive_Social_Capital =~ General_Trust + Institutional_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~~ Structural_Social_Capital
    Cognitive_Social_Capital ~ Age + Gender_1 + Education_1 + Education_2 + Religion_1 +
                                Religion_2 + Employment_1 + Employment_2 +Employment_3 +
                                Employment_4 +Employment_5 +Employment_6 + Employment_7 +
                                Marital_status_1 + Marital_status_2 + Marital_status_3 +
                                Marital_status_4 + Marital_status_5
    Structural_Social_Capital ~ Age + Gender_1 + Education_1 + Education_2 + Religion_1 +
                                Religion_2 + Employment_1 + Employment_2 +Employment_3 +
                                Employment_4 +Employment_5 +Employment_6 + Employment_7 +
                                Marital_status_1 + Marital_status_2 + Marital_status_3 +
                                Marital_status_4 + Marital_status_5
    level: 2
    Cognitive_Social_Capital =~ Institutional_Trust + General_Trust + Belonging
    Structural_Social_Capital =~ Political_Action + Participation
    Cognitive_Social_Capital ~~ Structural_Social_Capital
    Cognitive_Social_Capital ~ Control_of_Corruption + GINI_index + Poverty_gap +
                                Unemployment + GDP_per_capita
    Structural_Social_Capital ~ Control_of_Corruption + GINI_index + Poverty_gap +
                                Unemployment + GDP_per_capita
'
fit2 <- sem(model = model_multi2, data = data, std.lv = TRUE, cluster='country')
summary(fit2, fit.measures=TRUE)
```