Assessing measurement invariance across countries and regions in Europe: the case of political trust

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2021

Sixten M. Thestrup

School of Social Sciences

Department of Social Statistics

Contents

Number of words: 37,708

List of Figures

List of Tables

Abstract

This Thesis identifies issues with and proposes solutions for the measurement of political trust across space and time. We argue that political trust is a multidimensional concept which is best measured at the regional level. To illustrate this, we conduct three interconnected empirical studies. In the first study, we propose and test a new measure of political trust using multiple groups confirmatory factor analysis across 18 countries on the European Social Survey from 2016. We find that 'national political trust' and 'supranational political trust' are two empirically distinct measures, which are highly correlated in most countries. In the second study, we suggest that regional variation in political trust might be of higher importance than variation across countries. This is followed by considerations of changes to the standard modelling framework, which are necessary to successfully estimate a measurement invariance model at the regional level. In the third study, we provide a novel regional-level measurement model, which is estimated and tested using a Bayesian approach. The study pays particular attention to the choice of prior distributions and model fit and diagnostics. Our results show that there are both important and meaningful differences in political trust at the regional level. We propose that our results demonstrate the need for a fully Bayesian spatial multi-group confirmatory factor analysis model to uncover potential spatial effects and their relation to political trust at the regional level.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgements

I would like to show my sincere appreciation to my main supervisor and mentor, Dr Johan Koskinen, for invaluable support and guidance during four years of work, especially when things were tough and did not go my way. Without his help, this project would not have been possible. Also, I wish to show my gratitude to my supervisors Dr Bram Vahoutte and Prof Wendy Olsen for their much-appreciated feedback.

I wish to express my deepest gratitude to my wife, Kate, for her encouragement - the journey that we are on together is truly life-affirming. Also a big thanks to my parents, Merle and Stig, who always support me.

This work was supported by the Economic and Social Research Council North West Doctoral Training Centre and the School of Social Sciences, University of Manchester.

A big thanks to my friends, especially Jakob and Pablo, and my former PhD colleagues Georgia, Angelo, Joe, Pete, José, Javier and Adrian who are always up for a good discussion.

Finally, thanks to Morphy for his moral support every day.

Chapter 1

Introduction

Current research on and relating to political trust is of high relevance and has become increasingly popular in recent years (Cordero and Simón 2015; André 2013; Miller and Listhaug 1999). Many studies have introduced more advanced statistical tools to measure theoretical concepts across both time and space (Schneider 2016; André 2013; Märien 2011).

The theme of this thesis is to highlight issues and propose solutions regarding the measurement of attitudes - more specifically political trust. One of the great complications within the political and social sciences is that complex theoretical concepts are difficult to translate into an empirical modelling framework (Saris and Gallhofer 2014; Marozzi 2014). Not only in the initial transformation of the concept into for example a survey, but also the posthoc analysis. Often, researchers are concerned with the comparison of a measure across either time or space where the comparison may be cross-country or cross-regional which introduces additional complexity. It is common to find standard statistical approaches to measuring and comparing groups in the literature (e.g. linear regression, multilevel modelling, analysis of variance) (Miller and Listhaug 1999; Anderson et al. 2005; Torcal and

Montero 1999; Newton 2001). One issue with these methods is the question: do we measure the same construct across groups?

In the last few years, testing for measurement invariance has gained an increase in popularity, and we experience a larger number of researchers who are concerned about how invariant their measure of interest is across groups (Ariely and Davidov 2010; Allum, Read, and Sturgis 2018). This is found in studies using large-scale social surveys as well as small scale effect-measures and local studies. In short, measurement invariance is a property meaning that the same construct is measured across groups with different characteristics. If measurement invariance does not hold, it can have an effect on our substantive conclusions.

Regarding the research on social and political trust, establishing measurement invariance is paramount since the researcher is almost always faced with data involving different groups based on ethnicity, language, culture or nationality, just to name a few. In addition, changes in political trust happen frequently as it is highly dependent on fluctuations in the political landscape, which makes it one of the utmost interesting theoretical fields to investigate. We will discuss measurement invariance in more detail in Section 2.6.

Currently, research on political trust and similar topics face two major difficulties within the area of measurement invariance. The primary debates surround the case of unexpected non-invariance; is the study irrelevant? Should we abandon ship and try with some alternative data or theory? While some studies argue that the researcher can have a certain degree of non-invariance (partial invariance) and still draw sensible conclusions (Byrne, Shavelson, and Muthén 1989), others argue for an alternative application of the measurement invariance method to loosen the requirements (Shi et al. 2017). The discussion about how best to tackle unexpected measurement non-invariance continues to play an important role in the continued development and integration of statistical methods in the political and social sciences.

Second, we have not been able to find many studies that apply the measurement invariance methods on a very large number of groups (with a few notable exceptions, such as Zercher et al. (2015)). How might that be? When diving into the modelling part of the exercise, one immediately realises that current methods are simply not able to handle such a scenario. This may deter some from attempting to establish measurement invariance altogether, resulting in either the use of a simpler, yet sub-optimal, method or using aggregated measures instead.

1.1 Research questions

In this thesis, we will discuss and provide answers to the following research questions:

1. How can we measure political trust in a large cross-national survey, while ensuring that it is measurement invariant?
2. What are the current issues with the measurement invariance method when extended to many groups or groups with low sample size?
3. How can we develop a framework for testing measurement invariance when conventional methods do not work?

In the following, we will provide an overview of key concepts used throughout the thesis, and how their intersections can answer the research questions.

1.2 Political trust, measurement and Bayesian methods

In this section, we will first provide an overview of the key concept of political trust and how it relates to measurement modelling and invariance. Second, we will present the argument for why the hierarchy of geographical units plays a significant role when modelling political trust in particular. The purpose of this summary is to provide a clear overview of the process needed to answer the research questions.

1.2.1 What is political trust?

The concept of trust has been important in theoretical political and social science for a long time. A very basic definition of trust is that of ‘encapsulated interest’; an actors’ belief that others will not do him/her harm and, at best, act in his/her interest (Hardin 1998). A simple example of inter-personal trust would be as follows: A lends \$10 to B and expects that B will return the money in the future. In the best case, B returns the money, maybe even earlier than anticipated. In the worst case, B does not return the money, and A has lost \$10. If A find it likely that B will return the money, then we can say that A has trust in B. In this example, it is clear that trust involves some degree of risk, since B may act against the interest of A, no matter the level of trust which A has in B.

Social trust is another type of trust, which extends to civil society. It overlaps to some extent with other concepts, such as solidarity, empathy, toleration, respect etc. In this sense, social trust is related to the social interactions that actors engage in under co-presence (Giddens 1991). In other words, social trust is built upon the immediate experiences of others (Newton 2001).

Political trust is different from both inter-personal and social trust in the sense that it is built upon knowledge of others at a distance. Hence, political trust is not a result of direct social interactions between individuals, but rather of the evaluation of people/institutions/organisations formed by impressions and knowledge obtained through, for example, the media. Researchers have developed different definitions of political trust, and how it is best understood. Ken Newton is one of the more notable contributors within this field and defines political trust as "attitudes towards political institutions and leaders" (Newton 1999). Others, such as Hetherington (1998), defines political trust as the citizens' normative evaluation of government performance. A third definition is the citizens' belief in institutionalised practices and procedures as described by Sztompka (2003).

The difference between these examples of definitions is mostly tied to what is evaluated. In the Newton (1999) definition, political trust is the citizens' a priori belief that political institutions and leaders will act in their interest, whereas the Hetherington (1998) definition is an a posteriori evaluation of how a particular part of the political sphere, the government, performed against expectations. We will adopt the definition of Sztompka (2003). Citizens' belief in institutionalised practices and procedures lies in the intersection of the two other definitions. It is not an evaluation of performance or a belief in demarcated areas of the political sphere, but rather a belief in the framework on which the political world is built upon. The main difference is that a citizen can have a high level of political trust, even if governments, politicians or institutions are not performing according to expectations, simply because there is a belief that the practices and procedures eventually will result in outcomes that are in the interest of the citizen.

1.2.2 Why is political trust important?

One might ask why political trust is important in the first place. The key argument here is that we can think of high levels of political trust as a democratic resource, which contributes to a more stable democracy, which is also more economically efficient (Putnam 1995; Fukuyama 1995) and provides a more well-functioning civil society (Newton 2001). Putnam (1995) follows an argument where political and social participation over time will increase the level of trust, which in turn leads to increased levels of integration and stability. Fukuyama (1995) focuses more on the economic benefits of the expectations of cooperative behaviour, which he argues is directly related to levels of trust between citizen and their trust in institutions and the society as a whole. This is in line with Newton (2001) who argues that political trust is a 'litmus test', which detects if the political system is functioning as is should - something which will be reflected in the civil society as well. Furthermore, he argues (among others) that political trust is a necessity for the well-functioning of democracy as a whole (Newton 2001).

In this sense, researchers agree that a high level of political trust is a desirable asset. This of course makes it intriguing to investigate further in empirical studies; being able to identify what constitutes political trust, what empirical dimensions it contains and how it relates to outcomes of interest, could be of particular interest not only within academia but to decision-makers as well.

1.2.3 How do we measure political trust?

In existing empirical research within the social and political sciences, the majority of researchers rely on large-scale social surveys to attempt to understand how it varies across different groups. Even some natural experiments rely on the use of a large-scale social survey. One such example is in Ares and Hernández (2017), which uses the European Social Survey to measure the relationship between political trust in politicians and corruption, following the uncovering of a large corruption scandal in Spain (Ares and Hernández 2017). Two of the most used surveys are indeed the European Social Survey (European Social Survey 2017) along with the World Values Survey (Inglehart et al. 2014). Both of these surveys cover a large number of countries and contains question modules that are relevant to the measure of political trust. The European Social Survey contains the core module 'Politics', while the World values survey has the 'confidence' module (European Social Survey 2017; Inglehart et al. 2014). Examples of the usage of the World Values survey module is found in Delhey and Newton (2005), with the European Social Survey module being used by many, including Märien and Hooghe (2011). Other surveys also exist, such as the Eurobarometer series (Commission 2020) and the European Election Study (Schmitt et al. 2016), which is used by, for example, Klingemann and Weldon (2012).

However, one of the main drawbacks of using an existing survey is, that questions may not necessarily align with the definition of political trust adopted by the researcher. As a result, several studies turn out to be more explorative in nature, rather than confirmatory. One very popular approach to measuring political trust, and other theoretical concepts, is through a measurement model. Measurement models have their roots in psychometrics, namely through the development of classical test theory (see Traub (2005)), item response theory (Hambleton 1985)and the widely used Rasch model (Rasch 1993). One of the first applications of measurement models was on the measurement of intelligence, dating all the way back to Francis Galton (1822-1911). In its simplest form, a measurement model can be written as:

$$\boldsymbol{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

Here, $\boldsymbol{y}$ is a vector of observed variables, $\boldsymbol{\tau}$ is a vector of item intercepts, $\boldsymbol{\Lambda}$ is a matrix of factor loadings, $\boldsymbol{\eta}$ is a vector of latent variables and $\boldsymbol{\epsilon}$ is a vector of residual errors. In an explorative setting, the attention would be on finding the factor construct which provides the best fit for the data, given several survey items related to political trust. This is seen in, for example, Rothstein and Stolle (2008), Allum, Read, and Sturgis (2018), Schneider (2016), Marozzi (2014) and André (2013), which do not have a fixed number of dimensions of political trust in place before the empirical analysis, but rather extrapolate the dimensionality based on modelling outcomes. On the other hand, a confirmatory approach uses the theoretical definition of political trust as the basis of the model definition, thereby fixing the items and factor structure prior to the empirical analysis.

In this thesis, we approach the empirical analysis using a confirmatory approach using the European Social Survey, which is further described in Section 1.4. Our definition of political trust as citizens' belief in institutionalised practices and procedures aligns well with two of the three dimensions described by André (2013), which forms the basis of the empirical analysis: National Political Trust (trust in national parliaments, politicians and parties) and Supranational Political Trust (trust in the European Parliament and the United Nations). The empirical studies are described in more detail in Section 2.4.

1.2.4 What is measurement modelling and invariance?

One of the difficulties when modelling a theoretical concept like political trust in a large-scale social survey is the assumption that the questions asked in the survey carry the same meaning across groups (for example countries). As an example, consider 'National Political Trust' as described above, which entails three questions/variables. If we set up a simple factor/measurement model, how do we know if the latent mean for National Political Trust in country A can be interpreted similarly as the same latent mean in country B? One method, which has gained increased popularity over the years is measurement invariance testing through multiple groups confirmatory factor analysis (MGCFA) models (Steenkamp and Baumgartner 1998; Davidov et al. 2014). Compared to the simple model outlined above, the MGCFA approach extends to:

$$\boldsymbol{y}^g = \boldsymbol{\tau}^g + \boldsymbol{\Lambda}^g \boldsymbol{\eta}^g + \boldsymbol{\epsilon}^g$$

Here, the model is extended to $g = 1, ..., G$ groups, which can be any mutually exclusive entities such as countries, regions or even gender and age groups. To test whether the measurement model in group 1 and 2 can be interpreted similarly, we introduce measurement invariance testing, which is a rigorous way of detecting differences between parameters across groups. The most common procedure is to test three hierarchical levels of measurement invariance across the G groups: configural invariance (factor structure), metric invariance (factor loadings) and scalar invariance (factor intercepts). If it is possible to obtain scalar invariance, we can conclude that the factor score for an individual i is independent of group membership. As a result, we can make meaningful comparisons of factor means between groups. The classical measurement invariance method is discussed in detail in Section 2.6.

1.2.5 The relationship between political trust and measurement invariance

Within the literature on political trust, the MGCFA method described above is widely used to test for measurement invariance between countries. However, in several of them, it is not possible to obtain full scalar invariance across all countries in the respective studies (see for example Märien (2011)). This might be due to cultural, linguistic or political differences between countries, where the items are not interpreted in the same way by the respondents, or where the importance of items vary notably (Brown et al. 2015). While methods exist to still extract useful information from partially invariant models (Saris and Gallhofer 2014; Byrne, Shavelson, and Muthén 1989; Putnick and Bornstein 2016) it is by no means the desired outcome. But what does the lack of scalar invariance indicate when interpreting the model concerning political trust?

In short, it indicates that political trust is not measured in the same way in all groups. For example, considering Supranational Political Trust (trust in the European Parliament and the United Nations) mentioned above, we may discover that the intercept for trust in the European Parliament differs between the two countries. This means, that the relative importance of that particular item for the measure of Supranational Political Trust varies across countries. In other words, the estimated level of trust in the European Parliament for two individuals from two different countries with the same latent factor mean would not be the same. This naturally poses issues when trying to understand how political trust should be measured. However, it is also an important finding for the researcher, since the reason for specific groups not being measurement invariant may have a theoretical interpretation and justification.

1.2.6 Choice of a hierarchy of geographical units and political trust

When investigating political trust within a geographically confined area (such as Europe, which is the focus of this thesis), the researcher has to choose a level of geographical granularity. The most granular option is simply to consider the full sample as one unit and carry out aggregated measurements. However, especially for political trust, this does not appear as a reasonable strategy. Due to cultural, political, linguistic, ethnic and other differences across countries, we would not expect the level of political trust to be equivalent across Europe. Hence, most studies, which carry out measurement invariance models on large-scale social surveys, do so at the country level. One example is Arnold, Sapir, and Zapryanova (2012), who finds that trust in the institutions of the European Union is driven by trust in the domestic (i.e. country-level) institutions. European Union member states with a high level of country-level corruption also tend to trust institutions of the European Union less (Arnold, Sapir, and Zapryanova 2012). Another example is van der Meer (2010) who investigates the differences in trust in the national parliament across multiple European countries over time. He finds that corruption, type of electoral system and former regime type as the main predictors for country-level differences in trust in the national parliament, although the model does not capture the longitudinal changes in trust (van der Meer 2010).

The choice of further granulated groupings is very rare, but why? One reason may be that sub-country geographical information does not follow a consistent definition across countries in the chosen survey. This is the case in for example the World Values Survey, where the country-level regional divisions do not necessarily follow the same coding scheme for countries on different continents (Inglehart et al. 2014). Another reason might be, that the researcher believes that political trust as a measure is stable within countries and that further granulation does not improve the understanding of differences in political trust across groups. However, some studies show that regional differences can be present. One such example is in Charron and Rothstein (2014), who investigates a social trust measure and finds that '[...] the differences between regions in many countries are noteworthy and demonstrate clear limits of national level analyses.'

In the European Social Survey, each individual is classified into NUTS regions (further described in Section 1.4.1), making a sub-country (regional) division possible. As we will demonstrate in Chapter 4, it is possible to detect substantial regional differences in political trust in many countries. Investigating other, albeit related, construct, Iacono (2019), Schoene (2016), Rustenbach (2010) and Bäck et al. (2018) have also investigated variation across NUTS regions. Both theoretical arguments and empirical research points in the direction that political trust, opinion and behaviour is not necessarily independent of regional affiliation. On such example is differences in voting patterns during the Brexit referendum in 2016 - a case that we will use as a motivating example in Section 1.5.

Finally, a reason why measurement invariance models are not carried out at the regional level could be related to the methodological issues that the researcher encounters when setting up a model; regions with low sample sizes and a large number of parameters are potentially difficult to estimate. While this is not explicitly stated as a reason in any of the studies under investigation in this thesis, it is an issue that we encounter even for a simple measurement model. One way to handle the estimation issue is to apply Bayesian methods, more specifically Bayesian MGCFA models, which is the main topic of Chapter 4.

1.2.7 Bayesian inference

The basis of the modelling approach in Chapter 4 is Bayesian inference. In essence, Bayesian inference uses Bayes' theorem from probability theory:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)},$$

The Bayesian MGCFA models involves a prior distribution, the likelihood (the data generating model) and a posterior distribution. Bayes theorem is used to update the prior information about the parameters to a posterior. One of the difficulties of Bayesian inference is the choice of prior distribution. Also, simulating using Markov Chain Monte Carlo can be time-consuming and might require substantial tuning. Both of these issues will be explored in detail in Chapter 4. The main advantage is that posterior is well-defined, given that the prior is proper, which means that model estimation is always possible. Secondly, it offers great flexibility on high-dimensional data - something which is particularly useful in MGCFA models with many groups/parameters. In Chapter 4, it will be clear that Bayesian MGCFA is a necessity if we want to estimate model at the regional level, where the number of groups is large, or the group-level sample sizes are small.

1.2.8 Concluding remarks

In summary, the progress of the thesis is built upon the key concepts mentioned above. First, understanding the concept of political trust and its importance in theoretical research is necessary. Second, the task of measuring political trust in a multiple groups setting requires that we 1) establish how to empirically quantify political trust in a multiple groups setting and 2) how to detect differences between political trust at different empirical levels. Finally, once the theoretical definitions and modelling framework are established, we will investigate the impact of geographical granularity (i.e. political trust at the sub-country regional level) on 1) our understanding of political trust as a theoretical concept and 2) the modelling steps needed to be undertaken to obtain robust empirical results. The intersection of this process and the issues that come with it will form the key domain of the thesis. We will over the course of the thesis illustrate that there are indeed meaningful regional differences in political trust in Europe.

In the following, we provide a structured summary of each chapter in the thesis.

1.3 Thesis overview

The main part of the thesis consists of three empirical chapters that successively build towards answering how we should measure and analyse political trust across countries and regions in Europe. This is followed by a concluding chapter, summarising the key results and the implications for future research.

The thesis is using the European Social Survey (ESS) datasets (European Social Survey 2017). We provide an overview of the data in Section 1.4, listing details on the characteristics of the ESS and the different subsets used in different parts of the thesis.

1.3.1 Chapter 2

In Chapter 2, we provide a definition of political trust and offer an overview of how political trust has previously been investigated in empirical research. Next, we set up a simple measurement model for national and supranational political trust across 18 countries using the European Social Survey dataset from 2016 through multiple groups confirmatory factor analysis. This is followed by an overview of how to establish measurement invariance using the 'classical' approach. Finally, we discuss whether national and supranational political trust can be regarded as two empirically distinct concepts and whether the concepts are measurement invariant across countries. Although full measurement invariance across all countries is not achievable, we find that national and supranational political trust is not only two empirically distinct concepts, but they are also highly correlated in most countries. Whereas Chapter 2 establish partial measurement invariance, we suggest relaxing the assumption that countries are homogenous across regions.

1.3.2 Chapter 3

In Chapter 3, we introduce the substantive reasoning for applying the measurement model at the regional level (NUTS regions, further described in Section 1.4) rather than the country level. We are not able to estimate a full NUTS-level model using conventional methods. This highlights several issues which are discussed in turn: how to estimate a model with many small-N groups and potential floor effects in the data. More specifically, we focus on the drawbacks of the maximum likelihood estimation procedure, Heywood cases and negative estimates of the residual variances, which are a major hindrance when using conventional measurement invariance methods.

1.3.3 Chapter 4

Having established in Chapter 3 that analysing ESS on the subnational level poses a number of analytical challenges, Chapter 4 explores an alternative to the standard frequentist approach. We translate the already established measurement model set up in Chapter 2 and 3. More specifically, we apply a full Bayesian multiple groups confirmatory factor analysis approach with phantom-latent variables in order to overcome the estimation issues in the small-N case. This includes a discussion on the choice of prior distributions drawing on the often-overlooked work by Gelman (2006), as well as a treatment of model fit and model diagnostics in the Bayesian framework. In addition to the model results, we propose an alternative method of model comparison using the empirical cumulative distribution function on the posterior deviances put forth by Aitkin (2010) and Aitkin, Vu, and Francis (2016). We finish the chapter by applying the same modelling framework on already existing research by (Märien 2011), to highlight how a NUTS-level model may change the substantive interpretation differences in political trust across Europe. One of the main findings is, that in cases with many groups, a Bayesian modelling framework is the preferred approach, although it is important to pay special attention to the choice of priors on problematic parameters.

1.3.4 Chapter 5

In Chapter 5, we conclude on the discoveries in previous chapters and discuss possible future directions for research on political trust at the regional level. One of the more obvious approaches is to extend the model to account for spatial autocorrelation at the regional level to further investigate both within- and between-country differences in political trust. The chapter also includes a posthoc analysis of the final model established in Chapter 4.

1.4 Data description

Throughout the thesis, we will make use of the European Social Survey (ESS) dataset (European Social Survey 2017). The ESS is a biannual and European-wide survey that measures attitudes and beliefs across more than 30 nations. The data is collected by the national statistical offices through a stratified multi-stage probability sample design set out by the ESS. The data is repeated cross-sectional in the sense that a new sample is drawn every round. The details of the sampling design and procedure are available from European Social Survey (2017). The design means that while we may study country or regional-level differences over time, we have limited scope for analysing between-wave changes within units.

The ESS datasets are of particular interest when investigating the research questions put forward in this thesis. First, the rigorous methodology behind the survey ensures a much better comparison across geographical regions, let it be countries or other regions of interest, compared to, for example, gathering vaguely similar survey data from multiple different sources. Second, the ESS contains relevant questions related to political trust in the core module, which means that they are repeated every round. In comparison, some topics are part of the rotating modules, which may only be asked in one particular round. Third, the ESS record regional-level information of the respondents, which will be presented in Section 1.4. Finally, the ESS is the main choice for many studies and publications within the field of measurement invariance and measurement error. As a result, we believe that it will extend the relevance of the methods applied to the data throughout the thesis to a broader pool of researchers.

1.4.1 NUTS classification

Outside the fact that the ESS surveys a large number of countries, we will also make use of their division of countries into regional. Nomenclature of Territorial Units for Statistics (denoted 'NUTS'). NUTS regions are subdivisions of countries used for statistical purposes, which was developed by the European Union (European Commission/Eurostat) in 2003. The NUTS classification covers all of the EU member states and a few other countries (EU candidate countries and EFTA countries). The current NUTS classification operates a three different levels, NUTS-1, NUTS-2 and NUTS-3. NUTS-1 is the broadest classification, splitting countries into a maximum of 16 regions. NUTS-2 is nested within NUTS-1, while NUTS-3 (the most fine-grained regional division) is nested within NUTS-2. Depending on the country size, NUTS-1 may simply refer to the country itself, or the country may not have a NUTS-2/3 classification at all.

The division of countries into NUTS regions is very useful when analysing survey data, since it provides a common framework for regional-level comparisons. We will make extensive use of this regional division throughout the thesis.

1.4.2 Data subsets, NUTS levels and sample sizes

In the following we will provide a brief overview of the subsets of the ESS data used in different parts of the thesis, how the data are distributed across NUTS regions and the relevant sample sizes. We only analyse complete cases throughout the thesis, which may result in minor differences in sample sizes compared to the raw datasets.

1.4.2.1 Section 1.5

Section 1.5 serves as a motivating example for why it is relevant to explore political trust at the regional level. We will focus on the UK, which in the ESS is divided into 12 NUTS-1 regions. The data used to present the trends over time is ESS round 1 through 9, while the statistical models are based on ESS 8 (2016). An overview of the NUTS regions and samples sizes for each round is presented in Table 1.1 .

1.4.2.2 Chapter 2

In Chapter 2, we introduce a simple measurement model at the country level. For this particular section, we will focus on round 8 of the ESS, since round 9 was not fully released at the time of writing. The sample sizes are listed in Table 1.2.

1.4.2.3 Chapter 3

In Chapter 3, we apply the model developed in Chapter 2 to the NUTS level across 16 of the 18 countries using ESS round 5-8 (Austria and Italy are excluded). The reason for not including Italy and Austria is, that they were not surveyed in two of the ESS rounds. Hence, we are only analysing the countries with complete data in ESS round 5-8. The intial sample size for ESS 8 is equivalent to that listed in 1.2, split into NUTS regions. However, due to estimation issues it is necessary to remove some NUTS regions. The resulting sub-sample for each NUTS and ESS round is listed in Appendix 3.

United Kingdom		**ESS round**								
NUTS code	**NUTS name**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
UKC	North East	88	97	85	87	86	92	100	68	111
UKD	North West	220	189	246	251	211	200	233	246	229
UKE	Yorkshire and the Humber	144	131	166	192	193	157	159	149	201
UKF	East Midlands	142	124	140	188	154	151	164	156	155
UKG	West Midlands	165	163	165	155	177	177	172	165	139
UKH	East of England	158	182	221	200	192	170	190	174	216
UKI	London	191	130	164	193	180	194	191	163	199
UKJ	South East	258	212	287	323	291	249	300	234	319
UKK	South West	145	159	201	213	167	168	197	155	198
UKL	Wales	128	66	122	118	113	104	122	123	91
UKM	Scotland	160	135	203	150	209	186	191	155	171
UKN	Northern Ireland	57	58	58	61	48	67	56	66	63
Total		**1,856**	**1,646**	**2,058**	**2,131**	**2,021**	**1,915**	**2,075**	**1,854**	**2,092**

Table 1.1: UK NUTS classifications and sample sizes in ESS 1-9.

Country	NUTS level	Regions	Low N	Low N region	High N	High N region	SS	SS/Region
Austria	2	9	64	AT11	383	AT13	1,844	205
Belgium	2	11	47	BE34	275	BE21	1,710	155
Czech Republic	3	14	57	CZ041	267	CZ010	2,120	151
Estonia	3	5	171	EE007	776	EE001	1,786	357
Finland	3	19	11	FI200	5261	FI1B1	1,857	98
France	2	21	23	FRI2	242	FRI0	1,977	94
Germany	1	16	21	DEC	470	DEA	2,678	167
Hungary	3	20	14	HU313/HU233	275	HU110	1,409	70
Ireland	3	8	162	IE063	596	IE061	2,274	284
Italy	2	20	13	ITH2	314	ITC4	2,376	119
Lithuania	3	10	79	LT027	466	LT011	1,862	186
Netherlands	2	12	30	NL23	278	NL33	1,529	127
Poland	2	16	31	PL52	183	PL9	1,425	89
Portugal	2	5	39	PT15	419	PT11	1,145	229
Slovenia	3	12	35	SI038	250	SI041	1,186	99
Spain	2	18	7	ES23/ES63	311	ES61	1,633	91
Sweden	3	21	9	SE212	263	SE110	1,368	65
United Kingdom	1	12	61	UKN	235	UKD	1,773	148

Table 1.2: ESS 8 country sample sizes.

1.4.2.4 Chapter 4

In the main body of Chapter 4, we apply a Bayesian modelling framework to the NUTS-level equivalent to the measurement model put forward in Chapter 2. The sampled countries and sample sizes are the same as in Table 1.2. In the later part of the chapter, where we apply the same methodology on a different definition of political trust put forward by Märien (2011). The sample is extended to use all of the sampled countries which have a NUTS classification avilable. This, in addition to the 18 countries previously used, includes Iceland, Norway and Switzerland (all EFTA countries). The country characteristics and sample sizes are listed in the relevant section.

1.4.3 Survey questions

Throughout the thesis, we will use a few selected questions from the ESS survey. They are from the same battery of questions, which are introduced using the interviewer card:

"Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly… READ OUT…:"

1. *…[country]'s parliament?*
2. *…the legal system?*
3. *…the police?*
4. *…politicians?*
5. *…political parties?*

6. *...the European Parliament?*

7. *...the United Nations?*

In Section 1.5 we will focus on (1) and (6) above. In Chapter 2, 3 and 4, we will establish a measurement model where (1), (4) and (5) consitutes a measure for 'National Political Trust', while (6) and (7) constitutes a measure for 'Supranational Political Trust'. In the last part of Chapter 4, we will also include (2) and (3). The exact definitions and the theoretical background behind this choice is further explained in the relevant chapters.

1.5 A motivating example: why is it relevant to explore political trust at the regional level?

It has been a turbulent time in the EU, with the rise of nationalism, economic instability, and populism, something that at least partially has been linked to the public's trust in the institutions (Inglehart and Norris 2016). A recent and salient expression of these sentiments is represented by the EU referendum in the UK. On a local scale, 'Brexit' demonstrates the complexities in how different notions of political trust are related and how these change over time and differ across regions. We proceed to present the EU referendum as a motivating example that also serves to illustrate the relevance of the trust constructs on a regional level.

In June 2016, the UK voted to leave the European Union. This provides a unique opportunity to investigate the trends in political trust within the UK, which can then be linked directly to the outcome of the referendum in order to highlight why within-country differences are important.

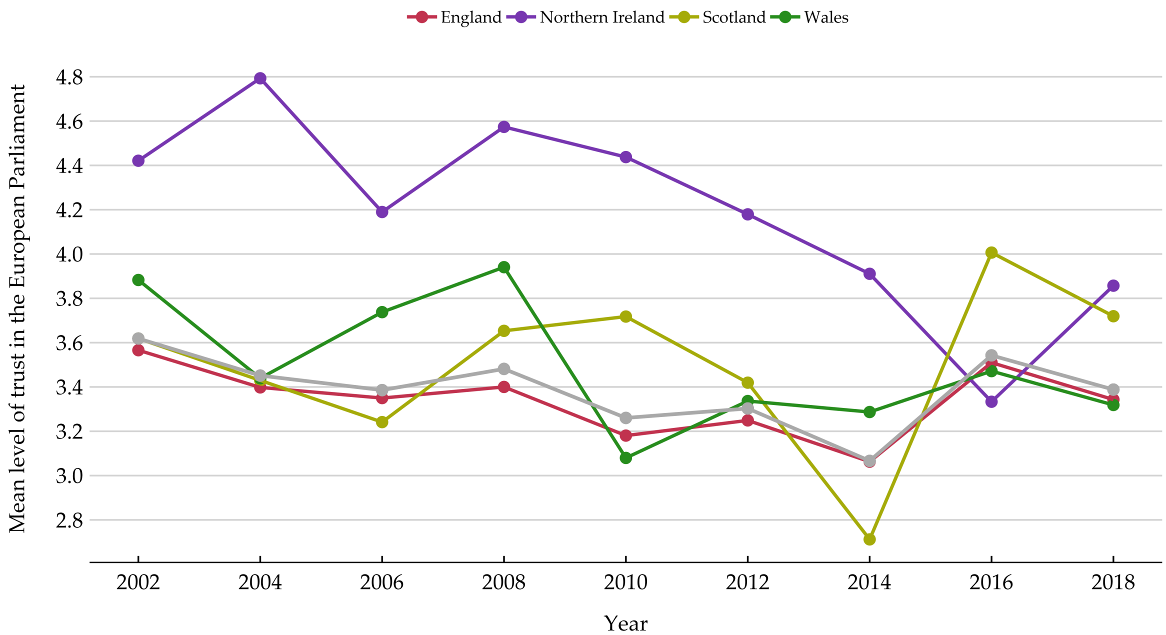


Figure 1.1: Trust in the European Parliament in the UK from 2002-2018.
The grey line represents the UK average.

We will focus on two questions from the European Social Survey, which asks about the respondents' trust in the national parliament and in the European Parliament, both on a scale from 0-10, where 0 indicates 'no trust at all' and 10 indicates 'complete trust'. In a simple figure, we can observe the development over time from 2002 to 2018 (corresponding to ESS 1 to ESS 9) as shown in Figure 1.1 and 1.2, where regions have been clustered into England, Scotland, Wales and Northern Ireland. In the sequel we use the term 'country' to refer to states and nations to be consistent with the rest of the thesis and thus we do not refer to the countries that make up the UK when we use 'country' here. Overall, the trust in the European Parliament thas gone down since 2008 in Wales, Scotland and Northern Ireland, while it has been more stable for England. However, it stands out that Scotland in 2014 reached an overall low point in trust for both the national and European Parliament. This may have to do with the fact that Scotland had a referendum for

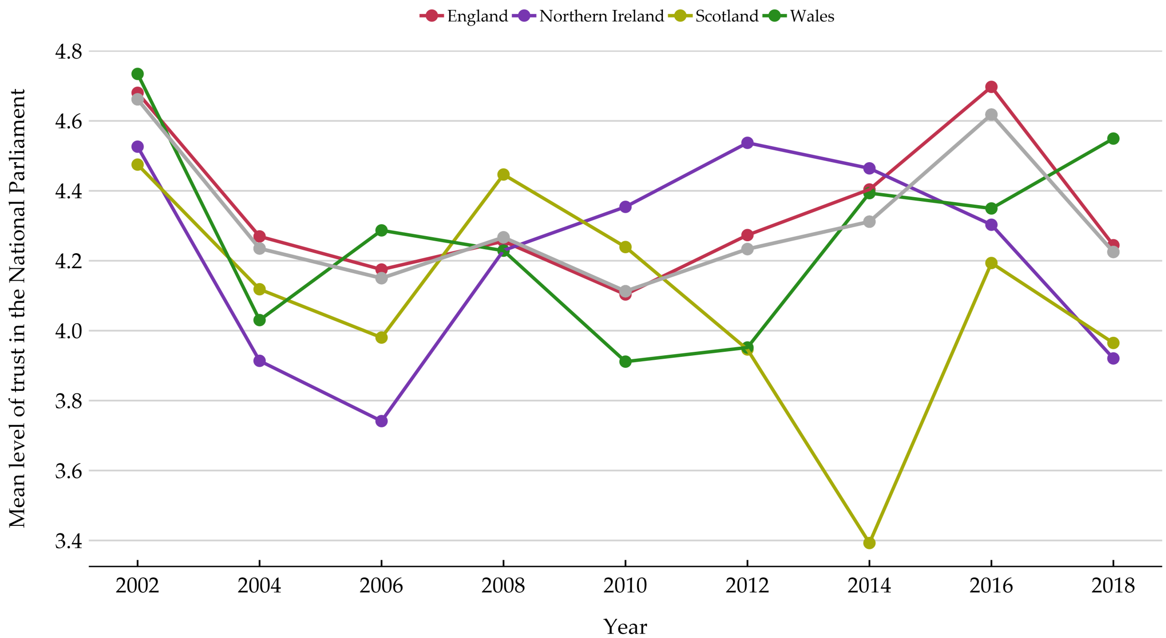


Figure 1.2: Trust in the National Parliament in the UK from 2002-2018.
The grey line represents the UK average.

independence, although this is just pure speculation.

It is clear that trust in the European Parliament cannot be considered independently of trust in the national parliament. Indeed, Brexit has widely been seen as a protest vote against the national parliament (e.g. Ryan 2016). To investigate how trust in the EU varies as a function of trust in the national parliament while taking into account the commonly cited effect of education on the leave vote (Hobolt 2016), we fit a regression. To account for regional variability, we let y_{ij} be trust in the European Parliament of individual i in region j, and fit $y_{ij} = \eta_{0j} + \beta X_{ij} + \epsilon_{ij}$, where $\eta_{0j} = \gamma_{00} + \zeta_{0j}$ is a regional random effect (Skrondal and Rabe-Hesketh 2004). The ESS has 12 regions in total recorded for the UK, which is strictly speaking too few regions to estimate a random effects model properly. However, this is sufficient for illustrative purposes, although we choose not to model the trend over time. We did fit a model with fixed time effects, but it did not affect the other parameter

	Single level model		Multilevel model	
Fixed effects	**Est**	**S.E.**	**Est**	**S.E.**
Constant	2.967***	0.282	2.984***	0.292
Trust in National Parliament	0.503***	0.020	0.505***	0.020
Gender (1=female)	0.357***	0.094	0.358***	0.094
Age	-0.029***	0.003	-0.030***	0.003
Citizen of country (1=yes)	-0.599**	0.215	-0.584**	0.214
Education (ISCED II)	-0.040	0.175	-0.063	0.174
Education (ISCED IIIb)	-0.047	0.182	-0.014	0.180
Education (ISCED IIIa)	-0.018	0.177	-0.006	0.176
Education (ISCED IV)	0.118	0.159	0.109	0.158
Education (ISCED V1)	0.266	0.171	0.268	0.170
Education (ISCED V2)	0.634***	0.168	0.627***	0.168
Random effects				
Intercept std. dev. ($\sqrt{\psi}$)	-	-	0.275	0.075
Log-likelihood	-3749.5		-3740.2	

Table 1.3: Simple single and multilevel models.
Outcome variable: Trust in the European Parliament.
The sample consists of 1,783 respondents across 12 regions.
** = $p < 0.1$, ** = $p < 0.05$, *** = $p < 0.01$*

estimates. Using year-dummies, we found only round 6 to have a significant effect on trust in the European Parliament (estimate: -0.32, z-score: -4.37). In order to further investigate a potential trend over time, it would be necessary to set up a growth curve model. However, this is outside the scope of this motivating example.

The single level and multilevel models are summarised in Table 1.3. The positive effect of trust in the national parliament on trust in the European parliament is highly significant, with each one-point increase trust in the national parliament corresponding to a 0.5 point increase in trust in the European Parliament on an 11-point scale, controlling for citizenship, educational level, gender, and age.

The scatterplot of the best linear unbiased predicted random intercepts at the

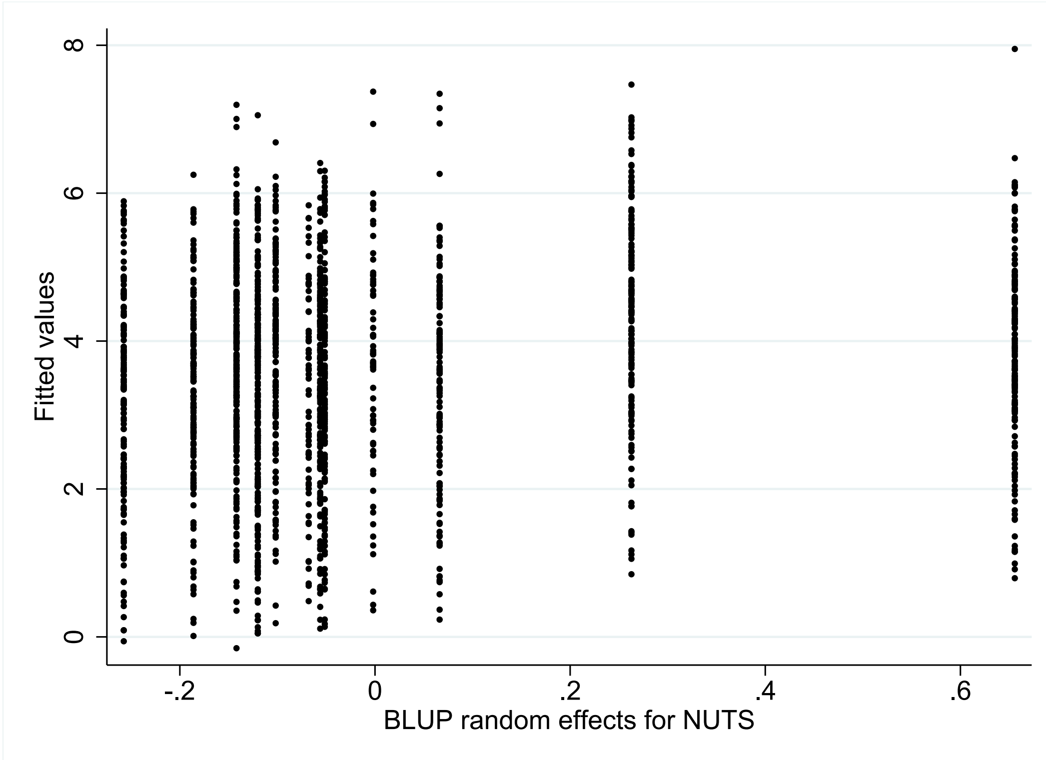


Figure 1.3: Multilevel model random effects and fitted values for 12 regions in the UK.

regional level as shown in Figure 1.3. The scatterplot shows high variability in the predicted random intercept for each region, indicating significant within-country variability. The exact values are listed in Appendix 1. The highest random intercepts are in Scotland (0.66), London (0.26) and Wales (0.07), while the lowest intercepts are in the West Midlands (-0.26) and the South West (-0.19). In other words, while it is tempting to draw conclusions about country differences (recall, here we use country for the state, i.e. the UK) and their explanatory factor based on the average levels, there are considerable differences between regions to the point where some regions in different countries may be more similar than regions in their own country. To further expand on the idea of regional differences in levels of political trust, we will continue to examine the recent UK referendum.

Previous studies have studied the reasons for the outcome of the Brexit referendum (Colatone and Stanig 2018; Clarke, Goodwin, and Whiteley 2017), how it varied across regions (Hobolt 2016), and what impact it might have on the UK social policies (Pinker 2017). Inglehart and Norris (2016) describe the Brexit vote as a result of populism and the cultural backlash against:

> “[. . .] increased tolerance among the younger cohorts and the college educated living in Western societies for the expression of diverse forms of sexuality, LGBT rights, same-sex marriage and varied family units, and more fluid gender identities; more secular values, habits, and ethical norms; open-mindedness towards migrants, refugees, foreigners, and multicultural diversity of lifestyles, foods, and travel; and cosmopolitan support for international cooperation, humanitarian assistance, and multilateral agencies like the United Nations and EU.” (Inglehart and Norris 2016)

They specifically use the ESS from 2002-2014 to support the cultural backlash thesis by examining how economic insecurity and cultural values act as predictors of voting for populist parties. This goes in line with Hobolt (2016) who have been analysing survey data to arrive at the conclusion that people in favour of Brexit, i.e. leaving the EU, were characterized by having lower levels of education, being older, poorer and expressing concerns about immigration compared to the ‘remain’-voters (Hobolt 2016). In addition, she finds that “[...] the Remain side did better in the larger multicultural cities (especially in London) and where there were more graduates, whereas the Leave side was strongest in the English countryside and in the post-industrial north-eastern towns with larger working class populations.”

(Hobolt 2016). This statement is further reinforced by the fact that England and Wales voted 53% Leave, whereas Northern Ireland (56%) and Scotland (62%) had a majority voting Remain.

Although the literature discusses both the difference between political trust in national entities and supranational entities (i.e. what Inglehart and Norris (2016) describes as multilateral agencies) and how the Brexit vote is associated with regional differences, they are not analysed in conjunction.

Figure 1.4 shows a scatterplot for all 12 UK NUTS regions of trust in the national/European Parliament and referendum results side by side. The correlation between trust in the European Parliament and the outcome of the referendum is slightly positive (0.42) while there is no sign of a correlation between trust in the national parliament and the outcome of the referendum (-0.03). This is to be expected under the assumption that trust in the national parliament has no effect on the outcome of the referendum, i.e. a clear difference between measures of political trust at different levels. However, looking at a scatterplot in Figure 1.5 of the level of the two trust measures and the outcome of the referendum shows some interesting results.

The correlation between the two measures of trust and the referendum results are surprisingly high (0.61). The relation between trust in the European Parliament on the referendum result can only be understood in relation to the level of trust in the national parliament. How large trust in the European Parliament is as a fraction of trust in the national parliament tells us something about how the two concepts are related to a tangible outcome (the difference between the two is not as strong a predictor). Not only does this indicate that the relationship between them is relevant, but it also shows the necessity of taking into account within-country differences when we try to understand political trust in general. Of course,

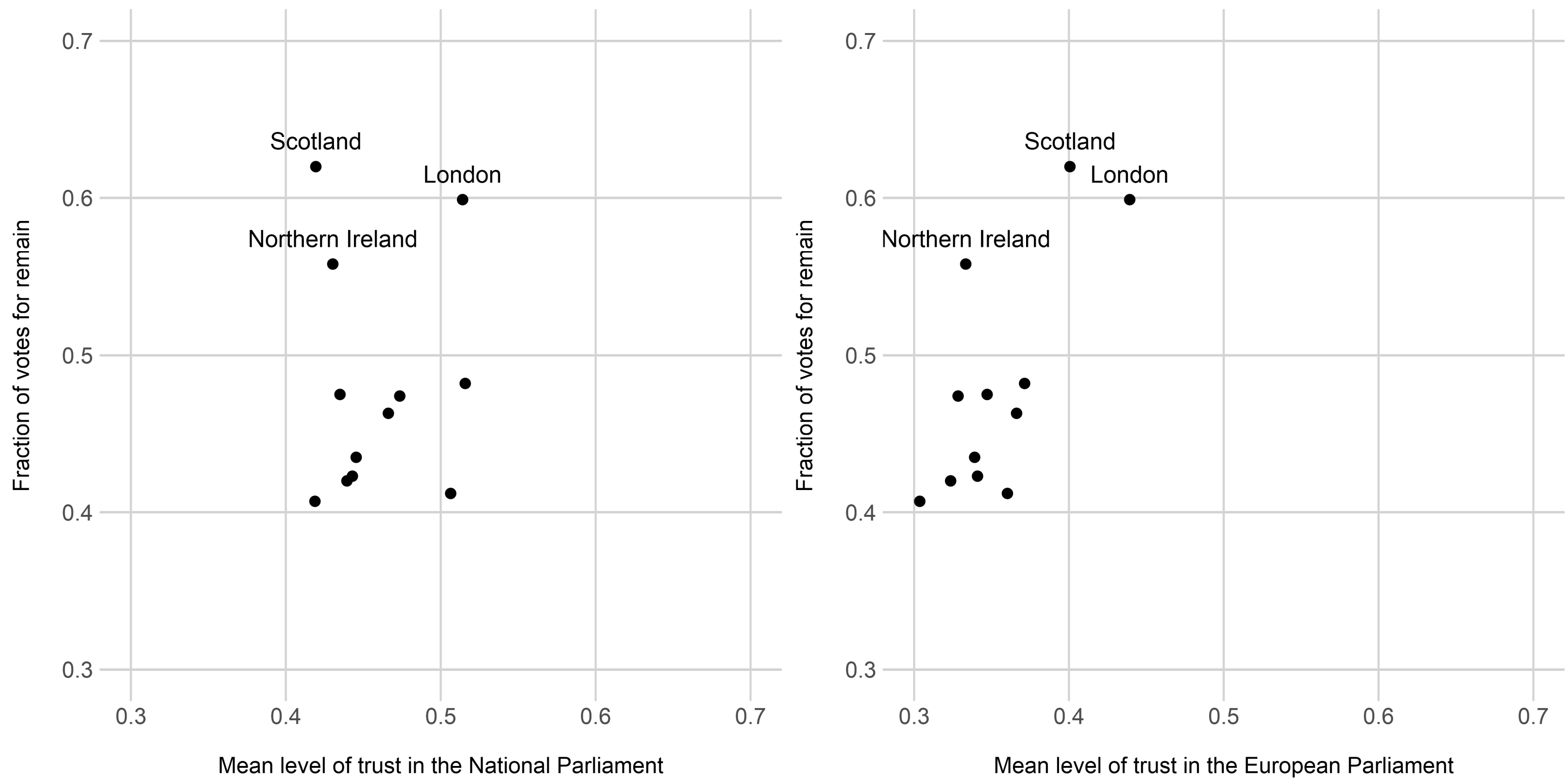


Figure 1.4: Scatterplot of trust in National Parliament and trust in the European Parliament vs. fraction of votes on remain in the 2016 UK referendum across NUTS regions.
Trust measures have been rescaled to [0 ; 1]

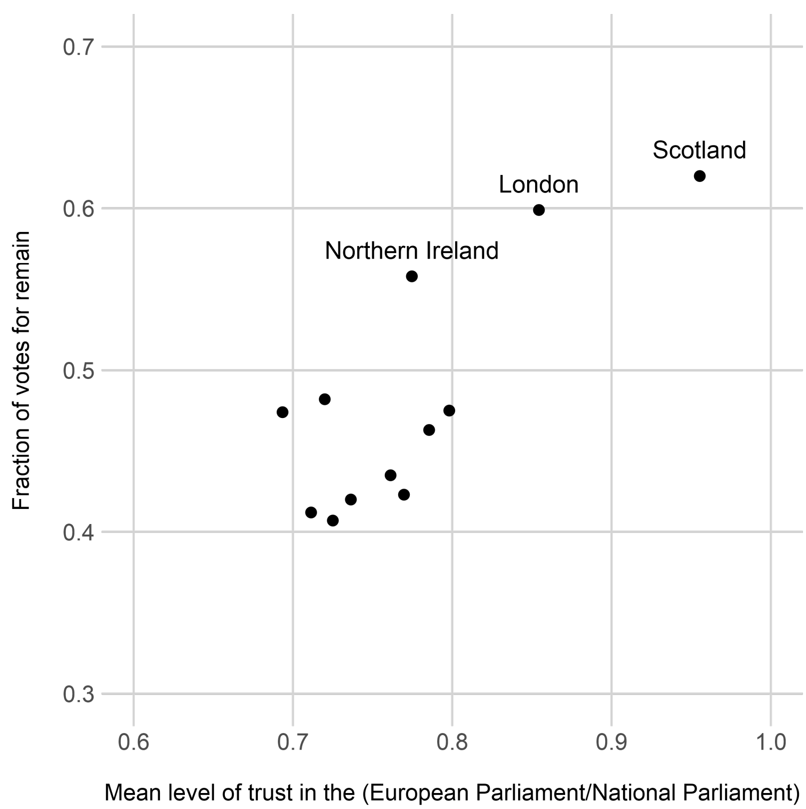


Figure 1.5: Scatterplot trust in the (European Parliament/National Parliament) vs. fraction of votes on remain in the 2016 UK referendum across NUTS regions.
Trust measures have been rescaled to [0 ; 1]

such scatterplots are based on aggregated regional means and not individual-level data, which clearly limits their descriptive power. However, they indicate that such a relationship on the regional level should be further investigated. Popular prediction models, as well as post-referendum analyses, pointed to the same predictors for voting to remain as predicted by trust in the national parliament in the simple model in Table 1.3 (Lauderdale, 2016). Since both measures of trust are clear antecedents to the referendum outcome, understanding the functional relationship between trust in the national and the European Parliament in affecting the vote informs us of the validity of any underlying measure.

Chapter 2

An investigation of measurement and association of different levels of political trust in 18 EU countries using the European Social Survey 2016

In this chapter we analyse the relationship between measures of national and supranational political trust, focusing on their correlation across 18 European Union (EU) member states. Using multiple groups confirmatory factor analysis (MGCFA) of the European Social Survey 2016, we give evidence for an empirical distinction between these two concepts of political trust. When investigating measurement invariance, full scalar invariance does not hold. However, by freeing a few intercepts, we can obtain partial scalar invariance. Our central finding is a strong correlation between national and supranational political trust across most

EU countries. Furthermore, no noticeable differences are found in levels of political trust based on a 'stable' versus 'new' democracies dichotomy within the EU. The results indicate the necessity of analysing national and supranational political trust simultaneously. Our measurement model is meant as a stepping stone for future analyses, which can further enlighten the dynamic between political trust at different empirical levels.

2.1 Trust in the EU

Trust in the EU is facing a range of challenges, partially due to events such as the 2008 financial crisis and the 2015 migration crisis. Furthermore, the 2014 European Parliament election had the lowest ever voter turnout and resulted in increased support for Eurosceptic parties across multiple countries. Empirical studies and cross-national surveys document the lack of trust in the EU and its institutions, demonstrating how it has reached a low-point with less than fifty per cent of the EU population 'tending to trust the EU' (European Commission 2020). However, the methodology of the Eurobarometer is disputed, so this number may be even lower ((Bennike 2019). This empirical evidence is a topic for philosophers and social scientists alike, who have been questioning the well-functioning of the EU and the ongoing legitimacy issues for a while (Habermas 2013; Zielonka 2014). With the most recent case of interest being the 2016 referendum on the UK's EU membership, possibly resulting in an EU member state leaving the Union for the first time in its history, the topic of political trust is as relevant as ever. Yet, while plenty of literature exists on political trust on the European as well as the national level, little research has been conducted on the relationship between political trust at different empirical levels.

Do low levels of political trust at the supranational level reflect lower political trust at the national level in the EU context? Or do national political factors have little relation with how much trust citizens have in supranational governance? We start by giving a definition of political trust and how it relates to social trust. A brief review of political trust in the EU follows, focusing on the potential differences in levels of political trust across different types of democracies. Next, we present the data and the model used to investigate the relationship between national and supranational political trust, followed by an explanation of the issue of measurement invariance. Finally, we present the model results evaluated in relation to the initial discussion on political trust and suggestions for relevant future work.

2.2 A definition of political trust

In the context of political systems, it is necessary to make a clear distinction between political trust and social trust. According to Newton (1999) "social trust is a feature of the most basic level of community, while political trust refers primarily to attitudes about political institutions and leaders". One key difference is that social trust is characterised by "social connections established in circumstances of co-presence" (Giddens 1991), while political trust is built upon citizens' knowledge of others, let it be leaders, institutions or politicians. In other words, social trust occurs within the private sphere, while political trust is a feature of the public political sphere (Newton 1999).

This aligns accurately with the idea presented by Giddens (1991, p. 83) that "[...] the nature of modern institutions is deeply bound up with the mechanisms of trust in abstract systems, especially trust in expert systems". According to Giddens, modern trust is expressed through faceless commitments, which can be understood as a commitment to, or faith in, symbolic tokens and abstract systems rather than personal knowledge of others. Consequently, possible definitions of political trust include citizens' normative evaluation of government performance (Hetherington 1998) or citizens' belief in political institutionalised practices and procedures (Sztompka 2003).

We shall adopt the latter as a working definition, based on the claim that political trust encompasses future expectations and belief in institutionalised practices and procedures, regardless of its current or past performance. The key reason why political trust is of particular interest when determining the well-being of a political system is due to it being the primary constituent of social capital, which is necessary to ensure both democratic stability and economic efficiency (Putnam 1995; Fukuyama 1995) as well as to provide the necessary conditions needed for a well-functioning civil society (Newton 2001).

From an EU perspective, we believe that it is necessary to consider political trust at different empirical levels. Following the idea of multilevel governance, where political decision-making processes and structures are defined to operate at multiple empirical levels (Bache and Flinders 2004), we also expect political trust to be specific to the level of governance in question. Hence, it is assumed that citizens' political trust is unique for each level, let it be EU institutions, national governments, regional entities etc. In the following, we focus our attention on two different notions of political trust: national political trust (political trust the national governments and institutions) and supranational political trust (political trust in supranational political institutions and organisations).

Empirical research surrounding the concept of political trust is already present, although limited. One example is Cordero and Simón (2015), who argue that political trust is directly linked to the degree to which citizens are satisfied with the way democracy works in their country. This statement aligns with André (2013) who points out how political trust in an empirical context should be considered a democratic resource. However, it is important to note that determining the reasons for distinct levels of political trust across countries is far from a trivial task. It depends both on how it is measured and the theoretical orientation of choice. Miller and Listhaug (1999) sums up how previous studies have managed to explain changes in trust in national governments as a result of either 1) government performance, which directly influences trust in political institutions and hence the regime legitimacy or 2) different levels of expectations from the citizen where high expectations lead to lower levels of trust in the said political institution.

In the context of this study, it is reasonable to assume that these explanations are possible both at the national and supranational level, albeit it is unknown whether the same explanation applies in both cases.

2.3 Political trust in a European perspective

Within a European context, it is worth highlighting the notable differences in levels of political trust and trust in democracy across countries in Europe. Numerous countries have undergone drastic changes to the democratic rule and the role of political institutions since the Second World War. Anderson et al. (2005) arrive at a classification of European countries based on regime change and levels of freedom; stable democracies (Western European and Southern European countries), new democracies (post-communist countries in Eastern Europe) and transitional democracies (Croatia, Belarus, Ukraine, Belarus) based on the Freedom House index. Within this three-way classification, support for democracy is highest in stable democracies and lowest in transitional regimes with the post-communist democracies somewhere in-between (Anderson et al. 2005).

Even though many years have passed since the Revolutions of 1989, the stable/new democracy dichotomy is still reflected in the length of EU memberships. Several countries have been members for a long time, including the Netherlands, France and Germany (founders in 1957) while some countries joined in the 1970s like the United Kingdom and Ireland. On the contrary, most post-communist countries have joined the Union much later, such as the Czech Republic in 2004, Hungary in 2004, Bulgaria in 2007 and Romania in 2007.

Existing research within this democracy classification shows how citizens from post-communist countries possess initial reduced levels of political trust in democratic institutions. One explanation is, that the political transition from an authoritarian regime results in performance issues for newly created democratic political institutions, which in turn leads to low initial levels of political trust (Mishler and Rose 2001). An example of the opposite scenario is the Nordic countries (Denmark, Norway, Sweden and Finland) which have a history of high levels of trust, let it be national or supranational. Delhey and Newton (2005) considers the Nordic countries to be the most trusting due to the well-functioning of the government and income equality. Consequently, citizens of countries with more than 20 years of being an established democracy show significantly higher levels of trust in political institutions (Delhey and Newton 2005).

However, an important question is whether differences in levels of political trust still should be characterised by a stable/new democracy dichotomy. First, all current EU member states have been democracies for more than twenty years and members of the EU for at least ten years (except Croatia). Second, it is unknown whether political (dis)trust manifests itself at the national or supranational level or both. While it is our expectation that national political trust is higher for stable democracies, it does not necessarily form a prerequisite for high levels of supranational trust. As a result, the EU is an interesting case due to the well-known observed cultural, historical and political differences between member countries.

2.4 Political trust in empirical research

Political trust is a complex and multidimensional concept which is difficult to measure (Saris and Gallhofer 2014). Within the political sciences, the majority of empirical work on political trust is either centred on how the concept is measured (Marozzi 2014), how it changes over time (van der Meer 2010; Klingemann and Weldon 2012) or how it relates to some outcome of interest (Torcal and Montero 1999; Newton 2001). It is noteworthy how different authors use different items, sometimes from the same survey, to identify measures of political trust. Rothstein and Stolle (2008) used principal component analysis to derive three different dimensions, namely the partisan institutions (political parties, politicians, government, civil services), the neutral and order institutions (army, police and legal institutions) and the power-checking institutions (press and television).

A different approach is taken in Allum, Read, and Sturgis (2018), who uses confirmatory factor analysis on the European Social Survey (ESS) to derive two dimensions; 1) trust in the parliament and politicians and 2) trust in the legal system, police, European Parliament (EP) and the United Nations (UN). Also using the ESS, Märien (2011) includes a broad range of indicators for political trust (trust in the parliament, politicians, political parties, legal system and police) and detects correlated errors between trust in the legal system and the police. A more recent study by Schneider (2016) takes an explorative approach to arrive at four different measurements models. Her main finding is that trust in the government, parliament and political parties belongs to a different dimension than trust in protective institutions (armed forces and the police) and trust in order institutions (courts and police) (Schneider 2016).

The approach taken in this chapter is closely aligned with André (2013), although with a different purpose. Her focus is to test equivalence in trust between EU natives and migrants using multiple groups confirmatory factor analysis (MGCFA) on the ESS and arrives as a three-dimensional (three-factor) model comprising 'distinctively political' (trust in politicians, political parties and politicians), 'order/neutral' (trust in the police and the legal system) and 'international' (trust in the European Paliament and United Nations) (André 2013). Since this study specifically investigates the correlation between political trust at the national and supranational level, 'distinctively political' translates to 'national political' and 'international' to 'supranational political'. We consider the UN and the EP to be supranational political institutions since they can be said to have political power which transcends that of a single nation-state.

To sum up, three indicators are used for national political trust (NPT). All responses on a 0-10 scale with 0 referring 'no trust at all' and 10 referring to 'complete trust':

1. Trust in the national parliament (PRL)
2. Trust in politicians (PLT)
3. Trust in political parties (PRT)

As a measure of supranational political trust (SPT), two indicators are used:

1. Trust in the European Parliament (EP)
2. Trust in the United Nations (UN)

It is worth noting that these indicators relate to trust in political institutions, rather than to political engagement or communities, thus following the definition of political trust as previously mentioned.

2.5 A measurement model for national and supranational political trust

The relationship between national political trust (NPT) and supranational political trust (SPT) is studied across all available EU countries in 2016 using data from the European Social Survey (wave 8), resulting in a total sample size of 31,952 across 18 countries. We have selected EU countries only, since SPT contains an item related to the European Parliament. Only complete cases are studied, i.e. observations with missing values on any of the indicators are disregarded.

The model investigates the relationship between two latent constructs (NPT and SPT) with the indicators stated above in a simple measurement model using the multiple groups confirmatory factor analysis (MGCFA) method. It is well known that confirmatory factor analysis (CFA) methods are superior to alternative approaches like bivariate correlations, univariate regression analysis and Cronbach's alpha (Brown et al. 2015). A list of countries, including sample sizes, mean country values and standard deviations for each item is listed in Table 2.1.

Initially, we observe that trust in the UN is highest (mean value of 5.14), followed by trust in the national parliament (mean value of 4.49) and trust in the EP (mean value of 4.37). In addition, the levels of trust in the different institutions vary considerably between countries. For example, trust in the national parliament

Country	**Sample size**	**National Political Trust**			**Supranational Political Trust**	
		Parliament	Politicians	Parties	EU Parliament	United Nations
Austria	1,844	4.98 (2.39)	3.91 (2.41)	3.84 (2.39)	3.82 (2.59)	4.41 (2.68)
Belgium	1,710	4.80 (2.21)	4.12 (2.17)	4.05 (2.16)	4.63 (2.40)	5.27 (2.33)
Czech Republic	2,120	4.33 (2.36)	3.64 (2.33)	3.58 (2.29)	4.16 (2.58)	4.93 (2.59)
Germany	2,678	5.24 (2.42)	4.10 (2.23)	4.12 (2.12)	4.36 (2.41)	4.84 (2.37)
Estonia	1,786	4.56 (2.42)	3.66 (2.21)	3.60 (2.15)	4.59 (2.41)	5.07 (2.48)
Spain	1,633	3.93 (2.58)	2.45 (2.29)	2.45 (2.23)	4.24 (2.57)	4.71 (2.69)
Finland	1,857	5.76 (2.26)	4.76 (2.18)	4.86 (2.14)	5.18 (2.25)	6.39 (2.14)
France	1,977	4.04 (2.31)	2.89 (2.10)	2.82 (2.06)	3.73 (2.37)	4.78 (2.46)
United Kingdom	1,773	4.63 (2.39)	3.72 (2.26)	3.80 (2.13)	3.58 (2.45)	5.23 (2.41)
Hungary	1,409	4.51 (2.57)	3.77 (2.57)	3.62 (2.49)	4.44 (2.50)	5.00 (2.49)
Ireland	2,274	4.54 (2.30)	3.79 (2.36)	3.78 (2.29)	4.98 (2.25)	5.70 (2.21)
Italy	2,376	3.24 (2.51)	2.31 (2.26)	2.31 (2.24)	4.00 (2.56)	4.43 (2.59)
Lithuania	1,862	3.76 (2.29)	3.52 (2.32)	3.30 (2.33)	5.44 (2.48)	5.58 (2.51)
Netherlands	1,529	5.52 (1.97)	5.07 (1.94)	5.11 (1.92)	4.62 (2.16)	5.67 (1.89)
Poland	1,425	3.44 (2.58)	2.55 (2.20)	2.51 (2.14)	3.93 (2.42)	4.88 (2.48)
Portugal	1,145	3.99 (2.67)	2.48 (2.20)	2.62 (2.28)	4.08 (2.64)	5.40 (2.71)
Sweden	1,368	6.00 (2.27)	4.77 (2.04)	4.84 (1.97)	4.77 (2.12)	6.16 (1.99)
Slovenia	1,186	3.36 (2.34)	2.46 (2.13)	2.48 (2.10)	3.94 (2.48)	4.34 (2.50)
Total/mean	**31,952**	**4.49(2.50)**	**3.58(2.38)**	**3.56(2.35)**	**4.37(2.48)**	**5.14(2.49)**

Table 2.1: Country sample sizes, item means and standard deviations in ESS 8.

is 6.00 in Sweden and only 3.36 in Slovenia, while trust in the EP is highest in Lithuania (5.44) and lowest in United Kingdom (3.58). In order to determine whether the items are a representation of two underlying latent variables, it is necessary to define the CFA and MGCFA models.

Given a $p \times 1$ vector of observed variables $\boldsymbol{y} = \left[y_1, \ldots, y_p\right]^T$ and a $q \times 1$ vector of latent variables $\boldsymbol{\eta} = \left[\eta_1, \ldots, \eta_q\right]^T$, a basic linear structural equation model (SEM) is defined as (Song and Lee 2012).

$$\boldsymbol{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{\tau}$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings and $\boldsymbol{\epsilon}$ is a $p \times 1$ random vector of residual errors. Since the first three items are linked to NPT and the last two are linked to SPT, the measurement model with a non-overlapping factor loading matrix becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

A graphical representation of the model is in Figure 2.1.

Both NPT and SPT are latent independent variables and hence we can set $\boldsymbol{\eta} = \boldsymbol{\delta}$, which is a $q \times q$ vector of residual errors with mean κ. In a simple measurement model, three key assumptions are necessary, namely that 1) the random vectors of residual errors $\boldsymbol{\epsilon}$ is independent and identically distributed (i.i.d) such that $\epsilon_i \sim N(0, \Psi_\epsilon)$, where Ψ_ϵ is a diagonal covariance matrix and 2) the random

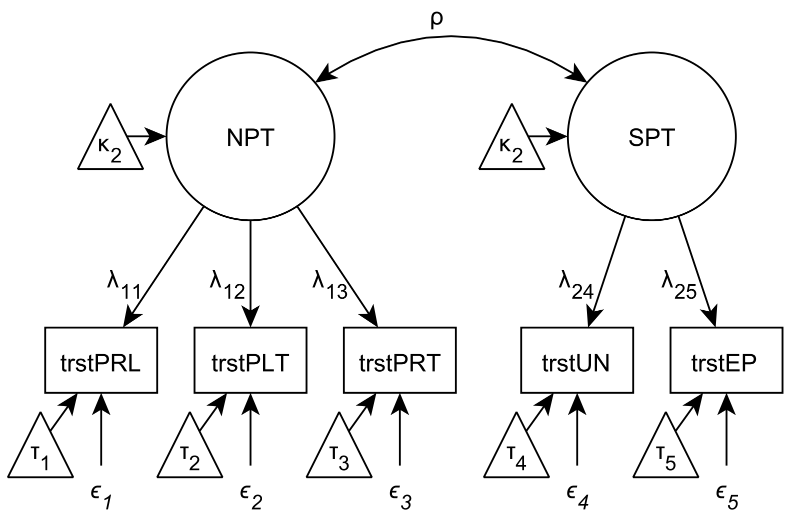


Figure 2.1: Simple measurement model.

vector of residual errors δ is i.i.d such that $\delta_i \sim N(0, \Psi_\delta)$, where Ψ_δ is a diagonal covariance matrix and 3) ϵ_i is independent of δ_i, that is $\text{Cov}(\epsilon_i, \delta_i) = 0$. Extending the equation to multiple groups is straightforward. Given a number of groups $g = 1, \ldots, G$ the extension is (Steenkamp and Baumgartner 1998; Davidov et al. 2014):

$$\boldsymbol{y}^g = \boldsymbol{\tau}^g + \boldsymbol{\Lambda}^g \boldsymbol{\eta}^g + \boldsymbol{\epsilon}^g,$$

where the notation follows that of equation 2.1. The following models are estimated using effects coding, that is setting $\sum_{i=1}^{I} \lambda_{iq}^{g} = I$, such that the sum of all factor loadings linked to a latent variable equals the total number of unique loadings for that latent variable in each of the G groups (Little, Slegers, and Card 2006). At the same time, by letting $\sum_{i=1}^{H} \tau_i^g = 0$, where H is the number of intercepts unique to each factor, the metric of the latent variables is set equivalent to that of the items. In other words, the latent means will be on the same scale as the survey questions, which has the clear advantage of making interpretations of the latent means more intuitive.

2.6 Assessing measurement invariance

When comparing groups, the multiple groups confirmatory factor analysis (MGCFA) method is applied, which allows for comparison of parameter estimates between countries. However, to compare the latent means of NPT and SPT in all 18 countries, the estimated model ideally should be measurement invariant across all groups. Previous studies such as Ariely and Davidov (2010) and van de Schoot, Lugtig, and Hox (2012) shows how it is necessary to obtain some degree of measurement invariance to not only guarantee that the latent constructs are measured in the same way across groups but also to bring in meaningful comparisons of the correlation between the latent means. If partial measurement invariance is impossible to obtain, significant differences in the measurement of NPT and SPT might as well be a result of measurement error.

The most common approach is to test three different models when assessing measurement invariance; the configural invariance model, the metric invariance model and the scalar invariance model (Steenkamp and Baumgartner 1998).

- Configural invariance model: This model only assumes that the factor structure is equal across groups, letting all parameters vary freely. If configural invariance is obtained, we can conclude that the same items measure NPT and SPT across all groups.
- Metric invariance model: Also called weak factorial invariance. The same as the configural invariance model, with the added constraint that factor

loadings are fixed to be equal across groups, that is setting $\Lambda^g = \Lambda \forall g \in G$. If metric invariance is obtained, we can conclude that the relationships between NPT, SPT and the items are the same across all groups.

- Scalar invariance model: Also called strong factorial invariance. The same as the metric invariance model, but with the added constraint that item intercepts are fixed to be equal across groups, that is setting $\tau^g = \tau \forall g \in G$. If scalar invariance is obtained, we can conclude that the mean values of the latent variables NPT/SPT correspond to a unique value on each of the items which is equal across all groups. Scalar invariance also makes it is possible to compare differences in latent means across groups.

It is important to note that assessing differences in the correlation parameter between the two latent variables across groups only requires metric invariance. Nonetheless, scalar invariance is pursued to motivate the interpretation of these differences. Often, it is difficult to obtain scalar invariance, due to model misspecification or non-normal item distributions (Bentler and Chou 1992). However, partial measurement invariance may be sufficient for comparisons to be valid across countries (Saris and Gallhofer 2014), for example, when only one item per factor is allowed to vary freely (Byrne, Shavelson, and Muthén 1989). However, current research on the effect of partial scalar invariance and its impact on latent mean estimation is limited and highly debated (Putnick and Bornstein 2016). We will take the approach of Chen (2008) in the case of partial measurement invariance, by comparing the substantive conclusions, including latent mean estimations, across relevant models and evaluate whether the non-invariance has any impact on the results.

For this chapter, it is necessary to determine whether national political trust (NPT) and supranational political trust (SPT) can be regarded as two empirically distinct concepts. This is done by testing a one-factor model encompassing all five indicators against the proposed two-factor model shown in Figure 2.1 and comparing the fit indices.

When examining the MGCFA models, numerous methods of assessing goodness-of-fit is applied. First, the raw χ^2 value for the model of interest is reported. However, it has been shown that model comparison based on differences the raw χ^2 values tend to favour potentially problematic models when the sample size is small and reject sensible models when the sample size is large (Hooper, Coughlan, and Mullen 2008). In this case, since the sample size is above 30,000, it is expected that an increase in model complexity will lead to a significant increase in chi-squared value. Second, the Comparative Fit Index (CFI) is reported, where values above 0.95 are considered 'good model fit' (tze Hu and Bentler 1999). When comparing competing models, Chen (2007) suggests that a decrease in CFI of less than 0.01 between models is adequately close to zero, thereby allowing one to draw conclusions based on the less complex model. Finally, two absolute fit measures, namely the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) are reported for model comparison. MacCallum, Browne, and Sugawara (1996) argue that RMSEA values below 0.05 indicate 'good' fit and values equal to or above 0.05 and below 0.08 indicate 'mediocre' fit. tze Hu and Bentler (1999) suggest that SRMR values below 0.08 indicate a good fit.

We have added to additional measures of fit, namely the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), which are both closely related. BIC is defined as

$$\text{BIC} = k \ln(n) - 2\ln(\hat{L})$$

where k is the number of estimated parameters in the model, n is the number of observations and $\hat{L}$ is the maximum likelihood estimate. This definition is closely related to the AIC, which is defined as:

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

While it is not possible to evaluate whether a model fit is 'good' or 'bad' by looking at the AIC and BIC, it is useful in model selection, since both measures penalise model complexity.

2.7 Model results

All models are estimated using the 'lavaan' package in R (Rosseel 2012) using standard maximum likelihood estimation with robust standard errors and Satorra-Bentler scaled test statistics to correct for potential non-normality in the distribution of the items (Satorra and Bentler 1994). In Table 2.2, the different models are reported alongside the global goodness-of-fit test statistics.

The first two entries in Table 2.2 compares the one-factor model and the proposed two-factor model. According to all fit indices, the one-factor model is inferior to the proposed two-factor model. However, when investigating the modification indices (MI) of the two-factor model, we find that the error terms of trust in politicians and political parties are highly correlated. Introducing the correlated errors increases model fit substantially as shown in Table 2.2. Substantially speaking, the correlated errors are meaningful: politicians are (most often) nested within political parties,

Model	**Chi-sq.**	**d.f.**	**CFI**	**ΔCFI**	**RMSEA**	**ΔRMSEA**	**SRMR**	**ΔSRMR**	**AIC**	**BIC**
Single-factor	-	5	0.919	-	0.225	-	0.075	-	635622.818	635748.397
Two-factor	778.127	4	0.992	0.073	0.078	-0.147	0.023	-0.052	623928.288	624062.24
Two-factor (corr. errors)	218.868	3	0.998	0.006	0.047	-0.031	0.007	-0.016	623065.709	623208.033
MGCFA (corr. errors)										
Configural	362.566	54	0.997	-	0.057	-	0.009	-	608041.205	610603.034
Metric	766.880	105	0.993	-0.004	0.06	0.003	0.029	0.020	608403.941	610538.799
Scalar	3056.886	156	0.967	-0.026	0.102	0.042	0.047	0.018	610907.33	612615.216
Partial scalar	1578.825	147	0.984	-0.009	0.074	0.014	0.037	0.008	609231.885	611015.118

Table 2.2: Goodness-of-fit indices for measurement invariance models.

and it thus seems reasonable to assume that any kind of response bias to one question will be reflected in the other. As a result, we adopt the slightly modified two-factor model with correlated errors in the subsequent models.

When examining the MGCFA models, both the configural invariance model, the metric invariance model and the full scalar invariance models are compared. The configural invariance model provides a good fit, although the RMSEA is slightly above the suggested cut-off point found in the literature for a well-fitting model. The metric invariance model also shows an acceptable fit to the data with only a slight decrease in CFI (ΔCFI $= 0.004$), slightly lower BIC and slightly higher AIC values. However, the full scalar invariance model results in a severe decrease in all model fit statistics.

Again, we look at the MI for potentially freeing up the intercepts of countries where one or more items appear to be non-invariant. We relaxed one intercept at a time until the reduction in fit between the metric and resulting scalar invariance model accoring to CFI was less than 0.01 (Chen 2007). This was achieved by relaxing the constraints with a MI > 50. The resulting changes for the item incepts are shown in Table 2.3.

We find that trust in the EP and UN is non-invariant across 7 countries, with Finland, United Kingdom, Netherlands and Sweden having a lower intercept for trust in the EP/higher intercept for trust in the UN compared to the rest of the countries. At the same time, Italy, Lithuania and Slovenia show the opposite pattern. With the Lithuania and Netherlands being invariant on only one item, it is necessary to exercise extreme caution when estimating the latent means.

The first partial scalar invariance model provides a better fit compared to the scalar invariance model. Compared to the metric invariance model, the ΔCFI $= -0.009$ is slightly below the suggested maximum reduction in CFI as suggested by Chen

Country/intercept	Parliament	Politicians	Parties	EP	UN
Finland				-1.12	1.12
United Kingdom				-1.22	1.22
Italy				-0.59	0.59
Lithuania	0.273	-0.04		-0.56	0.56
Netherlands	0.312		0.037	-0.98	0.98
Sweden				-1.18	1.18
Slovenia				-0.57	0.57
Invariant countries	**0.58**	**-0.35**	**-0.23**	**-0.76**	**0.76**

Table 2.3: Item intercepts for non-invariant items and countries.

(2007). The RMSEA, like the metric invariance model, is above the suggested cut-off point, although still within the boundaries of a 'mediocre fit' while the SRMR indicates a good fit. Consequently, the first partial scalar invariance model serves as the final model from where the estimated latent means and factor correlations can be extracted as shown in Table 2.4. Since some countries are non-invariant on one or more items, we have included the estimated latent means of the scalar and metric invariance models as well.

The differences in the estimated latent means between the partial scalar, metric and scalar invariance models are negligible for most countries. Compared to the scalar invariance model, the biggest differences (> 0.1) are observed on SPT for France, Spain, Portugal, Italy and Lithuania, of which only Italy and Lithuania were non-invariant. For NPT, the differences are 0 or close to 0 for all countries across all models.

Generally, the results show a very strong correlation (> 0.8) between NPT and SPT in most of the countries with the Benelux countries (Netherlands and Belgium) having a correlation above 0.9. Among the high correlation countries, Netherlands,

Country	**Partial scalar**			**ΔMetric**		**ΔScalar**	
	ρ_{12}	$\overline{NPT}$	$\overline{SPT}$	$\overline{NPT}$	$\overline{SPT}$	$\overline{NPT}$	$\overline{SPT}$
Netherlands***	0.93	5.24	5.14	0.00	0.05	0.00	-0.01
Belgium	0.91	4.30	4.95	-0.02	0.00	-0.01	-0.01
Austria	0.89	4.26	4.17	0.01	0.00	0.01	0.05
France	0.87	3.26	4.17	-0.03	0.00	0.01	0.14
Germany	0.87	4.48	4.65	0.01	-0.09	0.00	-0.01
Finland*	0.86	5.12	5.79	0.00	0.00	0.00	0.09
United Kingdom*	0.85	4.06	4.41	0.03	0.09	0.00	-0.01
Sweden*	0.84	5.18	5.46	0.00	0.05	0.00	-0.01
Spain	0.84	2.97	4.56	0.00	0.00	0.00	0.17
Czech Republic	0.83	3.84	4.52	0.02	0.06	0.00	-0.01
Portugal	0.80	2.99	4.58	-0.03	0.00	0.05	-0.22
Italy*	0.78	2.62	4.21	0.00	0.00	0.00	-0.13
Slovenia*	0.77	2.77	4.14	0.01	-0.06	0.01	-0.01
Estonia	0.76	3.94	4.88	-0.01	0.04	0.00	-0.01
Ireland	0.74	4.01	5.31	-0.04	-0.17	0.00	-0.01
Lithuania**	0.61	3.49	5.51	0.00	0.00	0.00	-0.10
Hungary	0.54	3.97	4.76	-0.02	-0.02	0.00	-0.01
Poland	0.41	2.84	4.35	-0.03	-0.03	-0.01	-0.01

Table 2.4: Correlation between NPT and SPT and estimated latent mean for the partial scalar invariance model and the differences in latent means between the metric and scalar invariance models.
**) Intercept for trust in the UN/EP non-invariant.*
***) All intercepts, except trust in political parties non-invariant.*
****) All intercepts, except trust in politicians non-invariant.*

Sweden, Finland, Germany and (to some degree) Belgium and Austria exercises the highest latent means of NPT while Finland, Lithuania, Sweden, Ireland, Netherlands and Belgium have the highest level of SPT. Countries with a weak correlation (< 0.7) include only Lithuania, Hungary and Poland, which are all characterised by having a high difference in mean levels of SPT and NPT. The results are further illustrated in Figure 2.2.

It looks as if the Nordic countries and the Netherlands fall into a category of themselves as having a consistently high degree of belief in both national and supranational political institutions. On the flipside, we observe the Southern European countries, Poland, Slovenia and France exercising lower-than-average levels of both NPT and SPT. Lithuania is an interesting case since it is the only country with below-average levels of NPT but very high levels of SPT.

One thing which is important to note is that country-level aggregates do not resemble individual levels of NPT/SPT, and as such the within-country variability could be notably different in countries with the same mean levels of NPT/SPT. Finally, the mean level of NPT (from 2.62 in Italy to 5.24 in the Netherlands) varies considerably more than SPT (from 4.14 in Slovenia to 5.79 in Finland).

2.8 Invariance and correlation between national and supranational political trust

The purpose of this chapter is to investigate if national and supranational political trust are empirically distinguishable concepts, and if we can reliably compare the latent variable means across countries, and to what extent levels of both types of trust are associated in different EU countries.

In this regard we arrive at three main conclusions: 1) national and supranational

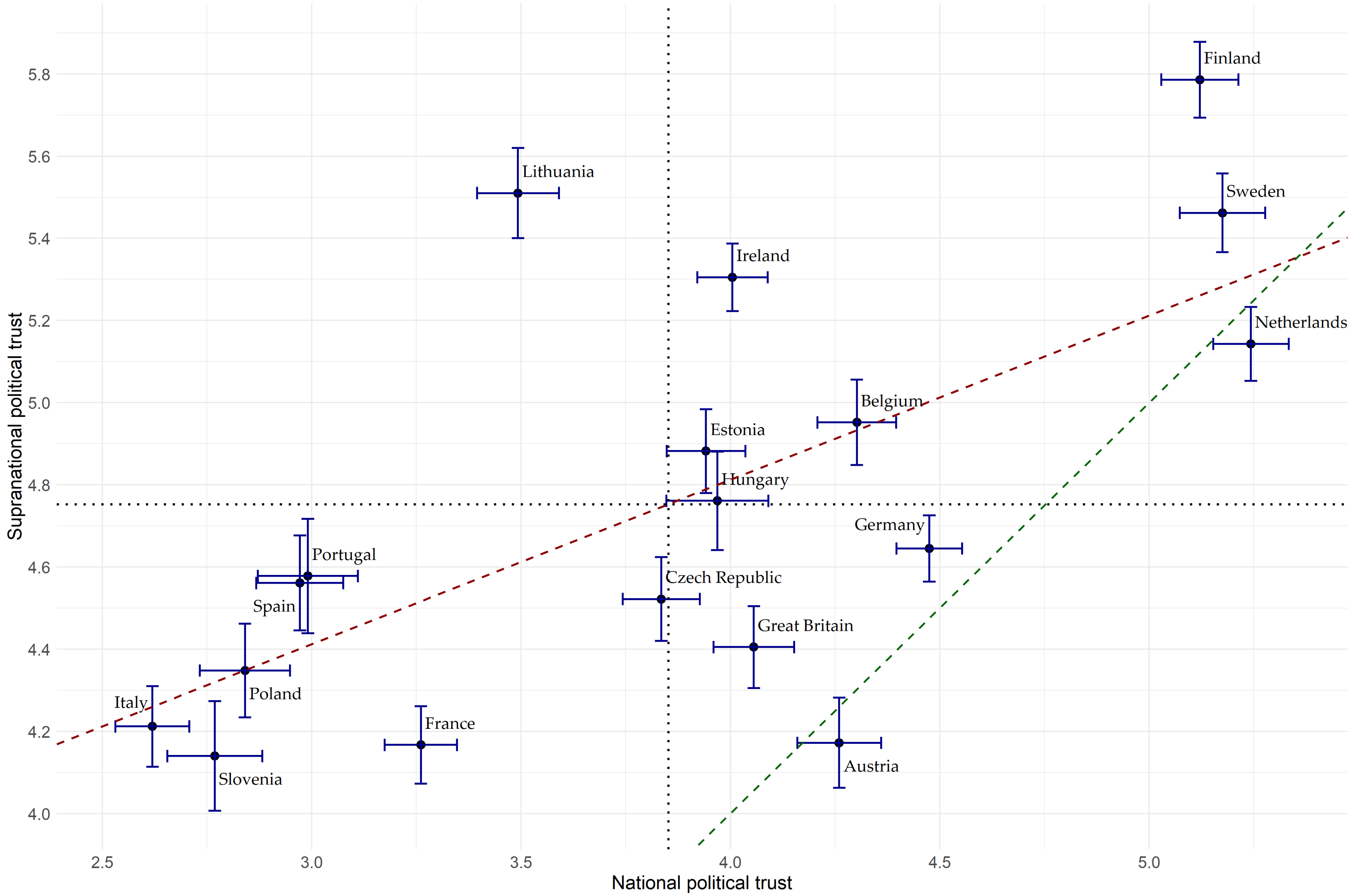


Figure 2.2: Estimated mean levels of national and supranational political trust in 18 EU countries.
NPT/SPT is measured on a 0-10 scale.
Blue error bars indicate the 95% confidence interval. Black dottted lines show the mean levels of NPT/SPT in the sample.
Red dashed line shows the best fitting OLS regression (SPT ∼ NPT).
Green dashed line indicates values for which NPT = SPT.

political trust are two empirically distinct concepts, which should be split into two latent variables, 2) NPT and SPT is measurement invariant in most countries with the latent means in non-invariant countries not affecting the substantive conclusions and 3) both NPT and SPT correlate strongly in the majority of the countries. We will discuss points 2) and 3) further below.

The modelling procedure clearly shows that full measurement invariance across all countries is not achievable. It is necessary to free up the intercept for trust in the EP/UN in multiple countries to obtain an acceptable fit. Freeing up the intercept for trust in the EP has the same effect on model fit since the measure for SPT is made from only two indicators (the intercepts must sum to 0).

Nonetheless, both the metric and the final partial scalar invariance model yield similar estimates of latent means, which in turn leads to similar substantive conclusions. Comparable results are found by Reeskens and Hooghe (2008), who uses MGCFA on two older waves on the ESS, finding it impossible to obtain full scalar invariance across all countries when measuring generalised social trust.

On the interrelationship between national and supranational political trust, the results show a strong correlation across most countries. Pursuing the evidence that NPT and SPT are empirically distinct, the correlation table shows how high levels of trust in national political institutions is strongly related with higher levels of trust in supranational institutions, such as the EU. This finding goes in line with that of Muñoz, Torcal, and Bonet (2011), who uses multilevel modelling of older waves of the ESS to claim that trust in the national parliament is positively correlated with trust in the European Parliament and Arnold, Sapir, and Zapryanova (2012) who '[...] confirmed the congruence hypothesis developed in the literature by showing that trust in domestic institutions fosters trust in the institutions of the European Union.' Interestingly, countries with low correlations between both forms of political trust are characterised by notably higher levels of SPT compared to NPT, and not the other way around (only in Austria and Netherlands is NPT slightly higher than SPT).

In turn, the Nordic countries and the Benelux countries have consistently high levels of political trust across the board, just as expected from the literature. With the exception of these countries, no clear dichotomy based on the stable versus new democracies seems to exist. Although earlier studies have shown how post-communist countries have an initial low level of political trust, this seems to be true only in a few cases for the 'new democracies', since both Hungary, Czech Republic and Estonia exhibit average or above-average levels of national political trust.

While it is not within the scope of this thesis to offer an explanation as to why this difference may or may not exists, the findings indicate that the NPT-SPT dynamic is quite different across countries. However, we find it relevant that the Southern European countries (Italy, Spain and Portugal) all have low levels of both NPT and SPT. This may indicate that a national- or cultural level bias exists in those specific countries. Summing up, we believe it imperative to evaluate political trust at both the national and the supranational level.

2.9 Extentions to the simple measurement model

Since the purpose of constructing a simple measurement model as in this chapter is to investigate the apparent correlation between two empirical concepts for between-country comparisons, no inference is made on NPT having a direct effect on STP and vice versa. Even though political theoretical models may imply such a relationship, corresponding assumptions cannot be made for the data used in this study. It would require not only longitudinal data but also an extensive understanding of historical events, which are known to affect NPT and/or SPT. Furthermore, a range of covariates are known to have an impact on political trust, let it be national or supranational, and including those in a structural equation modelling framework seems like the next logical step.

In the theoretically oriented literature on political trust, Kestilä-Kekkonen and Söderlund (2015) argue that engagement in the political system and the level of trust in the said system is non-constant across different age groups. In addition, Foster and Frieden (2017) argue that education is a strong indicator of political trust, with higher educated citizens having significantly higher levels of trust in the political system and hence exercising a higher level of political trust, national as well as supranational. They also claim that females are less trusting towards both national and European governments (Foster and Frieden 2017). The ESS offers a wealth of individual-level covariates, such as gender, age, and educational level, and there is plenty of scope for investigating and trying to explain the variations in trust that we have established here. However, while we have investigated variation across countries, this level of aggregation can be argued to be too simplistic and within-country variation may well reflect variation across further subdivisions, as we saw in the motivating example, as well as individual-level differences. Parsing out the variation across regions is the topic of the next chapter.

Chapter 3

Measuring the association between NPT and SPT across NUTS regions in the EU

In Chapter 2 we discussed and analysed a simple measurement model of political trust across countries using the ESS. One of the key discoveries was that NPT and SPT are strongly correlated within countries. However, as we argued in the motivating example, is it unreasonable to assume that countries are homogenous across regions, counties or other subdivisions? Since the ESS contains information on Nomenclature of Territorial Units for Statistics (NUTS) in the European Union (EU), we will be able to set up a NUTS-level model. By analysing a measurement model for political trust in smaller geographical areas rather than the country level, one can get a better and more fine-grained understanding of the multidimensionality of political trust and its spatial variability.

The purpose of this chapter is to outline the issues when moving from a country level model to a regional level model. Hence, it is important to note that the steps taken are not with the intention of following a rigorous modelling procedure, but rather to provide the basis for a Bayesian modelling framework put forward in Chapter 4.

Analysing the ESS on the NUTS-level is not new and has seen a number of applications in recent years (e.g. Algan et al. (2017), Iacono (2019), Schoene (2016), and Rustenbach (2010)). Iacono (2019) uses NUTS 2 on crime data, and is a strong proponent of using regional-level data. Of the many advantages, higher statistical power and capturing heterogeneity within countries are highlighted (Iacono 2019). Schoene (2016) and Rustenbach (2010) uses NUTS 2 in a similar way through different variants of multilevel models. However, the main reasoning for the application is to make use of regional variables such as population density and GDP. While demographic information and NUTS-level aggregates are important in certain modelling frameworks, these studies do not provide any theoretical justification for why an analysis at the sub-country is both relevant and necessary.

One recent study that touches upon regional variation within a theoretical framework is Bäck et al. (2018). They claim, in line with Rustenbach (2010), that regions with high levels of generalised trust "constitutes more hospitable environments of receptions for immigrants rather than regions where the overall level of trust is low" (Bäck et al. 2018). This is not only novel from a theoretical perspective but sets the stage for a more detailed analysis of the potential between-regional mechanics in play. In that sense, we agree with Ziller (2014) in a study on the longitudinal impact of immigration-related diversity:

> "Previous pan-European studies have often relied on country-level indicators, which is a quite far-removed perspective. Using regions instead will allow the modelling of important variations in ethnic con-

text, while still generating comparable results for a broader European context." (Ziller 2014)

In light of the above, the purpose of this chapter is to extend on the idea that theoretical concepts such as political trust should be investigated at the NUTS level. However, as pointed out in Chapter 2, it is necessary to establish a measurement model to ensure that we are measuring the same concept across all regions; something which the aforementioned studies do not consider. Hence, we provide one possible way of analysing the association between national- (NPT) and supranational political trust (SPT) in the European Social Survey (ESS) from 2010-2016 across NUTS regions. We extend on Chapter 2 with an example to highlight 1) considerations and issues regarding coding of NUTS over time, 2) motivation for examining measurement models at the NUTS-level, 3) results of between-NUTS and between-yearly differences of SPT and 4) a discussion of drawbacks of the chosen model, including suggestions on how they can be resolved.

3.1 NUTS coding over time

The data used in the following analysis is ESS round 5 through 8 (2010-2016). The total sample sizes across countries for each year are listed in Table 3.1. For this particular analysis, we are only interested in EU countries who are participating in all ESS rounds, hence the slight reduction in number of countries compared to analysis in Chapter 2.

When analysing data at the NUTS level, the aim is to preserve the highest level of detail possible (i.e. NUTS 3 > NUTS 2 > NUTS 1), while at the same time ensuring

Country	ESS round				Total
	5	**6**	**7**	**8**	
Belgium	1,639	1,814	1,713	1,710	6,876
Czech Republic	2,173	1,831	1,996	2,120	8,120
Germany	2,653	2,677	2,845	2,678	10,853
Estonia	1,433	1,956	1,819	1,786	6,994
Spain	1,666	1,657	1,593	1,633	6,549
Finland	1,781	2,091	1,971	1,857	7,700
France	1,650	1,876	1,793	1,977	7,296
United Kingdom	1,899	1,803	1,997	1,773	7,472
Hungary	1,299	1,720	1,474	1,409	5,902
Ireland	2,033	2,173	1,975	2,274	8,455
Lithuania	1,218	1,665	1,886	1,862	6,631
Netherlands	1,689	1,698	1,790	1,529	6,706
Poland	1,429	1,515	1,358	1,425	5,727
Portugal	1,714	1,871	1,127	1,145	5,857
Sweden	1,231	1,666	1,581	1,368	5,846
Slovenia	1,179	1,046	1,028	1,186	4,439
Total	**26,686**	**29,059**	**27,946**	**27,732**	**111,423**

Table 3.1: Country sample sizes in ESS 5, 6, 7 and 8. Countries in the analysis highlighted with bold.

consistent coding over time. Prior to round 5, the ESS did not incorporate official NUTS coding for countries but instead provided country-specific region variables, which in some cases did not follow official NUTS classification. Hence, we have chosen data from ESS 5 and forward, excluding ESS 9, which at the present time does not include data from a sufficiently large number of countries. Furthermore, several countries have adopted different coding over time under the NUTS 2010, 2013 and 2016 classifications, with a new classification expected to be released in 2020. In cases where NUTS classification has changed over time, previous classifications have been re-coded to follow the current standard where possible. Only two NUTS regions were not possible to recode: PL91 (Warszawski stołeczny) and PL92 (Mazowiecki regionalny), due to boundary changes. In this particular case, they have been merged into the NUTS1-region of PL9 (Makroregion Województwo Mazowiecki). In order to carry out the following analysis, recodings have been made for Finland, Slovenia, Ireland, France, Lithuania, Hungary and Poland, as listed in Appendix 2. One of the major drawbacks of ensuring consistency over time when analysing NUTS regions is the loss of depth when re-coding to a higher NUTS level is necessary. Therefore, if one finds only a specific year interesting for further analysis, the highest possible NUTS level should be applied.

In the analysis, only NUTS regions which have data available for all ESS rounds of interest and have more than 40 observations in every round are included. This is due to issues with the multiple groups confirmatory factor analysis method, which will be further discussed in Section 3.4 and the following chapter. The reduced number of NUTS regions, including the number of observations per NUTS is provided in Appendix 3.

The total number of NUTS regions in the sub-sample is 163 where the largest samples are found in the Norte region in Portugal in ESS 5 (PT11 - 726 observations),

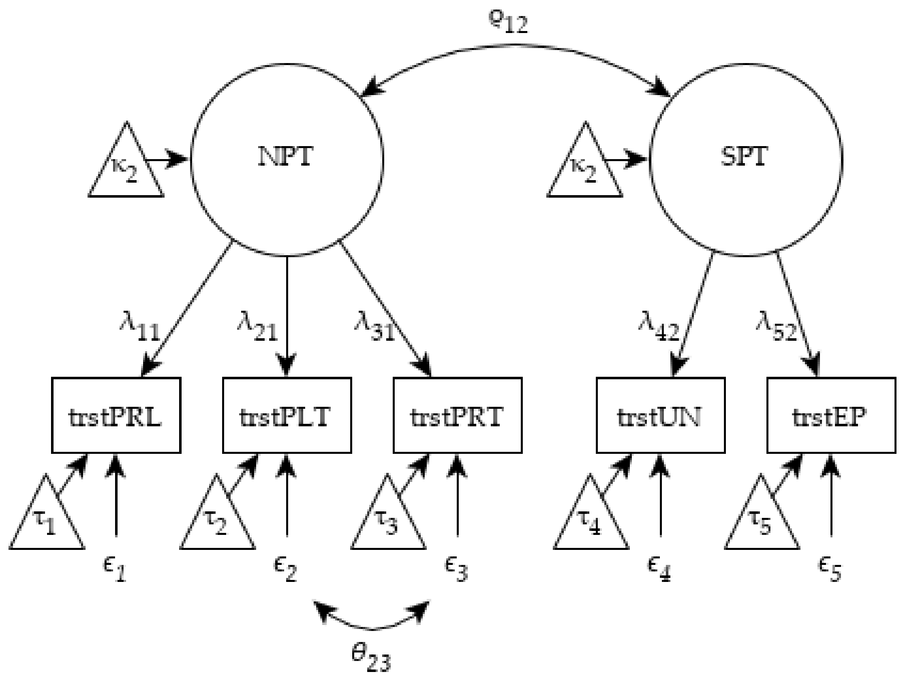


Figure 3.1: Measurement model of NPT and SPT with correlated errors.

Lisbon metropolitan area in Portugal in ESS 6 (PT17 - 746 observations), and Harju County in Estonia in ESS 7 and ESS 8 (EE001 - 823 and 776 observations respectively). The smallest region across all four ESS rounds is in Dalarna County in Sweden (SE312), which has a total of 176 observations. The mean sum of observations per NUTS is 628 across all four ESS rounds (157 per round). This mean value would, of course, be smaller if the full sample was used.

3.2 Model setup and reasoning

The model of interest is shown in Figure 3.1, which corresponds to the MGCFA models in Chapter 2.

The first three items act as indicators for national political trust (NPT) and the last two as indicators for supranational political trust (SPT). As shown in Chapter

2, this model fits well to the ESS data and is superior to the one-factor counterpart. Previous analysis has mainly focused on between-country differences in the correlation between NPT and SPT and not investigating these differences at the NUTS level. The reasoning for having a more in-depth look at this measurement model is two-fold. First, both theoretical arguments and empirical work suggest that political behaviour and opinion is not independent of regional affiliation within a country, which was also pointed out in the introduction to this chapter. As shown in the motivating example from the introduction, a large majority of citizens in London and Scotland areas voted to remain, indicating some regional differences in voting patterns (see also Hobolt 2016; Colatone and Stanig 2018; Clarke, Goodwin, and Whiteley 2017). Second, since this data is available across years, examining NUTS as the analytic unit of interest makes it possible to broaden our understanding of with-country changes over time. It might be that estimated country-specific levels of NPT and SPT respectively are estimated to be constant over time, even though said country experiences significant regional changes. As such, analysis involving measures such as NPT and SPT is best done at the lowest possible geographical level.

One thing to keep in mind is the relative difference in detail for each country. While some countries are divided into many NUTS regions with few observations in each, other countries have only a few NUTS regions. The difference in sample sizes is not an issue when the aim is to compare different regions, but as it is the case with all comparative research, one has to be cautious when examining aggregates.

3.3 NPT and SPT across a subset of NUTS regions in 2010-2016

Models in this section were run using the R-package 'lavaan' in Microsoft R Open 3.5.1 (Windows). One of the limitations when moving to the NUTS level, where the number of groups increases significantly, is the computational part of the modelling procedure, which requires more physical memory than normally available on standard personal computers. We ran each model separately. Each of the year-specific models was able to run with 32 GB of RAM with an additional 25 GB of virtual memory pagefile. The computational issues are less severe in Linux, which will allocate RAM automatically to the disk in case of overflow in R under the right setup, although this may slow down the estimation notably.

In Figure 3.2 to 3.5, the estimated mean level of NPT and SPT across all four rounds are plotted onto maps using freely available software (QGIS and GeoDa). The legends of NPT and SPT are different since the estimated means of SPT are consistently higher in most NUTS regions. Many countries lack several NUTS regions, due to the restrictions imposed by subsetting the data, which reminds us that this is an illustrative example rather than a rigorous modelling procedure. All models were run as metric invariance models (i.e. $\Lambda^g = \Lambda \forall g \in G$). It was not possible to compute a scalar invariance model since the maximum likelihood estimator is not able to converge.

When eyeballing the trends on the maps, it does not appear that NPT and SPT are consistent, neither between countries nor within countries over time. While the few Swedish regions, Finland regions and the Benelux countries' regions exhibit consistently high estimated levels of NPT and SPT, other countries indicate a high level of variability in these measures. One example is Lithuania and Hungary,

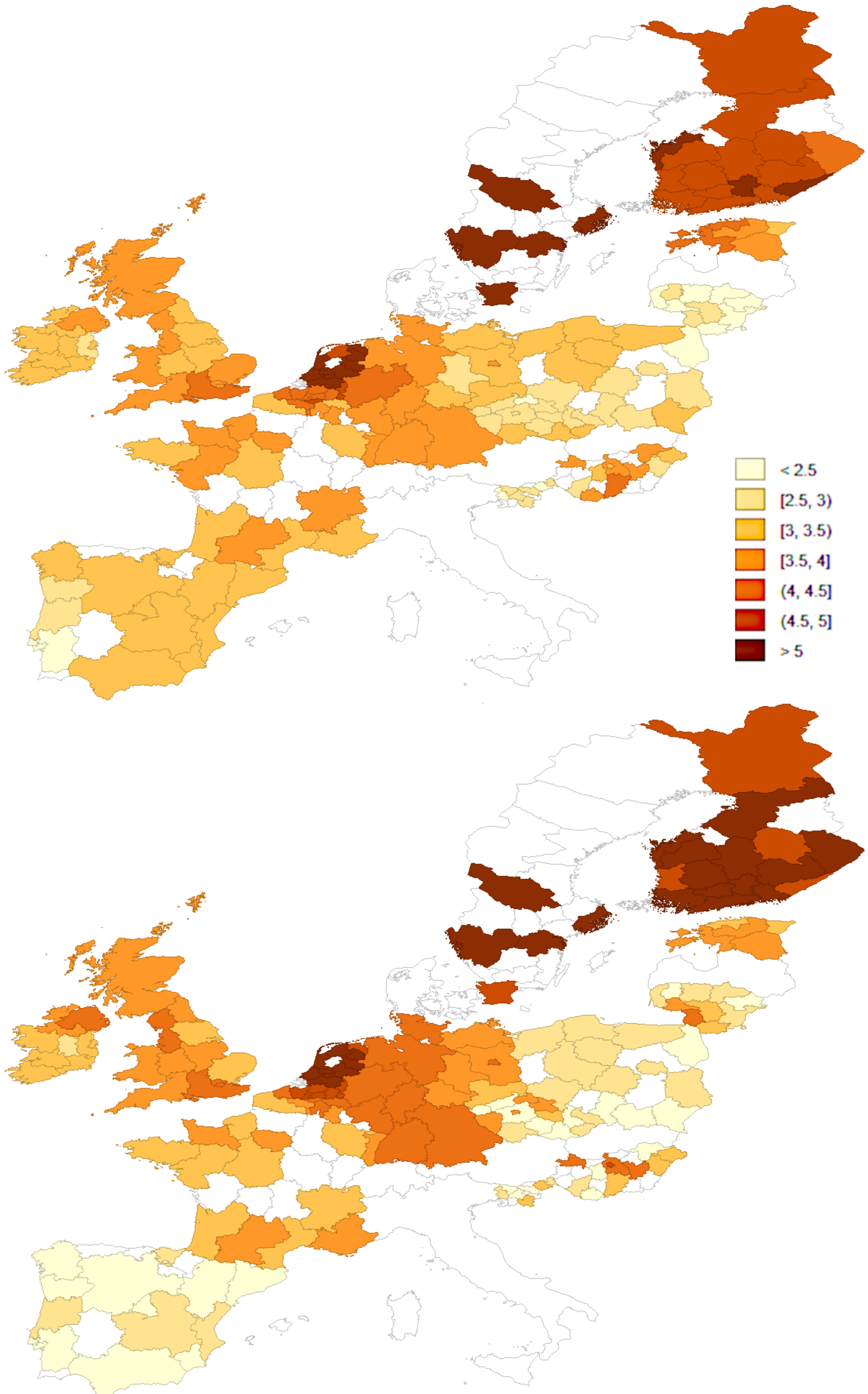


Figure 3.2: Estimated mean level of NPT across NUTS regions in ESS 5 (top) and ESS 6 (bottom).

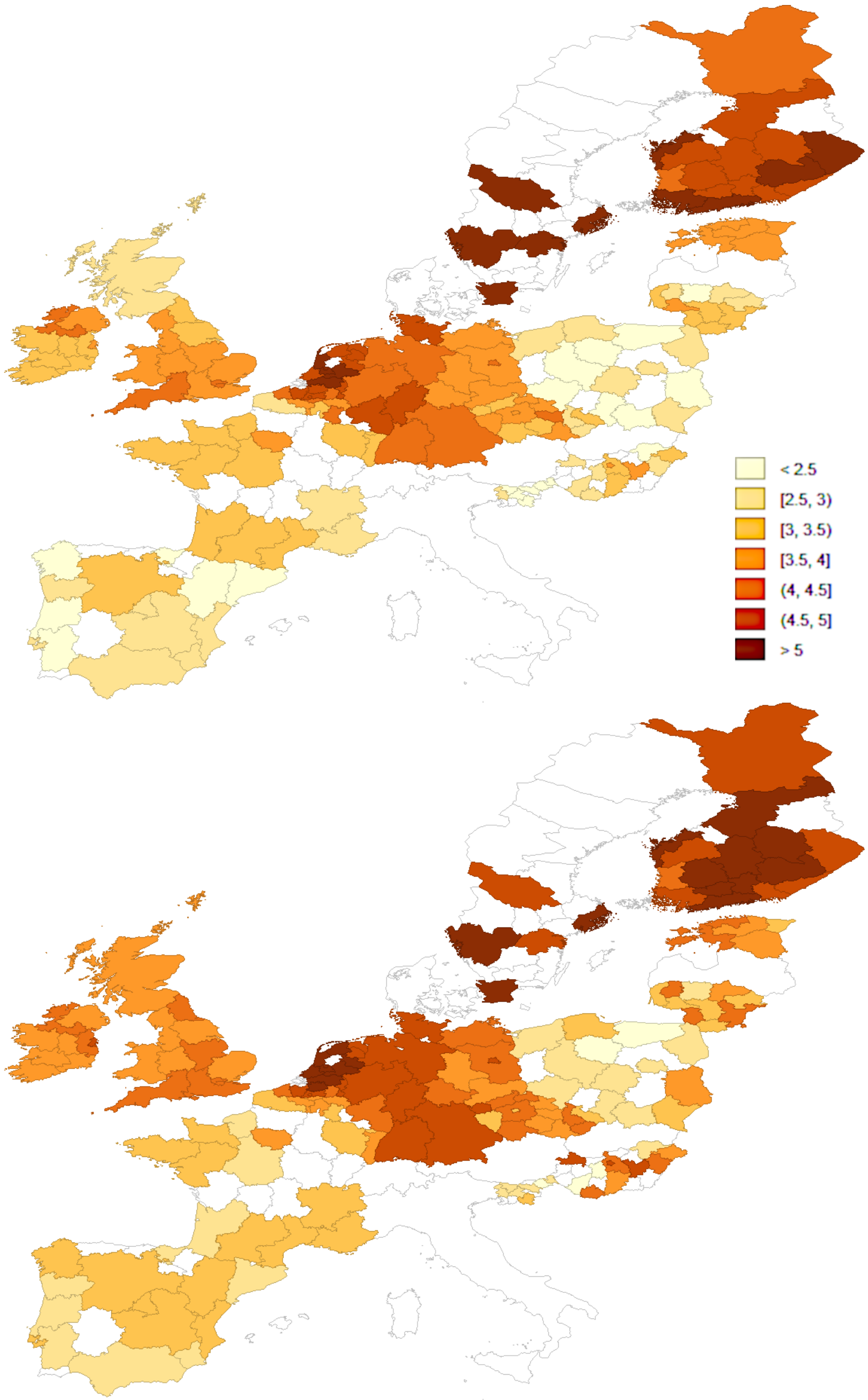


Figure 3.3: Estimated mean level of NPT across NUTS regions in ESS 7 (top) and ESS 8 (bottom).

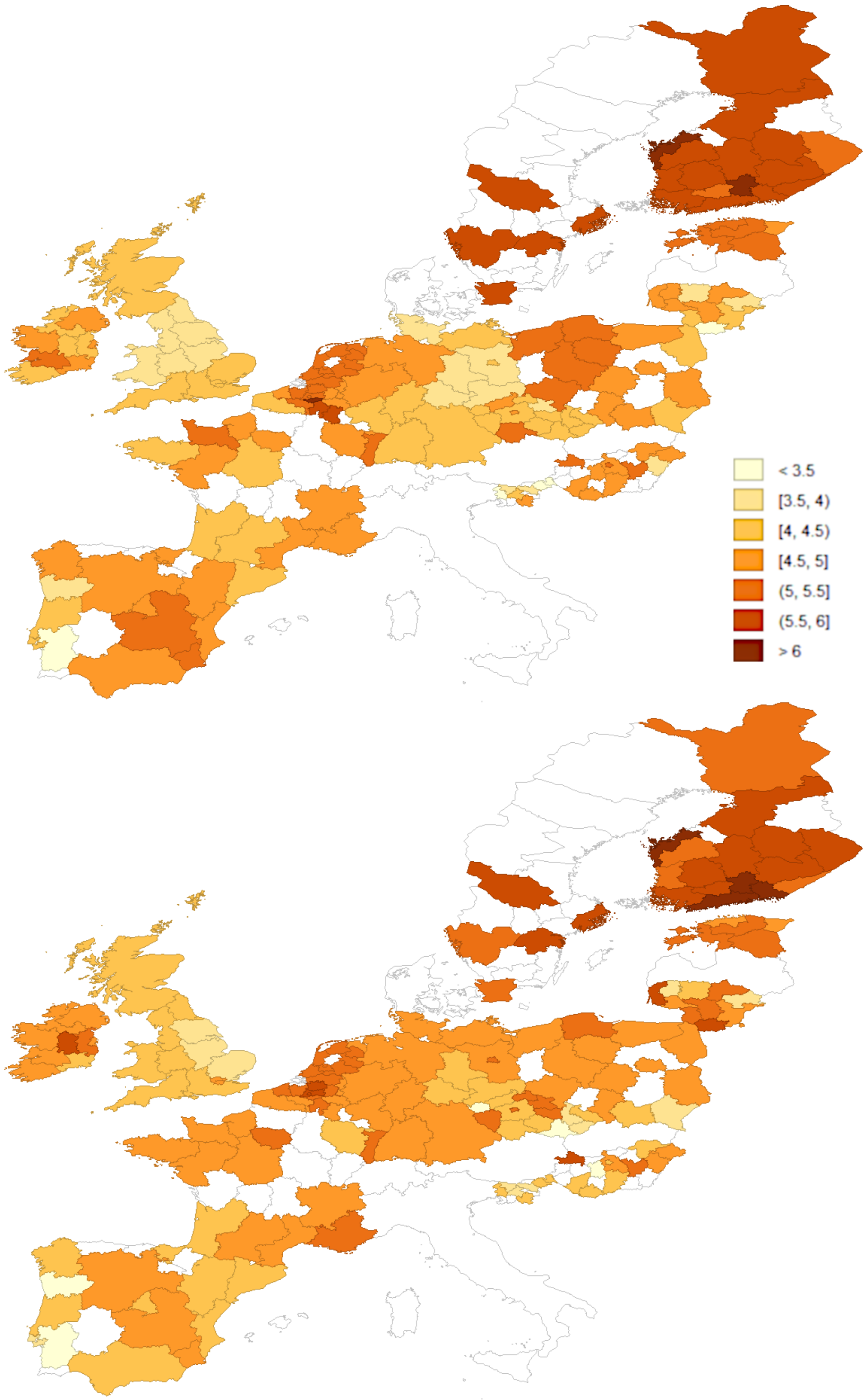


Figure 3.4: Estimated level of SPT across NUTS regions in ESS 5 (top) and ESS 6 (bottom).

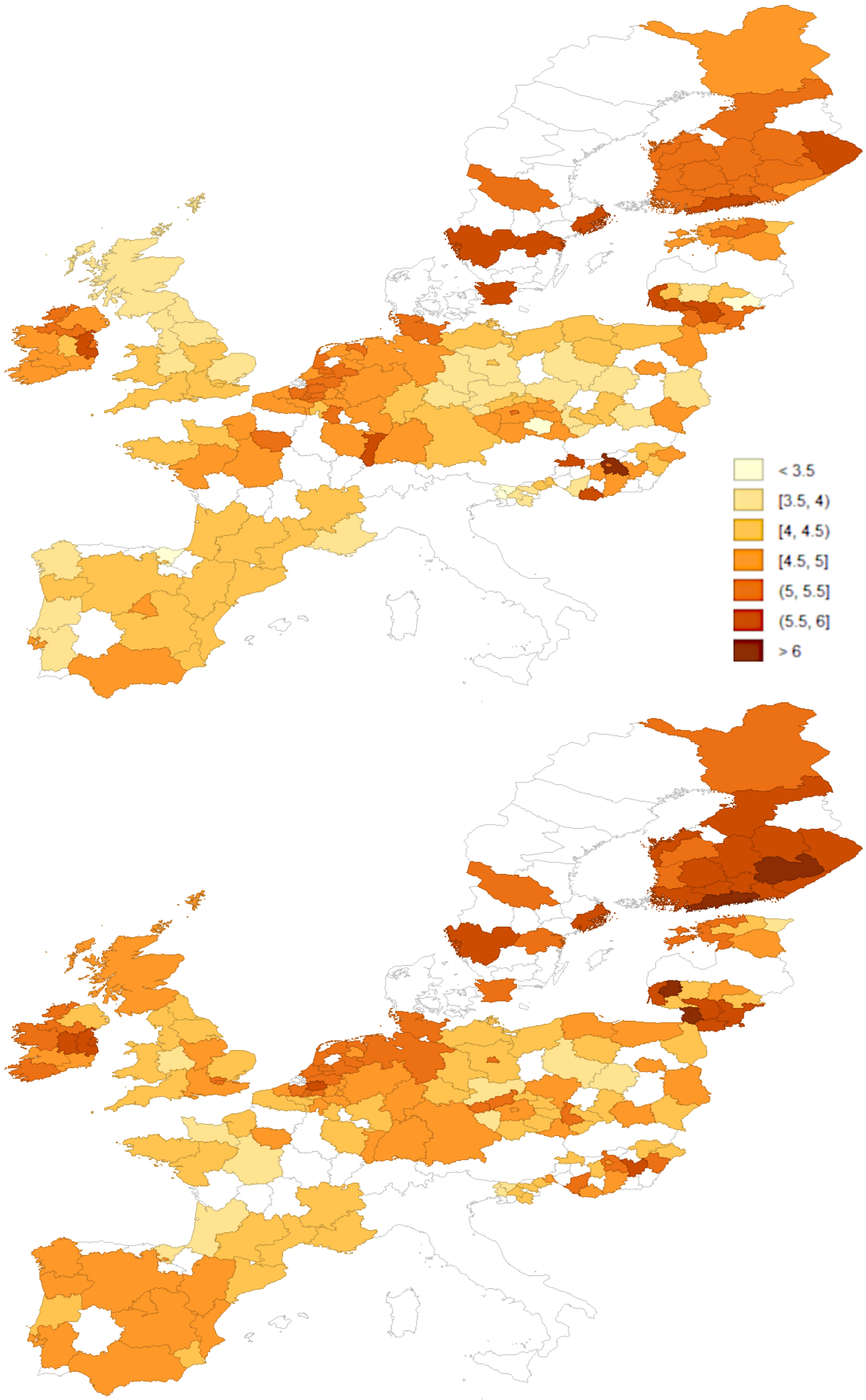


Figure 3.5: Estimated mean level of SPT across NUTS regions in ESS 7 (top) and ESS 8 (bottom).

where SPT varies across NUTS regions for all ESS rounds, albeit less for NPT. Finally, we observe a 'capital effect' on both measures of trust, with the capital region of France, Ireland, Germany, Finland and Czech Republic having a higher level of NPT than the surrounding regions. Such an effect have been observed in previous studies as well (e.g. Rahn and Rudolph (2005)). While it is not the purpose of this study to investigate the potential substantive reasons for such an effect, it is worthwhile to keep in mind for future analyses.

Again, it is necesary to highlight that the aim of this excercise is not to provide a meaningful modelling framework, but rather to highlight apparent issues when moving to a regional level model. One possibility would be to aggregate small-N regions, i.e. cluster regions together, to avoid estimation issues. However, we have chosen to stick with the most detailed NUTS level division possible.

3.4 Issues and solutions

Several issues present themselves when running a multiple groups confirmatory factor analysis (MGCFA) model like the one presented above, which shall be discussed in turn.

First, the conventional maximum likelihood estimator is not able to estimate parameters if the number of items is equal to or greater than the number of observations, which makes more complex models increasingly more difficult to estimate. Hence, when such NUTS regions appear in the data they must simply be removed in order for the maximum likelihood estimation procedure to be successful. In this specific analysis, it was even necessary to reduce the sample to only consist of NUTS regions with 40 or more observations to avoid complications in the estimation procedure. Second, so-called Heywood cases are a common appearance with such a vast number of groups (Gagne and Hancock 2006). A Heywood case appears when the residual variance of a given item is estimated to be negative. Previous research has developed methods for diagnosing the reason for their appearance, while some choose to simply fix the variance to be 0. If not handled properly, the implication can be wrong parameter estimation and thus wrong substantial conclusions. A full list of cases in the present analysis (ESS8 only) is presented in Table 3.2, which of course does not include regions which did not enter the analysis due to extremely small sample sizes (N < 40). This indicates that Heywood cases, while prevalent in small-N groups, can also appear in larger N groups. Heywood cases will be an important issue for discussion in Chapter 4.

Third, in some cases, the model implied co-variance matrix for a given NUTS is not positive definite. What that means is that some values exceed their natural range (correlation between latent variables is above 1 or below -1). While the correlation between the latent variables has not been investigated further in this analysis, literature suggests that it can be due to either model misspecification or estimated parameters which are very close to the parameter boundary.

Solving these issues are not trivial within the frequentist modelling framework, and hence we suggest that it is preferred to analyse MGCFA models on a large number of groups within a Bayesian framework, an approach which is accomplished in Chapter 4. As Gelman (2006) pointed out, specifying priors for variances is a non-trivial problem, but following his modifications for the standard prior specification, it is possible even with a very small number of observations. At the same time, Heywood cases can be eliminated by choosing a suitable prior distribution for the variance parameter and the correlation parameter.

3.4.1 Sensitivity to extreme cases

A key premise for the validity of the fit indices and the inference is general, are

Region	Region SS	Item		Total
		trstEP	trstPRT	
CZ010	267	1	0	1
CZ041	57	1	0	1
CZ072	100	1	0	1
DE4	154	0	1	1
EE007	169	0	1	1
ES21	71	1	0	1
ES24	45	1	0	1
ES41	94	1	0	1
ES42	61	1	0	1
ES52	172	1	0	1
FRG0	122	1	0	1
HU110	275	1	0	1
HU211	64	1	1	2
HU221	57	1	1	2
HU232	42	0	1	1
HU321	46	1	1	2
LT011	466	1	0	1
LT021	82	1	0	1
LT024	101	1	0	1
LT027	79	1	0	1
NL13	51	0	1	1
PL42	56	0	1	1
PL51	113	0	1	1
PL81	92	0	1	1
SE123	71	0	1	1
SE312	50	0	1	1
Total		**17**	**12**	**29**

Table 3.2: Heywood cases at the NUTS level in a subset of ESS 8.

the distributional assumptions of the model. In particular, the sphericity and continuous level of measurement that stems from the normally distributed error terms. A number of Monte Carlo studies have investigated the effects of departures from normality on estimators and tests of model fit (Muthén and Kaplan 1985; Curran, West, and Finch 1996; Lei and Lomax 2005) as well as empirical evaluation of standard methods to counter non-normality (e.g. Andreassen, Lorentzen, and Olsson 2006). Typically these studies have operationalised departure from normality in terms of skewness and kurtosis, but for the trust items in ESS there seems to exists a very specific censoring issue with flooring and ceiling effects on extreme response values. These extreme values will affect skewness, much like outliers as identified by Gao, Mokhtarian, and Johnston (2008), it is uncertain whether transformations or alternative estimation procedures would alleviate the problem (alternative estimation procedures would also not address departure from sphericity for the fit measures). For similar reasons, it is not clear whether an alternative model, like the model for ordered categorical outcomes employed by Lubke and Muthén (2004) would be appropriate.

Due to the nature of the scale provided by the ESS, which goes from 0 to 10, some items have a very high percentage of 'no trust at all' answers. For example, 10.61% of the respondents have 'no trust at all' in the European Parliament, while only 1.04% have 'complete trust'. These numbers are 15.2% and 0.42% for trust in politicians and 14.41% and 0.38% for trust in political parties respectively. However, it is only 6.92% and 1.98% for trust in the UN.

While the items that make up the trust scale are treated as continuous, it is not clear that different categories of people have the same reference points. For grouped or coarsened data, floor effects is a known statistical problem (e.g. Heitjan (1989)). A floor effect comes into play when data gets limited by the lower bound of the scale. For example, if two respondents answer '0 = no trust at all', which is the floor of the scale, they may still exhibit different levels of trust. While we do not go into the deeper psychological motivations, in attitudinal surveys, if people feel really negative to a topic, they might revert to extreme response styles (Hyman and Sierra 2012; Lavrakas 2008; Lu and Bolt 2015). The attitude to national and supranational political trust is clearly a polarising issue and we cannot discount the possibility that while some individuals rate their trust on the required scale, other people might feel very strongly that there is no value on the scale that is negative enough. The measurement model relies crucially on sphericity and floor effects may adversely affect the inference.

In other words, since the scale is limited to the range 0-10, it is assumed that some respondents answering 0 would answer below 0 if allowed. Since the distributions of all items are heavily right-skewed, the resulting model does not adhere to the assumption of multivariate normality, thus making it suspect to estimating wrong standard errors. Consequently, the realisation of answers is simply a truncated normal distribution. Thinking of the distribution as being left-censored, we can impute values such that the items better resemble a normal distribution. The procedure is to draw from the tail of a truncated normal distribution with mean and variance for each item, replacing the values of 0 with a random draw. Technically speaking, an estimation of μ and σ of a truncated multivariate normal distribution is performed using the generalised method of moments (GMM) given the lower (0) and upper (∞) truncation points (Wilhelm 2015). Details on the technical aspects of truncation is given by Lee (1979).

We chose the metric invariance model from Chapter 2, in order to test the performance of the imputations method. In total, the imputation procedure was run with 30 repetitions. The full code is available in Appendix 4. We use the imputation here only to test sensitivity to model assumptions and we are not interested in a pooled analysis (Rubin 1987). Figure 3.6 shows the original distribution of the items in the raw data alongside distributions in random choice of imputed data (repetition 15). The imputed values range from -11.2959 to -0.0002. The extreme low values of the imputation are very rare, since they are in the tail of the distribution. When the number of 0 responses is high, the resulting imputed data will have a long left tail. This can be seen by comparing trust in politicians and trust in the UN in Figure 3.6. Unsurprisingly, the fit indices improve when using imputed data, as shown in Figure 3.7. Across 30 imputed data sets, the CFI ranges from 0.9798-0.982, the RMSEA from 0.0738-0.0782 and the SRMR from 0.0386-0.0414 when estimating the metric measurement invariance model, which is at least as good a fit in the vast majority of cases, compared to the raw data.

Although the fit of the model improves, it is only marginal. An overview of changes to the country-level correlations between NPT and SPT are listed in Appendix 5, which shows some changes for certain countries. It is promising in that allowing for the full range better captures the correlation between NPT and SPT. This is a consequence of the fact that a coarsening of values, here the flooring effect, only can reduce correlation. The effect of extreme response bias is perhaps not that great on the overall analysis, but nevertheless, given the prevalence of a flooring effect, it would seem like a worthwhile task to predict and explain if individuals are judging the questions on a 'different scale from others'. This could possibly have been done by a mixture model, where a latent class indicates whether the respondent is a '0 respondent' or a respondent who utilises the full extent of the scale. Furthermore, it is likely that the response category carries meaningful information that can be explained and analysed through more detailed individual-level analysis. This is beyond the scope of the current thesis, however.

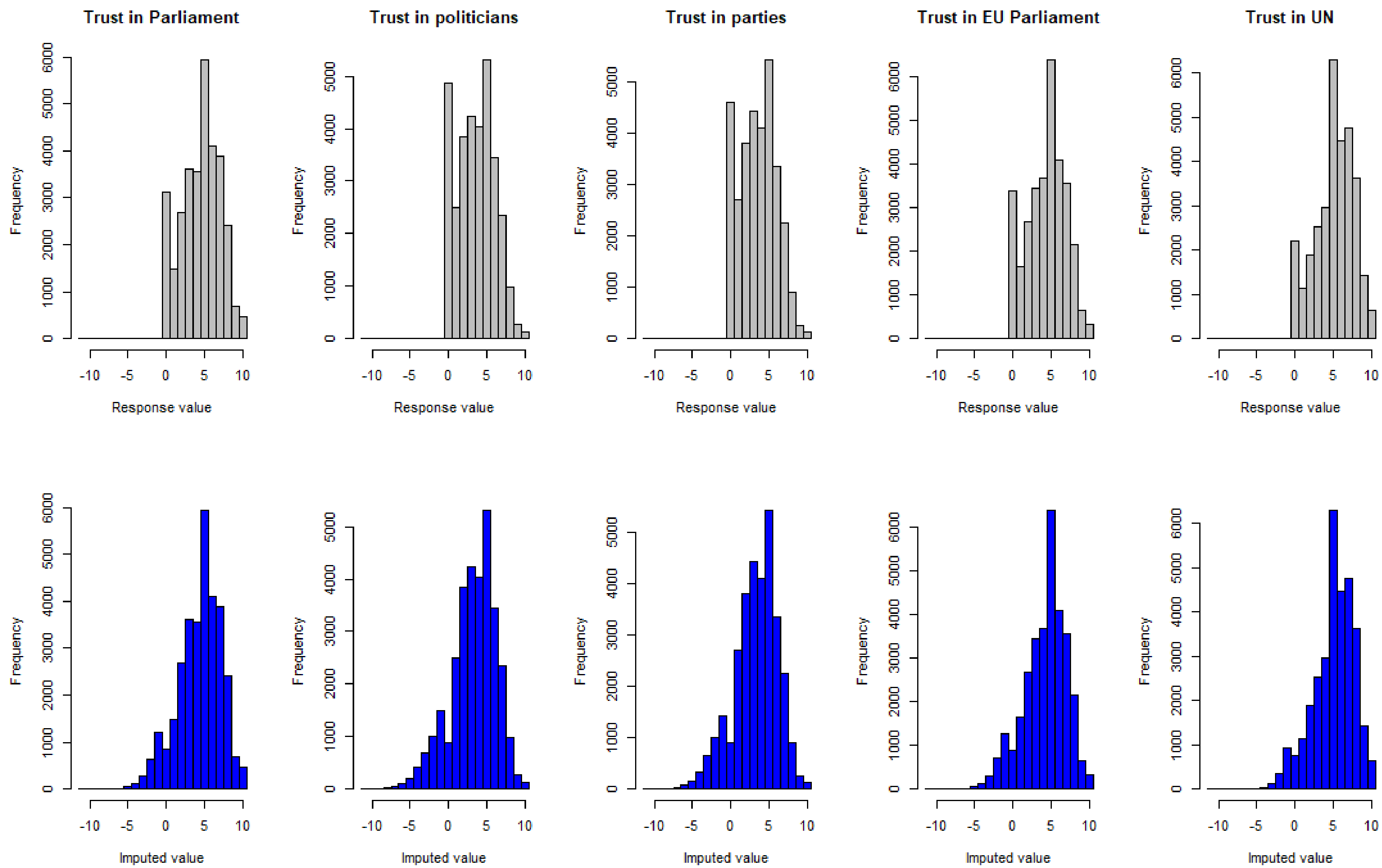


Figure 3.6: Distributions of items in the raw ESS 8 data and imputed data.
Repetition 15 of the imputation data is selected.

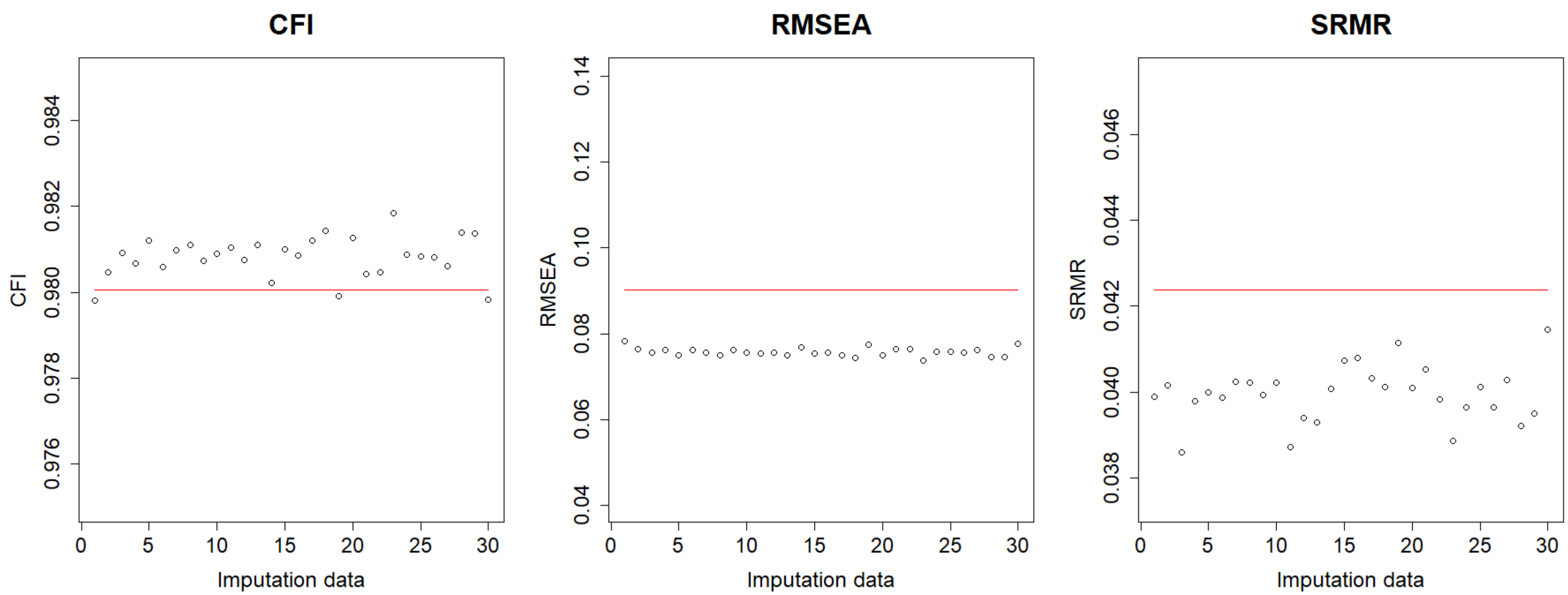


Figure 3.7: Comparison of CFI (left), RMSEA (middle) and SRMR (right) between raw data and 30 imputed datasets. *The red line indicates the model fit of the raw data.*

Chapter 4

A Bayesian approach to measuring political trust across NUTS regions in Europe

As previously mentioned, applied research in the field of social and political sciences has experienced an increase in the number of studies applying multiple groups confirmatory factor analysis (MGCFA) models to investigate measurement invariance of theoretical concepts across groups. However, many studies do not utilise available data to its full potential. A range of large-scale cross-national surveys contain detailed individual-level information on region membership within countries, yet many studies rely on country-level aggregates to investigate between-country differences in latent constructs. While MGCFA models comparing a large number of groups undoubtedly allow for a more nuanced analysis of a given theoretical measure, the researcher faces several obstacles when attempting to estimate such models. One issue is a convincing way of summing up the results in order to make useful comparisons between groups. Another, more technical issue, is the frequent occurrence of small samples within groups, which can result in conventional frequentist maximum likelihood estimation failing to converge or produce 'Heywood cases', which yields improper solutions (Heywood 1931; Kolenikov and Bollen 2012).

The purpose of this chapter is to advocate the use of MGCFA models at the regional level when available, in order to gain additional insights. In order to solve estimation issues, we suggest applying a Bayesian multiple groups confirmatory factor analysis (BMGCFA) approach. First, we restate the principles of the MGCFA method used in the previous chapters to identify the key issues. Second, the issue of the small-N case is presented, followed by a review of the alternative BMGCFA approach, with a particular focus on the choice of priors for the residual variance parameters. Third, we present an application of the method based on the two-factor model used in Chapter 2 and 3. Fourth, the analysis is carried out on a model presented in previous research by Märien (2011) for a one-factor model of political trust measured across 265 NUTS regions in Europe using the European Social Survey, round 8. Finally, we suggest a graphical way of comparing models within the BMGCFA framework using the posterior deviances, alongside a discussion of the potential theoretical insights gained by estimating the two models at the regional level.

4.1 Multiple groups confirmatory factor analysis and its applications

Recall the standard linear SEM model which is defined as

$$\boldsymbol{y} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon},$$

for $j = 1, \ldots, p$ where $\boldsymbol{y} = [y_1, \ldots, y_p]^T$ is a $p \times 1$ vector of observed variables, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_q]^T$ is a $q \times 1$ vector of latent variables, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings and $\boldsymbol{\epsilon}$ is a $p \times 1$ random vector of residual errors.

As previously illustrated, MGCFA is of particular interest in cases where the researcher wants to establish measurement invariance (Jöreskog 1971; Davidov et al. 2014). With the increased access to computing power, even complex models are estimated fairly quickly. Recently, the use of MGCFA models within the social and political sciences has increased significantly. For example, Ariely and Davidov (2010) use MGCFA to investigate measurement invariance in attitudes towards democracy in 36 countries using the World Values Survey. Likewise, Märien (2011), which we will evaluate in Section 4.7, assess measurement invariance for institutional trust across multiple countries on the basis of four waves of the European Social Survey using a one-factor model. However, in both cases, lower-level groupings (NUTS regions) in data is simply ignored, and consequently, their results rely on country level aggregates. Especially within the political and social sciences, aggregate measures of multidimensional theoretical concepts may not be acceptable when it is well known that they encompass complex within-country dynamics. However, moving from a country-level to a regional-level MGCFA model is not necessarily straightforward, as seen in the previous chapter. Besides the substantial difficulties, which arises when interpreting differences between a (large) number of groups, the maximum likelihood estimation procedure also struggles in certain scenarios. This issue shall be discussed in the following.

4.1.1 Heywood cases

One issue when working with small sample sizes is the possible occurrence of

Heywood cases (Gagne and Hancock 2006). A Heywood case is identified when variance estimates are negative or correlation estimates are greater than one in absolute value (Kolenikov and Bollen 2012). Within the CFA framework, a Heywood case typically occurs in the estimation of negative item error variances or non-positive definite covariance matrices (van den Bos 2007; Kolenikov and Bollen 2012). A covariance matrix can be thought of as 'rescaling' a vector parsing out the variance due to different dimensions and as such must be positive definite. By definition, a matrix $\boldsymbol{W}$ is positive definite if $\boldsymbol{z}'\boldsymbol{W}z$ is strictly positive for all real vectors z. In terms of the covariance in the normal distribution, this can be thought of as the distance among a set of points from the mean vector. If the covariance is positive definite, it is possible to calculate its Cholesky decomposition. Besides the obvious issue of the estimates yielding nonsense values, improper solutions potentially influence other parameters in the model thus making model inference unreliable.

The reasons for Heywood cases' occurrence can be many. Most often, the issue can be narrowed down to be related to either model misspecification, low sample sizes or variables misbehaving, i.e. the variables having an uncommon non-normal distribution, extreme values or a data error. Studies on the issue of Heywood cases on error variances has proposed a range of possible workarounds, including fixing the error variance to 0 (Chen et al. 2001). This solution is typically justified in cases when the confidence interval for the negative error variance estimate contains 0 (i.e. when an estimate is negative and close to 0). Other solutions include the application alternative estimation methods for the variance parameters or making use of bootstrapping or robust sandwich estimators. While the ad hoc application of a different estimator might solve the issue, it is by no means guaranteed. Hence, a range of studies chooses to partially ignore the issue and simply fix the error variance instead, which used to perform 'reasonably well' (Dillon, Kumar, and Mulani 1987). This approach is not recommended however, since fixing the variance has the potential to bias other parameter estimates in the model, making model inference questionable at best (Kolenikov and Bollen 2012).

When using MGCFA models across a large number of groups, the likelihood of Heywood cases appearing increases, which might deter researchers from proceeding with such studies. However, we suggest handling the small N case using a Bayesian approach, which not only avoids this issue altogether but introduces a more flexible and explicit modelling framework.

4.2 The application of BMGCFA

From a substantive point of view, there are several appealing reasons for employing Bayesian methods in the MGCFA framework. First, it enables the researcher to directly incorporate prior knowledge into the model. Second, the use of Bayesian methods allows for increased model flexibility through the specification of prior distributions. Third, we are able to overcome issues with the maximum likelihood estimator such as model convergence and Heywood-cases.

Bayesian Multiple Groups Confirmatory Factor Analysis (BMGCFA) models have been used in a variety of fields and topics with some recent studies pointing out some of the advantages of applying them to well-known models. In more general terms, Muthén and Asparouhov (2012) advocate for the use of BSEM in substantial research. They point out how ML and likelihood-ratio chi-square testing relies on strict models, which often results in the rejection of sensible models following the general null-hypothesis testing. The more specific applications are plenty and is highlighted by the fact that Bayesian estimation can help overcome convergence problems (Depaoli and Clifton 2015; Hox, van de Schoot, and Matthijsse 2012;

Fontanella et al. 2016). However, as noted by McNeish (2016) the influence of weak priors on parameter estimates for small samples can, in some cases, lead to worse results compared to a conventional frequentist approach. Hence, it is necessary to exhibit caution in the prior specification.

In addition to the standard application of measurement invariance testing, Bayesian approaches have recently found their way to the literature. More specifically, some studies now incorporate 'approximate measurement invariance'. Instead of fixing parameters to be exactly equal across all groups, a zero-mean prior with small variance is applied for the difference in intercept or loadings between groups. One example is from Verhagen and Fox (2012) who applies Bayesian approximate measurement invariance using the ESS on attitudes towards immigrants - an approach which is further developed in Shi et al. (2017). However, it is important to stress that this does not necessarily solve the estimation issues discussed above, namely the issue of Heywood cases. While it does move away from an 'exact measurement invariance' approach and makes it easier to argue for measurement invariance when the actual differences in parameter estimates are small, the maximum likelihood estimator may still struggle in cases where the sample sizes are small.

4.3 The case: re-measuring political trust in European countries

We will now introduce the development of a BMGCFA model based on the established MGCFA model presented in Chapter 2 and 3 using the five questions from the European Social Survey questionnaire; trust in the county's parliament, politicians, political parties, European Parliament and the United Nations.

4.3.1 Data

We will use the same sample of the European Social Survey, round 8 (2016) as previously, consisting of 18 European countries divided into a total of 249 regions, where only complete cases are studied (31,952 observations in total).

4.3.2 Defining a BMGCFA model with phantom-latent variables

Consider the simple measurement model $y^g = \alpha^g + \Lambda^g \eta^g + \epsilon^g$ for individual $i = 1, \ldots, N$ in group $g = 1, \ldots, G$. In traditional BMGCFA we assume that $\epsilon^g \sim N_p(0, \Theta^g)$ and $\eta^g \sim N_m(\nu^g, \Psi^g)$, where Θ^g and Ψ^g are residual covariance matrices. To obtain posterior estimates of ν^g on the same scale as y^g we apply effects coding in a similar way as in Section 2.5 for frequentist MGCFA models. Specifically, we set $\sum_{i=1}^{N} \alpha^g = 0$ and $\sum_{i=1}^{N} \lambda^g = 1$ for all $g \in G$. In the following, we drop the notational dependence on g whenever it is not needed to distinguish group-specific effects.

We recall that the model measures political trust on two latent variables using five observed items (that is $p = 5$ and $m = 2$), which are all measured on a scale from 0-10. Also, the model assumes correlated residuals between trust in politicians and trust in political parties (item 2 and 3). As a result, the observed residual covariance matrix is

$$\Theta = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & \theta_{23} & 0 & 0 \\ 0 & \theta_{32} & \theta_{33} & 0 & 0 \\ 0 & 0 & 0 & \theta_{44} & 0 \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

Since the MCMC estimation of the off-diagonal elements Θ is known to be difficult

Merkle and Rosseel (2018), we can introduce a set of phantom-latent variables D with associated factor loadings Λ_D such that $\epsilon = \Lambda_D D + \epsilon^*$where Λ_D is a $p \times v$ matrix of factor loadings, D is a $v \times 1$ vector of latent variables and ϵ^* is a $p \times 1$ vector of residuals. This separation strategy was proposed by Bernard, McCulloch, and Meng (2010) and has since been implemented in, for example, the Bayesian SEM package 'blavaan' (Merkle and Rosseel 2018). By specifying the non-zero entries in Λ_D correctly we get that $D \sim N_v(0, \Psi_D)$ and $\epsilon^* \sim N_p(0, \Theta^*)$ where both Ψ_D and Θ^* are diagonal. In our case, we get the following working model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda_{D21} \\ \lambda_{D31} \\ 0 \\ 0 \end{bmatrix} D1 + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \epsilon_3^* \\ \epsilon_4^* \\ \epsilon_5^* \end{bmatrix}$$

In the working model, we set $\psi_D = 1$, and in addition

$$\lambda_{D21} = \sqrt{|\rho_{23}|\,\theta_{11}} \quad \text{and} \quad \lambda_{D31} = \operatorname{sign}(\rho_{23})\sqrt{|\rho_{23}|\,\theta_{22}},$$

where ρ_{23} denotes the correlation parameter belonging to the covariance θ_{23} and $\operatorname{sign}(\rho_{23})$ is 1 if $a > 0$ and -1 otherwise. The working residual variances are then

$$\theta_{22}^* = \theta_{22} - |\rho_{23}|\,\theta_{22} \quad \text{and} \quad \theta_{33}^* = \theta_{33} - |\rho_{23}|\,\theta_{33}.$$

As a result, we are only required to set priors for the residual variances θ_{22}, θ_{33} and the correlation parameter ρ_{23} in the inferential model. The full code for defining a configural invariance model using the phantom-latent variable specification can be found in Appendix 6.

There are several advantages of applying the separation strategy in BMGCFA models. Besides the desirable decrease in computing time (because the sampler only has to evaluate diagonal matrices), we are primarily concerned with the increased prior flexibility for the residual standard deviations through the specification of univariate prior distribution. If one were to assign a prior to the raw covariance matrix, the possibilities are severely limited due to the restriction of the prior, $\mathrm{p}(\Theta)$ being semi-positive definite. A popular choice is to assign $\mathrm{p}(\Theta) \sim \mathrm{Wishart}(I, d)$or $\mathrm{p}(\Theta) \sim \mathrm{InverseWishart}(I, d)$, where I refers to the identity matrix and d to the degrees of freedom (Gelman et al. 2013). We immediately realise that our primary way of expressing the uncertainty around Θ is to vary d accordingly, which applies to all entries in Θ (Barthelmé 2012). Furthermore, we may have little or no prior knowledge about ρ but a strong prior idea about of the distribution of Θ. Unfortunately, we have no straightforward way of incorporating varying prior knowledge in Θ, unless we apply a separation strategy like the one mentioned above.

4.3.3 Choice of prior distributions

We shall apply the following priors for the intercepts, loadings, latent means, latent variance/covariance and correlations:

$$
\begin{aligned}
\alpha_p &\sim N(\mu_\alpha, \frac{1}{\tau_\alpha}) \\
\lambda_p &\sim N(\mu_\lambda, \frac{1}{\tau_\lambda}) \\
\nu_p &\sim N(\mu_\nu, \frac{1}{\tau_\nu}) \\
\rho_D &\sim -1 + 2 \cdot \text{Beta}(a, b) \\
\Psi_p &\sim \text{InverseWishart}\,(I, 3)
\end{aligned}
$$

For illustrative purposes, we set $\mu_\alpha = 0$ and $\mu_\lambda = 1$, since we apply effects coding with the implied restriction that $\sum_{i=1}^{N} \alpha^g = 0$ and $\sum_{i=1}^{N} \lambda^g = 1$. Furthermore, we set $\mu_\nu = 5$, which is the middle point of the original item scale. However, the choice of μ has minimal impact on the posterior distributions, as long as the prior variance is large, making it a favoured weak (or 'flat') prior for parameters with support on $-\infty$ to ∞. Hence, we set the precision $\tau_\alpha, \tau_\lambda$and τ_ν to 0.01, resulting in the normal priors being close to flat in the interval of interest for each parameter. We have chosen the symmetric $-1 + 2 \cdot \text{Beta}(1, 1)$ prior for the correlation parameter, which has support on $[-1; 1]$. This is true, since every $\text{Beta}(a, b)$ can be shifted to be symmetric on the interval (c, d) if we multiply it by $(d - c)$ and add c (L'ecuyer and Simard 2006). Although the $-1 + 2 \cdot \text{Beta}(1, 1)$ prior is equivalent to a $\text{Uniform}(-1, 1)$ prior, the $-1 + 2 \cdot \text{Beta}(a, b)$ prior, in general, allows us to control the tail behavior of the distribution if we so wish. We also choose to assign an inverse Wishart prior to the latent covariance matrix. Previous models did not indicate any issues with the latent variances being negative. With the inverse Wishart distribution being the conjugate prior for the covariance matrix for the multivariate normal distribution, it is the natural choice. However, if the latent variances exhibited the same issues as the residual variances, the same separation strategy as mentioned above could be applied.

Setting the priors for the residual variances is more involved. Running the model using MLE results in a range of Heywood cases, which, as previously mentioned, occurs in small-N groups. In this section, we explore a range of different priors on the residual variances (or standard deviations), followed by a brief justification for the prior to be used in the following BMGCFA models. We shall investigate the posterior distribution for θ_{11} in a low $N = 9$ group (S212: Kronoberg County in southern Sweden) for a selection of commonly used priors, which can be found in Figure 4.1 to 4.4. Please note that the scale of the axes are not the same between the different posterior distributions. We aim to make an informed choice of prior when only little data is available, based on the posterior distribution, while ensuring that the choice of prior does not distort the posterior distribution in groups where more data is available. For these illustrative runs, all models are estimated using JAGS using the R2jags version 0.5-7 in R with two chains, burnin = 2,000, iterations = 200,000. We considered accounting for the floor effects discussed in Section 3.4.1 by specifying zero values on the items as missing values truncated to the right in 0, something that is straightforward in JAGS. However, we chose to omit this in favour of simplicity. The posterior draws for each iteration are represented as a red and black line for chain one and two respectively.

4.3.3.1 InverseGamma $(\varepsilon, \varepsilon)$ on the residual variances

The inverse gamma distribution is commonly used as a default choice of prior for the residual variances in Bayesian estimation software (Lunn et al. 2000). One of the persuading properties of the inverse gamma prior is its conditional conjugacy for the gamma likelihood, which allows one to analytically derive the conditional

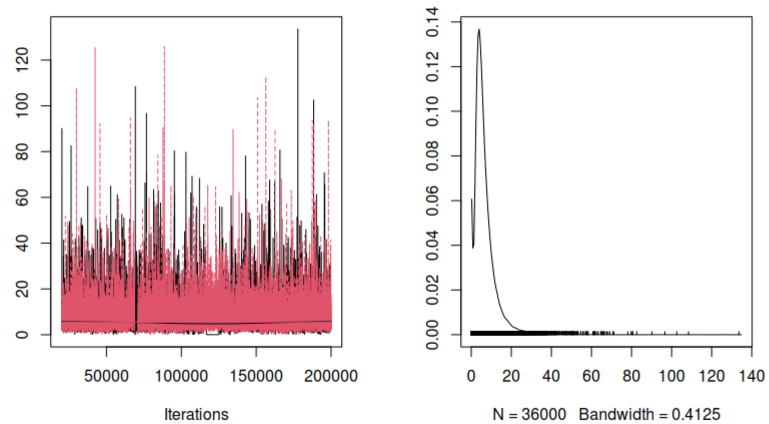


Figure 4.1: Inverse Gamma prior on the residual variance.

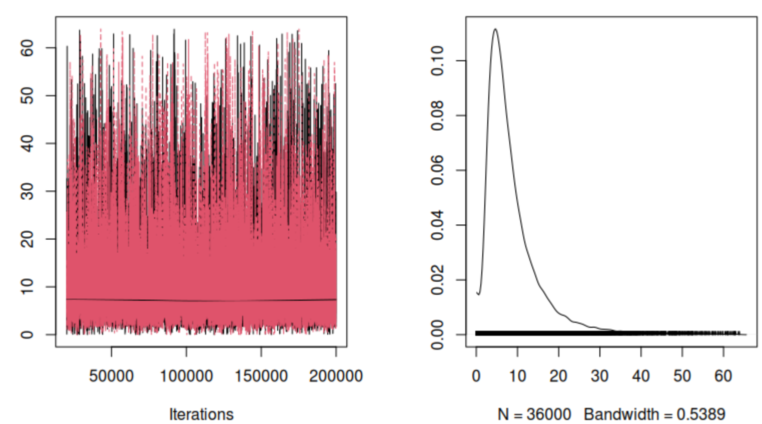


Figure 4.2: Uniform prior on the residual standard deviation.

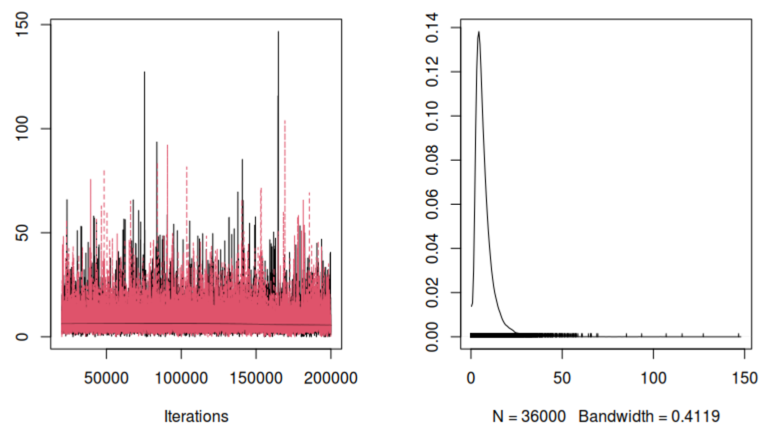


Figure 4.3: Half-Cauchy prior on the residual standard deviation.

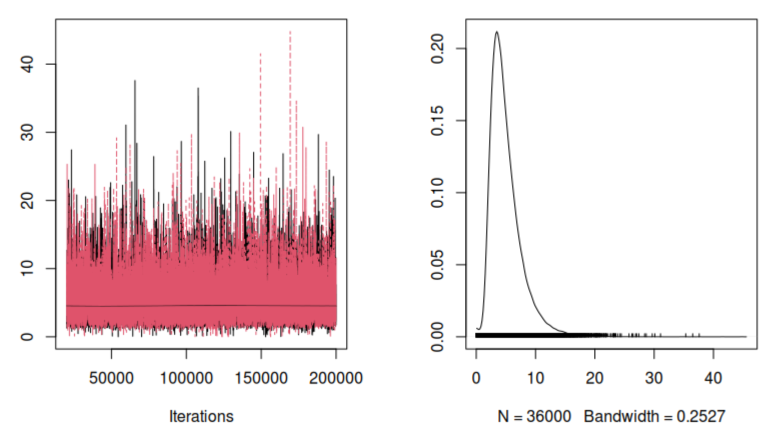


Figure 4.4: Half-t prior on the residual standard deviation.

posterior distributions. However, recent studies have highlighted some potential issues with the inverse gamma distribution when data is sparse. For example, Gelman (2006) describes this:

> "[...] a difficulty of this prior distribution is that in the limit of $\varepsilon \to 0$ it yields an improper posterior density, and thus must be set to a reasonable value. Unfortunately, for datasets in which low values of σ_α are possible, inferences become very sensitive to the choice of ε, and the prior distribution hardly looks noninformative [...]".

Figure 4.1 shows the posterior draws from θ_{11} in the model for NUTS S212 (posterior mean = 6.69). The distribution has a sharp peak but encompasses occasionally extreme values of θ_{11}. Even though the inverse gamma, $IG\,(0.001, 0.001)$ is frequently used as a weak prior for θ, it performs badly in this case, since the posterior estimate for the residual variance is inflated in the case of a low sample size.

4.3.3.2 Uniform$(0, 8)$ prior on the residual standard deviations

An alternative choice of prior for the residual standard deviations is the uniform prior on the interval $(0, A)$ where A is set to some reasonable value. The main advantage of the uniform distribution over the inverse gamma is the behaviour in the limit $\sigma \to 0$. Figure 4.2 shows the posterior distribution for a Uniform $(0, 8)$ prior on θ_{11}, which corresponds to a range of [0;64] for the posterior residual variance. Compared to Figure 4.1, the distribution takes on values close to 0 but exhibits an unrealistically long tail (posterior mean = 8.8). It appears that the uniform distribution leads to over-estimation of θ in the case of small sample size due to the heavy right tail. Although we have the option to adjust the value of A, it is completely contingent on an arbitrary cut-off value, which is not desirable if we have little to no prior knowledge about the distribution of θ. If we set A to be 'high', we will experience posterior draws close to $A/2$ when the sample size is small, while if we set A to be 'low' we do not allow for posterior draws which would normally occur at the tail of the distribution.

4.3.3.3 Half-Cauchy and half-t prior on the residual standard deviations

Investigations made by Gelman (2006) have demonstrated how the folded half-t family of priors on the residual standard deviations is an alternative to the inverse gamma and the uniform priors mentioned above. Consider a half-Cauchy (HC) prior with scale parameter A which is a special case of the half-t (HT) distribution with one degree of freedom. The standard Cauchy distribution takes values $\theta \in (-\infty, \infty)$:

$$f(\theta; \theta_0, A) = \frac{1}{\pi \left[1 + \left(\frac{\theta - \theta_0}{A}\right)^2\right]}$$

Since for variances $\theta \in (0, \infty)$ the half-Cauchy is defined on positive numbers as a special case of a HT:

$$f(\theta; A) = \frac{2}{\pi A \left[1 + \left(\frac{\theta}{A}\right)^2\right]}$$

We observe that the probability density function decreases as a function of the distance from 0 as the denominator increases as θ gets larger. As $A \to \infty$, the HC approaches a uniform distribution, while it has a broad peak at 0 for lower values of A and a much heavier tail compared to, for example, the half-normal distribution. Note that the mean and variance of the half-Cauchy are undefined. However, this does not mean that the posterior means and variances are undefined. Figure 4.3 shows the posterior draws from a $HC(10)$ prior on θ_{11} in the low N group (posterior mean = 7.19). The distribution resembles that of the inverse gamma IG $(0.001, 0.001)$ prior, with a shorter tail. We can apply a more informative prior by defining a HT distribution with d.f. > 1, which has the probability density function:

$$p(\theta) \propto \left[1 + \frac{1}{v}\frac{\theta^2}{\sigma^2}\right]^{-\frac{v+1}{2}}$$

Here, v denotes the degrees of freedom. Figure 4.4 shows posterior draws with d.f. $= 20$ and scale 2. Compared to the weaker HC distribution, the extreme samples in the tail are eliminated, while we keep the characteristics of the distribution (posterior mean = 5.05). In the following models, we shall make use of the HT distribution for both the country-level and the NUTS-level invariance models.

4.4 Model fit indices and diagnostics

In order to compare BMGCFA models, several model selection criteria have been developed. One of the most common indices is the deviance information criterion (DIC), which has been applied by researchers for a number of years and is incorporated in several software packages (Lunn et al. 2000; Plummer, Stukalov, and Denwood 2016; Merkle and Rosseel 2018). The DIC is also used for other types of models, and can also be used as an alternative to fit indices in MGCFA models. In addition, recent developments include both the widely applicable information criterion (WAIC; also called the Watanabe-Akaike information criterion) and LOOIC (leave-one-out information criterion).

The DIC is determined by calculating the deviance, adding the effective number of parameters, that is $DIC = \overline{D(\theta)} + p_D$. The deviance is defined as minus two times the log likelihood:

$$D(\theta) = -2\ell(\theta; x)$$

The mean and variance of the deviance is calculated across the posterior draws of θ. Next, p_D can either be calculated as $p_D = \overline{D(\theta)} - D\left(\bar{\theta}\right)$,which was suggested by Spiegelhalter et al. (2002) or as $p_D = \frac{1}{2}\overline{var\left(D(\theta)\right)}$ later suggested by Gelman et al. (2013). We have chosen the latter in this analysis but refer to the original texts for more detail on the DIC calculations.

One of the criticisms of the DIC is that it is not entirely Bayesian since it is a based on point estimate (van der Linde 2005) and can produce negative estimates (Plummer 2008). Hence, Vehtari, Gelman, and Gabry (2016) provides a strong argument for using WAIC and LOOIC instead, which is incorporated into Bayesian software package 'Stan' (Carpenter et al. 2017) and the R package 'loo' (Gabry 2019). They sum up the advantages as follows:

> "WAIC is fully Bayesian in that it uses the entire posterior distribution, and it is asymptotically equal to Bayesian cross-validation. Unlike DIC, WAIC is invariant to parametrization and also works for singular models. Although WAIC is asymptotically equal to LOO, we demonstrate that PSIS-LOO is more robust in the finite case with weak priors or influential observations." (Vehtari, Gelman, and Gabry 2016)

Both DIC, WAIC and LOOIC measures parsimony of fit, adjusted for model complexity. Hence, the values of the information criterions only carry meaning when comparing competing models. As a result of this, it is important that each model is run with the same data and number of iterations. We chose to report both the DIC, WAIC and LOOIC in the following.

4.4.1 Computational considerations

All models were run with 25,000 iterations with burnin = 1000 and thinning = 5 across two chains, resulting in 10,000 saved iterations per model (5,000 per chain). The full R code for running a configural invariance model is available in Appendix 7. The mean run time per model was 8.5 hours on a laptop with a 12-core i9-8950HK processor and 32 GB of RAM (64 GB including swap space) running Arch Linux. The models were run using the 'saveJAGS' package, a wrapper for 'rJAGS', where each 1000 iterations were dumped to disk (for debugging and RAM saving purposes).

The reason for running the model with more than one chain is two-fold. First, from an efficiency perspective, two chains can be run in parallel, which saves computing time per iteration. Second, it is possible to diagnose potential between-chain differences, which could be a result of model misspecification or convergence issues. A convergence issue could be, that global convergence was not achievable with each chain converging to different values. This is a particularly useful metric when looking at parameters with low sample efficiencies, such as parameters in small NUTS regions.

4.4.2 Model fit

Running the configural and metric invariance models on the country-level (18 countries) and NUTS-level (249 regions) gives the DIC, WAIC and LOOIC as listed in Table 4.1.

Model	**DIC**	**WAIC**	$SE(WAIC)$	**LOOIC**	$SE(LOOIC)$
NUTS configural	658408	659164	*900*	658178	*836*
NUTS metric	661123	662828	*867*	661714	*840*
NUTS scalar	659066	663236	*975*	659561	*849*
Country configural	631579	631358	*864*	631533	*871*
Country metric	660796	662455	*901*	660766	*849*
Country scalar	636424	635132	*843*	636071	*869*

Table 4.1: DIC, WAIC and LOOIC for country-level and NUTS-level models in 21 countries and 249 regions.

We remember that the indices in Table 4.1 are not the same as the fit indices used in the frequentist models in Chapter 2. According to the information criterion in Table 4.1, the country-level models appear to provide a better fit than the NUTS-level model for both the configural, metric and scalar invariance models. However, the comparison of a country and NUTS level model is only possible due to the application of Bayesian MGCFA; Chapter 3 illustrated how it is not possible to fit a NUTS level model using conventional maximum likelihood. It is also necessary to keep in mind that the difference between, for example, the NUTS metric and country metric invariance model is small, if even relevant, considering the size of the standard error of both the WAIC and LOOIC estimates. Although we will not dive deeper into the intricacies of model evaluation and selection using cross-validation, we follow the general guidelines given by Vehtari, Gelman, and Gabry (2016). In the following, we will be evaluating and interpreting on the NUTS-level models, which is the primary goal of this chapter.

Looking at the NUTS level models, the difference between the NUTS scalar invariance and configural invariance model is negligible relative to the size of the standard error. In other words, when we run a more restrictive model, the decrease in number of parameters outweighs the flexibility of having factor loadings and intercepts vary between groups.

Interestingly, the NUTS metric invariance model provides the worst fit (DIC, WAIC and LOOIC all above 659,000 which is higher than the corresponding configural and scalar invariance models). This highlights the complications of interpreting the information criterion when the number of groups and/or parameters is large. We extract the posterior draws of 360 parameters in the country-level model and 4980 parameters in the NUTS-level models, which entails all non-deterministic nodes in the model specification. When comparing the configural and metric invariance models, we would expect a more restrictive model to provide a better fit if the NUTS regions were relatively similar, due to the reduction in model complexity. On the other hand, if NUTS regions exhibit extreme differences, the imposed restrictions may force the remaining model parameters to follow a distribution which decreases model fit overall. In essence, we can think of this as a balancing issue:

1. Relative differences across NUTS regions
2. Violation of distributional assumptions
3. Model complexity

Even in a model where (2) is not an issue, the balance between (1) and (3) is difficult. We do not have an intuitive way of determining the optimal tradeoff

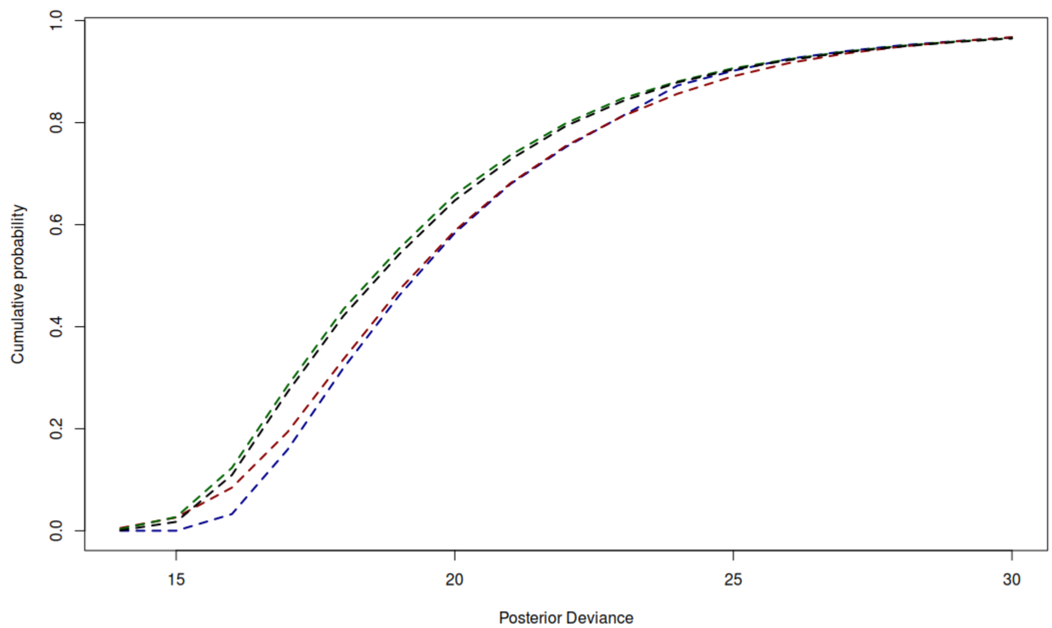


Figure 4.5: ECDF of the posterior deviances for the country-level configural invariance model (green), country-level scalar invariance model (black), NUTS-level configural invariance model (blue) and NUTS-level scalar invariance model (red).

between capturing differences between NUTS regions, model complexity, and its effect model fit when imposing restrictions on model parameters. We believe that this is an interesting point of investigation for future research dealing with a large number of groups in a BMGCFA setting.

One very useful way of graphically comparing models is to consider the empirical distribution function (ECDF) of the posterior deviance, shown in Figure 4.5 (Aitkin 2010). In other words, since the deviance is a function of the parameters, we obtain the posterior distribution of the deviance by plugging in our posterior parameter draws in the deviance function. As seen in Figure 4.5, the two NUTS-level and the two country-level ECDFs look close to identical with the majority of the deviances being in the range of 15-25. However, for both pairs, the distribution of the configural invariance model is slightly left-skewed, indicating a better fitting model even though this strictly speaking is inconclusive according to the criteria set out by Aitkin, Vu, and Francis 2016. Compared to using, for example, Bayes factors as a model comparison tool, the ECDF has the advantage of being both intuitive to interpret and easy to calculate. In addition, when the prior is improper, it is not possible to use the Bayes factor. However, the ECDF does not suffer from improper priors, since it depends solely on the posterior. To our knowledge, this approach has not been applied to measurement models in a Bayesian framework before and might serve as a useful approach for future studies within this field.

4.4.3 Model diagnostics

With the application of Bayesian methods, it is necessary to investigate model convergence and diagnostics. In addition, it is possible to detect potential model issues by comparing trace plots between low-N and high-N groups, which is the main focus of this analysis. As an example, we have present the trace plots and posterior densities for the $\alpha_{1,2,4}$ and $\theta_{1,2,4}$ parameters in two groups in the configural invariance model (since α_{1-4} does not vary across groups in the scalar invariance model):

- Group A (NUTS 154): ITH2 (Trento) with N = 13.
- Group B (NUTS 72): ES23 (La Rioja) with N = 7.

For group A, the two chains in Figure 4.6 appear to mix quite well and are stable right after burnin. We recall that α_3 is defined as a function of α_1 and α_2 due to effects coding. The same is true for α_5 being defined as a function of α_4. The trace plots for group B shows equally good mixing. It is relevant to point out that the 95% HDI (highest density interval) for the draws ranges from around -20 to 20 in both groups. This is a direct result of 1) applying a flat prior on α and 2) small sample sizes. If we had a better prior idea about the distribution of α in the two groups, this could have been incorporated directly into the modelling framework.

Trace plots and posterior densities for the error variances θ are shown in Figure 4.7. Both groups show decent MCMC mixing, with a reasonably short tail. However, we see occasional draws from the tail of the distribution in both groups. The primary reason is, that the residual variances for these groups are small (minimum draw less than 0.002). Imposing stronger priors on the precision/variance parameters would be an option to further reduce the number of draws from the tail of the distribution.

4.4.4 Model estimates

In Table 4.2, we summarise the NUTS-level scalar invariance model with the mean value of the parameters of interest, their standard deviations and the highest and lowest observed values for the latent means and latent correlation. The highest mean level of NPT is found in Åland (11 respondents), an autonomous archipelago province in Finland, which only contain around 0.5% of the Finnish population. The highest level of SPT is in Telšiai County (92 respondents), a relatively small region of Lithuania, containing around 5% of the Lithuanian population. For both NPT and SPT, the lowest latent means are in Valle d'Aosta (31 respondents); an autonomous region in north-western Italy, which is the smallest, least populous,

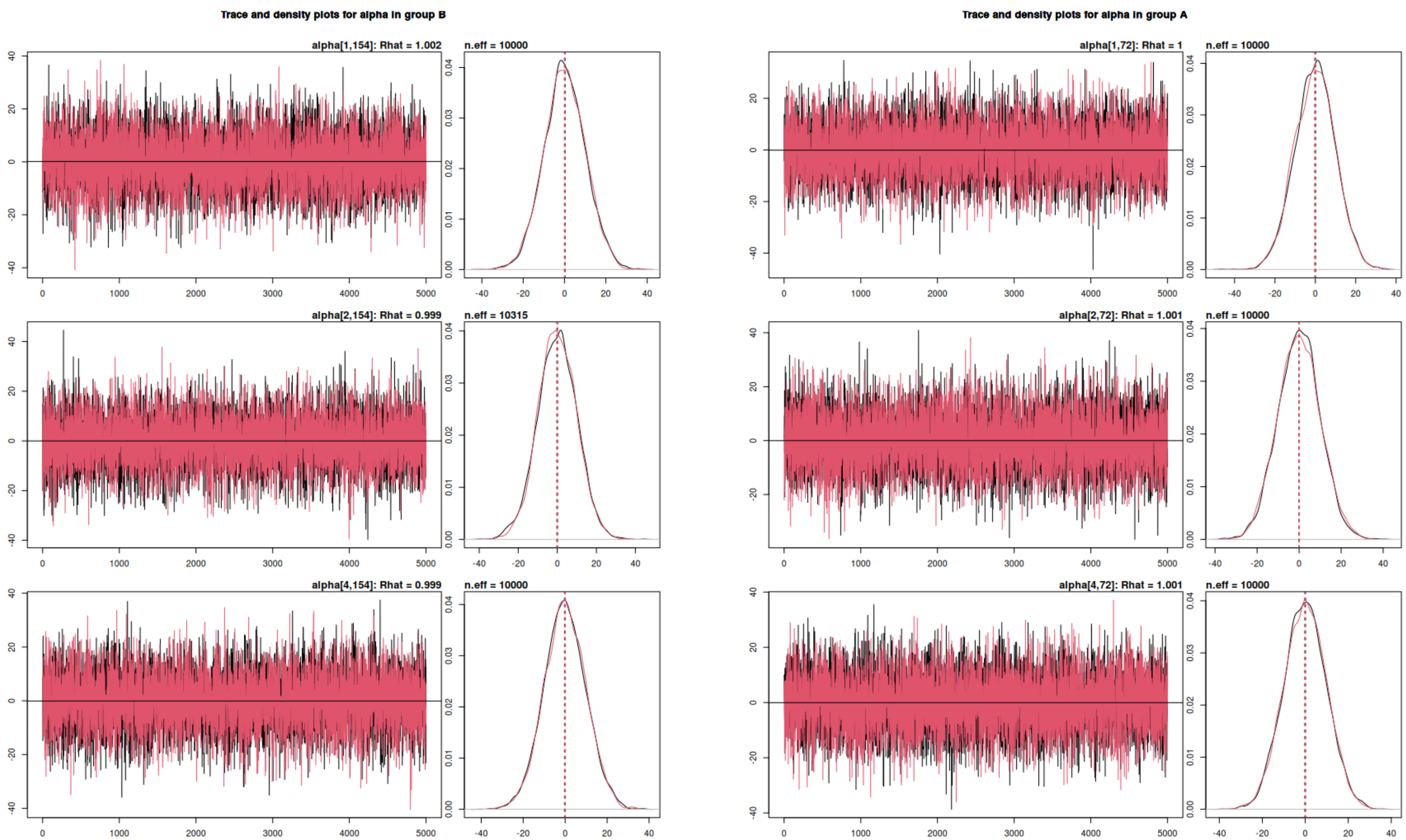


Figure 4.6: Trace plots and posterior densities for α_1,α_2 and α_4 in Group A (left) and Group B (right).

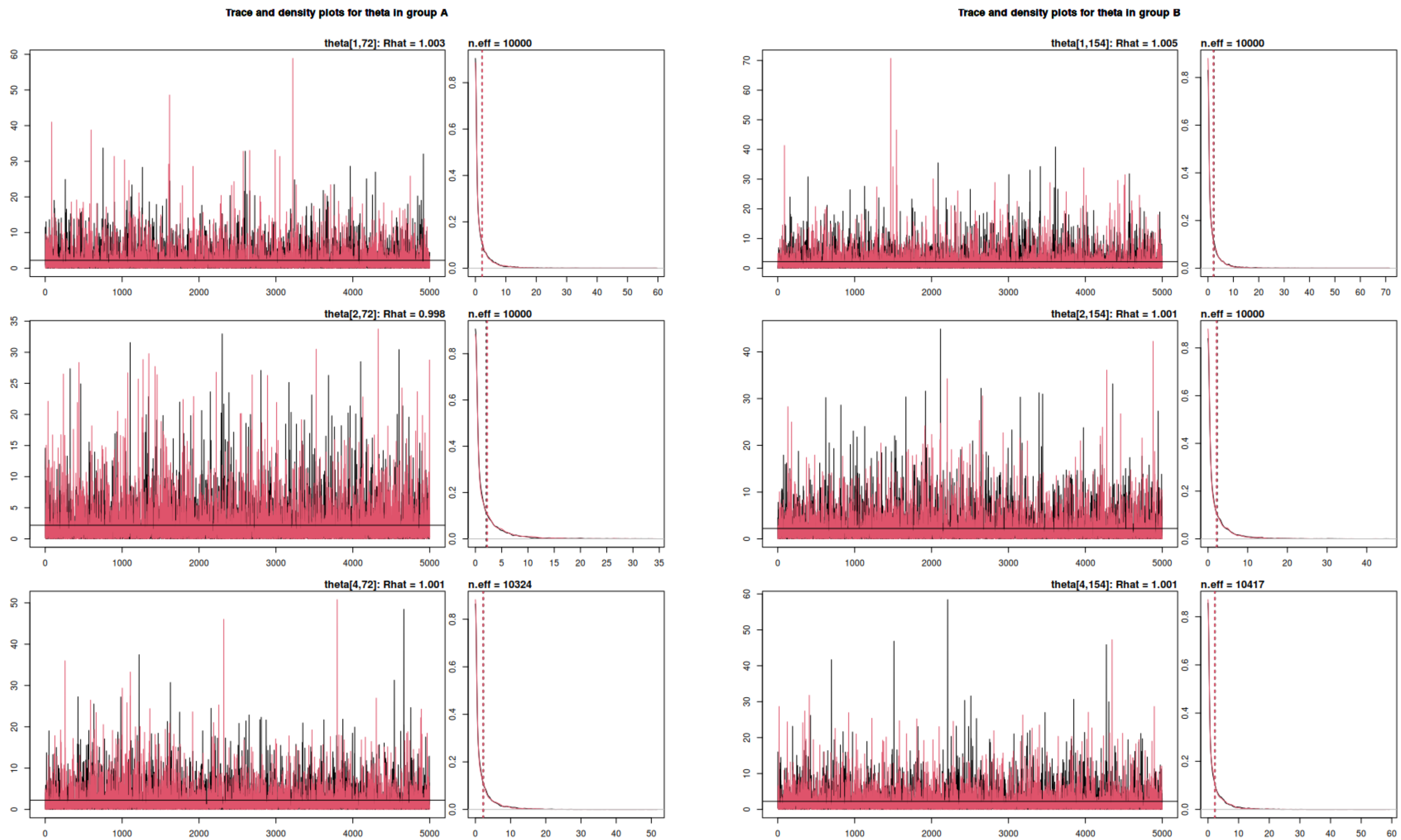


Figure 4.7: Trace plots and posterior densities for θ_1,θ_2 and θ_4 in Group A (left) and Group B (right).
Red dotted line indicate the posterior mean variance. 'n.eff.' is the effective sample size. 'Rhat' is a diagnostic of within and between-chain convergence, where Rhat < 1.05 indicates a good mix.

and least densely populated region of Italy. Looking at the raw NUTS-level data, the low level of trust in this particular region is unsurprising, given that two-thirds of the respondents answered “0: no trust at all” on all trust measures.

For the correlation between NPT and SPT, the lowest value is found in Somogy County (42 respondents). The low correlation between NPT and SPT is due to a high number of 0 responses on NPT, while SPT is relatively high (NPT latent mean: 2.88, SPT latent mean: 5.27). The highest correlation between NPT and SPT is in Cantabria, a community in Northern Spain (17 respondents). For both NPT and SPT, the distribution in the raw data is notably right-skewed (NPT latent mean: 3.27, SPT latent mean: 2.85).

Parameter	Mean	Std. dev.	Minimum	Min. NUTS	Maximum	Max. NUTS
ν_1	3.84	0.69	1.84	ITC2	5.48	FI200
ν_2	4.67	0.74	2.04	ITC2	7.06	LT028
$\rho_{1,2}$	0.70	0.15	0.22	HU232	0.95	ES13
λ_1	0.95	0.18	-	-	-	-
λ_2	0.59	0.41	-	-	-	-
λ_3	1.45	0.50	-	-	-	-
λ_4	1.06	0.01	-	-	-	-
λ_5	0.93	0.01	-	-	-	-
α_1	0.81	0.70	-	-	-	-
α_2	1.29	1.56	-	-	-	-
α_3	-2.10	1.97	-	-	-	-
α_4	-0.67	0.05	-	-	-	-
α_5	0.67	0.05	-	-	-	-

Table 4.2: Summary of posterior estimates for the NUTS scalar invariance model.

One big advantage of the NUTS level model appears when looking at the trace plots and estimated latent means. Departure from normality may be difficult to detect when using a country level model, due to the large sample size. However, in the NUTS level model, it is more obvious which parts of a country have a high degree of distrust. Even though the NUTS level modelling procedure can be quite complex due to the careful selection of priors and large number of parameters to be estimated, it carries the benefit of revealing where bad fit may be a result of scale limits.

4.5 Regional differences in political trust

A graphical overview of the estimated latent mean level of NPT is shown in Figure 4.8, which is based on octiles of the distribution of latent means across all NUTS regions. The estimated latent mean level of SPT is shown in Figure 4.9, and the latent variable correlation is shown in Figure 4.10. All maps show octiles of the distribution.

We find notable within-country differences in NPT and SPT in some countries, while other countries exhibit a more homogenous distribution of trust across NUTS regions.

4.5.1 NUTS-level differences in NPT

Looking at Figure 4.8, the Nordic countries, Benelux countries and Germany contain NUTS regions which are all on the upper half of the distribution for NPT. On the other hand, NUTS regions in Italy, Poland, France and Spain all have low average levels of NPT. Finally, the distribution of NPT in NUTS regions within

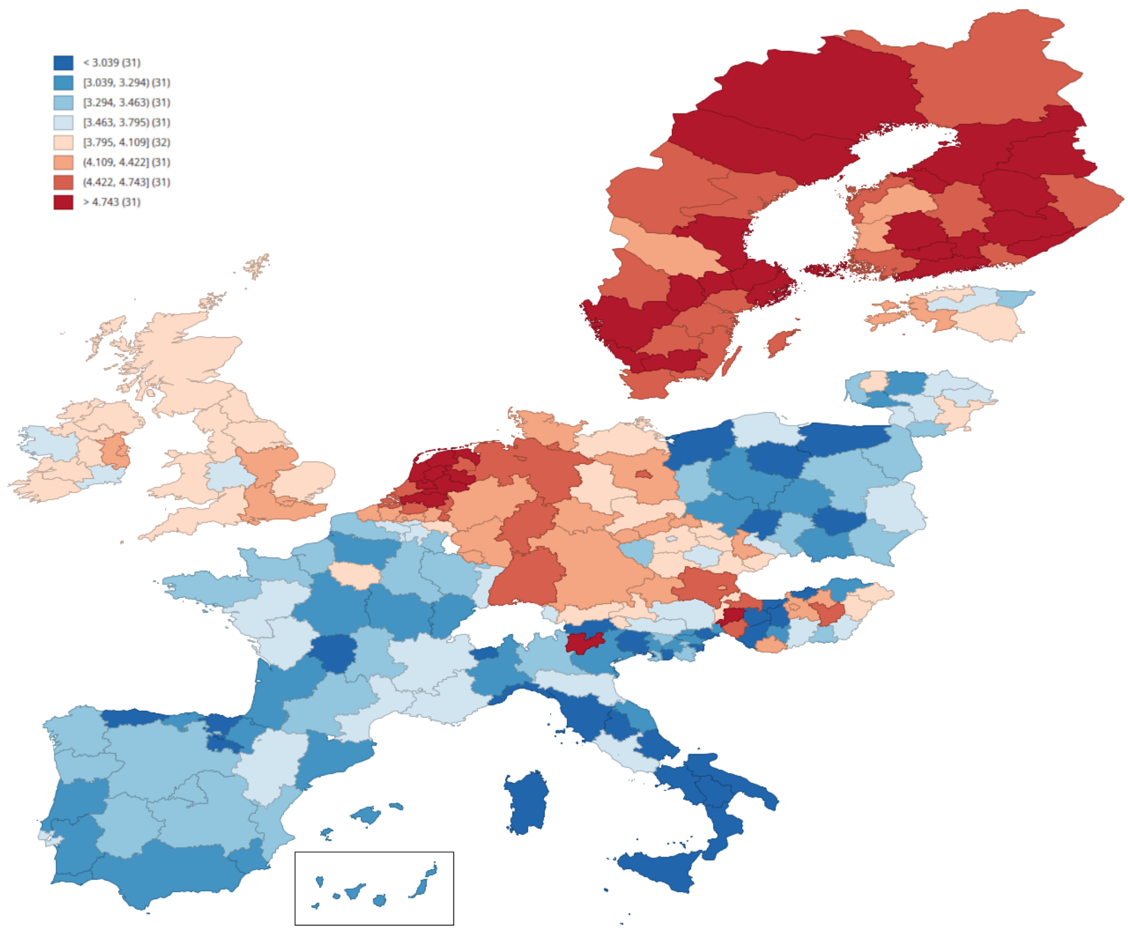


Figure 4.8: Posterior latent mean level of NPT in the NUTS BMGCFA scalar invariance model.

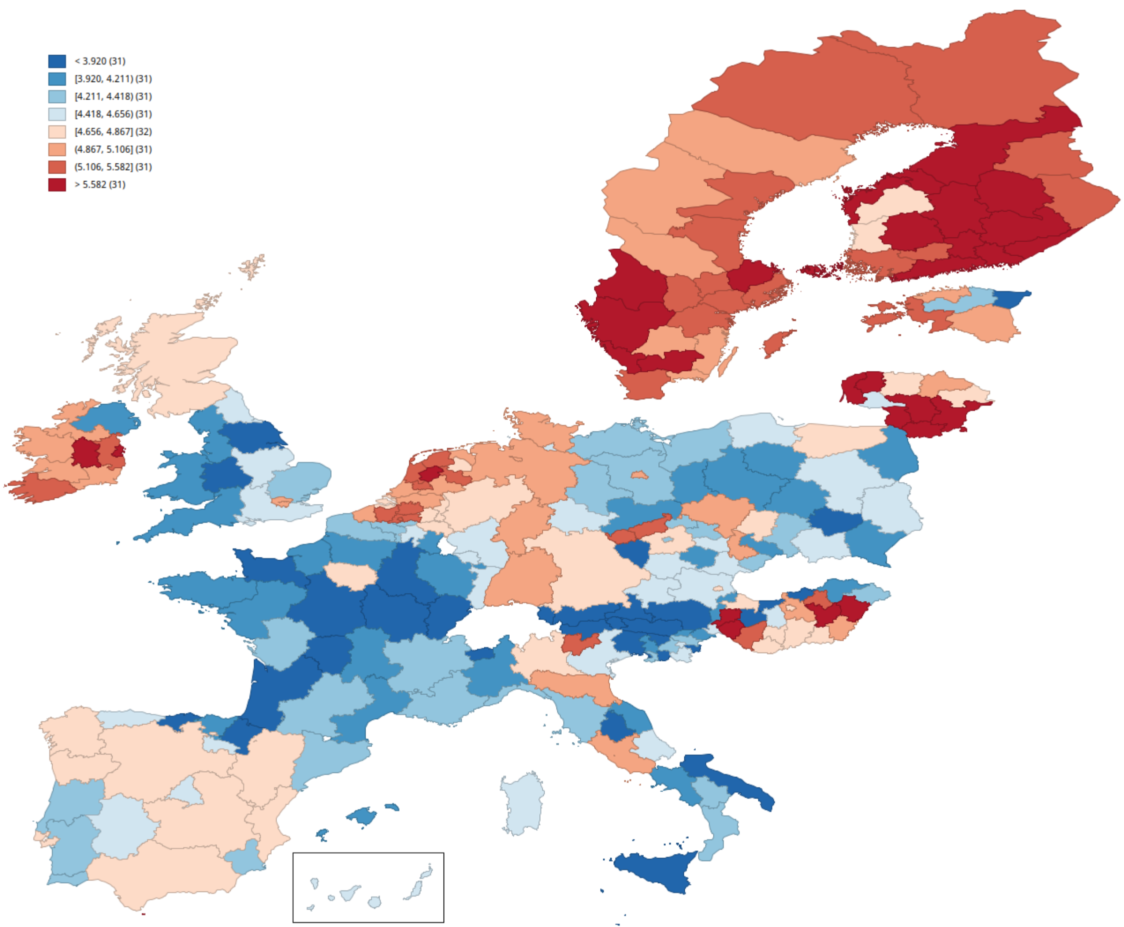


Figure 4.9: Posterior latent mean level of SPT in the NUTS BMGCFA scalar invariance model.

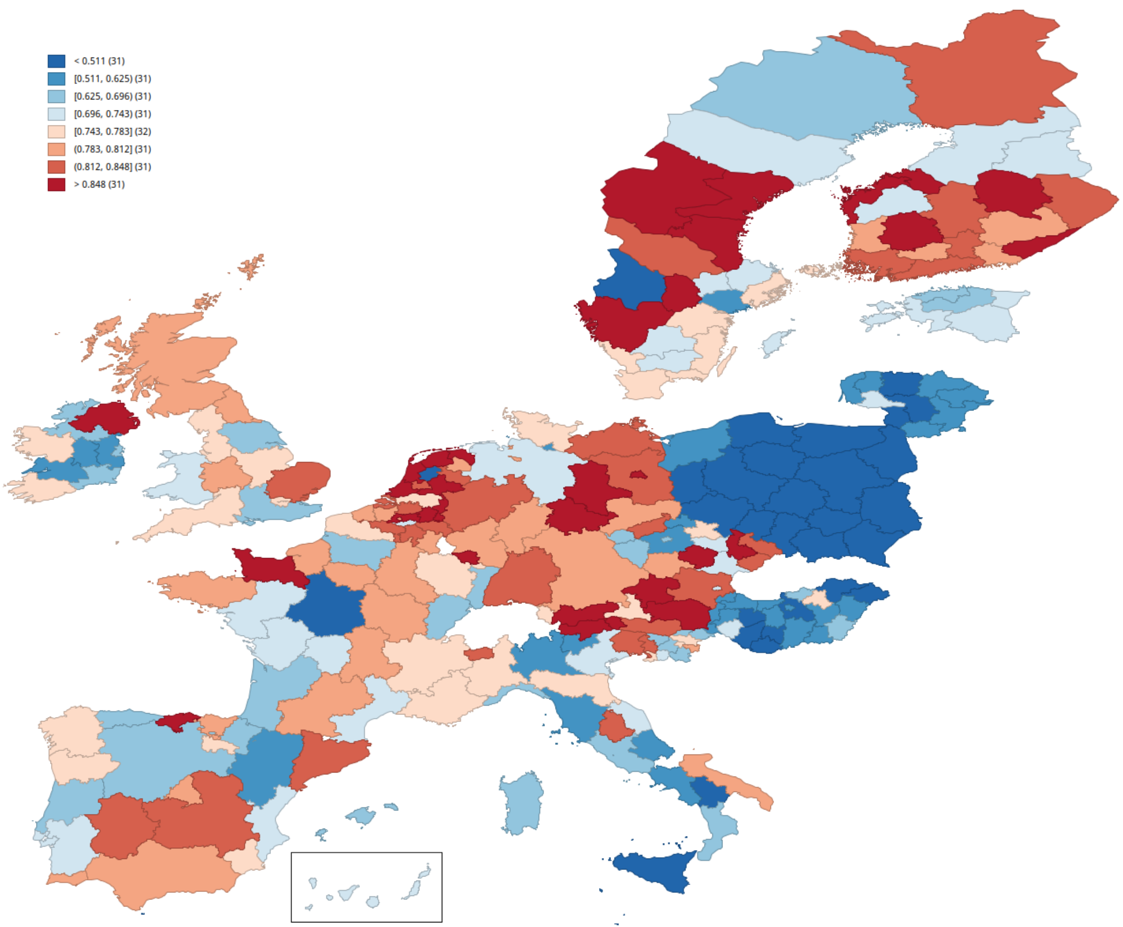


Figure 4.10: Posterior latent variable correlation between NPT and SPT in the NUTS BMGCFA scalar invariance model.

for example the United Kingdom, Ireland, Austria, Estonia and Hungary is wider, indicating regional-level differences. By modelling a NUTS-level model, we are able to catch these within-country differences, even though they may not be notable across all countries.

4.5.2 NUTS-level differences in SPT

Looking at figure 4.9, the map indicates more notable differences within several countries. On average, the level of SPT is higher than NPT. While the Nordic and Benelux countries have consistently high levels of SPT, the majority of countries exhibit differences which we expect are due to the unique characteristics of the region and/or country in question. For example, regions in West Germany have a higher level of SPT compared to East Germany, excluding the capital region of Berlin. In Estonia, the western regions have high levels of SPT while Ida-Viru County to the east have low levels of SPT. One explanation for this difference may be that Ida-Viru County population is 73% Russians and only 19% Estonians.

For the United Kingdom, we can refer back to the initial motivating example in Section 1.5. The level of SPT is in the upper half of the distribution only for Scotland and London. This is in line with the observations by Hobolt (2016) when analysing the outcome of the Brexit vote. Larger multi-cultural cities are more in favour of Remain, while countryside regions were in favour of Leave. One interesting discrepancy is the relatively low level of SPT in Northern Ireland. This highlights the fact that SPT in this study is not a direct measure of trust in the European Union as such, but entails a broader understanding of supranational political trust.

The 'capital effect', as seen in London, is also present in other countries, most notably in Germany, France and Italy. On the other hand, the effect is 'reversed' in Spain, Austria and Poland, just to name a few. It is outside the scope of this thesis to provide theoretically founded reasons for this 'effect' in each country, but we believe that it may be relevant to investigate further for researchers who are experts in the country-specific dynamics of political trust.

4.5.3 NUTS-level correlation between NPT and SPT

Figure 4.10 shows the latent correlation between NPT and SPT across all NUTS regions. First, it is noteworthy that not a single NUTS region has an estimated correlation parameter below 0. Hence, we are not evaluating if a positive correlation exists, but rather we are evaluating the magnitude of the positive relationship between NPT and SPT. Naturally, a low correlation can be due to either high level of NPT with a low level of SPT or vice-versa. The same goes for a high correlation; for example, ITC2 (Valle d'Aosta) has the lowest level of both NPT and SPT in the sample (see Table 4.2), but a correlation between NPT and SPT of 0.82.

The only countries where correlation parameters are comparable between NUTS is Poland, Estonia, Lithuania and to some extent Hungary. We are not able to provide any reason for why this pattern emerges. However, what we can say is, that the NPT-SPT 'dynamic' is not homogenous within the majority of countries.

4.6 Comparison with the frequentist model estimates

In Chapter 2, we found notable differences between countries on the mean level of NPT, SPT and the correlation between the two. However, it was necessary to free up a number of intercepts, mainly for trust in the EP and UN, to obtain reasonable fit. Hence, it was not possible to achieve full scalar invariance, but only partial scalar invariance.

In this chapter, we fitted the BMGCFA equivalent model on the same data, mainly with the purpose of highlighting NUTS-level differences. It is worthwhile to consider how the country-level models compare to the estimates obtained in Chapter 2. From the results in Chapter 3, we found that it was necessary to move to a Bayesian modelling framework to solve estimation issues, since it was not possible to estimate a full NUTS level frequentist MGCFA model at all. Only when removing NUTS regions with less than 40 respondents and manually handling Heywood cases was it possible to make frequentist maximum likelihood estimation work. The drawback obviously being that not all NUTS regions were estimated. The Bayesian BMGCFA model in this chapter not only solved the estimation issues, but also highlighted how deviations from normality is an issue that needs to be accounted for in certain NUTS regions. We recall that the introduction of a suitable prior distribution on the residual variances ensures that they are strictly positive, hence making Heywood cases impossible.

The information criteria in Table 4.1 indicate that the country level configural invariance model is preferred over both the metric and the scalar invariance model. Only through the application of Bayesian MGCFA did this become possible - again due to the fact that estimating a full frequentist NUTS level MGCFA model was not possible. However, what we can do, is to compare the parameter estimates for the frequentist country level MGCFA model and the country level BMGCFA model. We have listed the country level partial scalar BMGCFA model estimates of the latent means and correlations in Table 4.3 next to the frequentist estimates from Chapter 2.

The parameter estimates for the frequentist and Bayesian country level partial scalar invariance models are almost identical, which is to be expected given the

Country	MGCFA partial scalar			BMGCFA scalar		
	ρ_{12}	$\overline{NPT}$	$\overline{SPT}$	ρ_{12}	$\overline{NPT}$	$\overline{SPT}$
Netherlands	0.93	5.24	5.14	0.93	5.20	5.10
Belgium	0.91	4.30	4.95	0.91	4.30	4.96
Austria	0.89	4.26	4.17	0.89	4.25	4.18
France	0.87	3.26	4.17	0.87	3.27	4.18
Germany	0.87	4.48	4.65	0.87	4.46	4.65
Finland	0.86	5.12	5.79	0.87	5.09	5.62
United Kingdom	0.85	4.06	4.41	0.87	4.05	4.32
Sweden	0.84	5.18	5.46	0.86	5.14	5.33
Spain	0.84	2.97	4.56	0.84	2.99	4.57
Czech Republic	0.83	3.84	4.52	0.83	3.84	4.53
Portugal	0.80	2.99	4.58	0.80	3.01	4.59
Italy	0.78	2.62	4.21	0.78	2.66	4.34
Slovenia	0.77	2.77	4.14	0.77	2.80	4.24
Estonia	0.76	3.94	4.88	0.75	3.94	4.89
Ireland	0.74	4.01	5.31	0.73	4.01	5.31
Lithuania	0.61	3.49	5.51	0.62	3.45	5.71
Hungary	0.54	3.97	4.76	0.53	3.96	4.77
Poland	0.41	2.84	4.35	0.40	2.86	4.36

Table 4.3: Comparison of frequentist MGCFA partial scalar invariance and Bayesian MGCFA scalar invariance models at the country level.

large sample size per country and weak priors. Although we did apply a somewhat more restrictive prior on the residual variances ($HC(20,2)$ prior), it did not impact other parameter estimates to an extent where it would have any impact on model interpretation. In conclusion, we believe that it is not necessary to formulate a full BMGCFA model when not encountering any issues related to estimation, specific parameters or Heywood cases. In this particular case, the BMGCFA model took around 10 times longer to run, which is not desirable for simple models with no prior information about the distribution of the parameters.

4.7 An application on existing research: an altenative definition of political trust

We have demonstrated the extension of the model in Chapter 2 to a NUTS level analysis and the necessary considerations that needed to be made in relation to the Bayesian modelling framework. In the following, we will provide an equivalent extension to a model published in a closely related field Märien (2011). Märien (2011) measures political trust differently, namely using five questions from the European Social Survey questionnaire; trust in the county's parliament, politicians, political parties, legal system and the police. The answers are given on a 0-10 scale. The model is illustrated in Figure 4.11 . It is different to the model presented in Chapter 2 and 3, due to the one-dimensional notion of political trust and the selection of different trust measures. Hence, we have chosen the sample to not be restricted to consist of EU countries alone but will incorporate all countries in the ESS, round 8 which has a NUTS classification available.

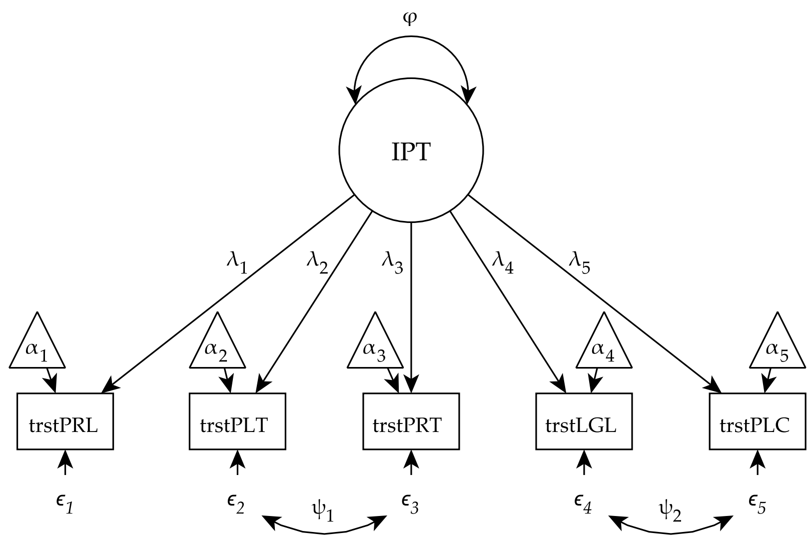


Figure 4.11: Marien (2011) measurement model of political trust.

4.7.1 Data and model definition

We will use a sub-sample of the European Social Survey, round 8 (2016) consisting of 21 European countries divided into a total of 265 regions, where only complete cases are studied (37,918 observations in total). As can be seen from Table 4.4, the sample size within each region ranges from 7 (Ceuta) to 808 (Põhja-Eesti). In most countries, capital regions have the largest sample sizes (for example Wien, Budapest, Île de France, Stockholms län and Osrednjeslovenska) while the smallest sample sizes are mostly found in the least populated regions. Apart from a few, every country contains at least one region with a sample size of less than 70.

In line with the model in Section 4.3.2 we set up the the simple measurement model $y^g = \alpha^g + \Lambda^g \eta^g + \epsilon^g$ for individual $i = 1, \ldots, N$ in group $g = 1, \ldots, G$ with effects-coding. The proposed model by Märien (2011) measures political trust on one latent variable using five observed items (that is $p = 5$ and $m = 1$). Also, the model assumes correlated residuals between trust in politicians and trust in

Country	NUTS level	Regions	Low N	Low N region	High N	High N region	SS	SS/Region
Austria	2	9	70	AT11	391	AT13	1,962	218
Belgium	2	11	47	BE34	283	BE21	1,739	158
Czech Republic	3	14	61	CZ041	275	CZ010	2,222	159
Estonia	3	5	179	EE007	808	EE001	1,935	387
Finland	3	19	14	FI200	526	FI1B1	1,908	100
France	2	21	25	FRI2	249	FRI0	2,028	97
Germany	1	16	22	DEC	488	DEA	2,775	173
Hungary	3	20	15	HU313	286	HU101	1,541	77
Iceland	3	2	350	IS002	509	IS001	859	430
Ireland	3	8	194	IE012	625	IE021	2,604	326
Italy	2	20	13	ITH2	333	ITC4	2,514	126
Lithuania	3	10	83	LT007	442	LT002	2,022	202
Netherlands	2	12	32	NL23	299	NL33	1,624	135
Norway	2	7	114	NO02	346	NO01	1,523	218
Poland	2	16	34	PL52	185	PL12	1,567	98
Portugal	2	5	42	PT15	445	PT11	1,219	244
Slovenia	3	12	36	SI018	275	SI021	1,251	104
Spain	2	18	7	ES63	362	ES61	1,820	101
Sweden	3	21	9	SE212	272	SE110	1,485	71
Switzerland	2	7	51	CH07	343	CH02	1,420	203
United Kingdom	1	12	69	UKN	256	UKD	1,900	158

Table 4.4: ESS 8 sub-sample for the Märien (2011) model.

political parties (item 2 and 3), and trust in the legal system and trust in the police (item 4 and 5). As a result, the observed residual covariance matrix is:

$$\Theta = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & \theta_{23} & 0 & 0 \\ 0 & \theta_{32} & \theta_{33} & 0 & 0 \\ 0 & 0 & 0 & \theta_{44} & \theta_{45} \\ 0 & 0 & 0 & \theta_{54} & \theta_{55} \end{bmatrix}$$

Using the same separation strategy as in Section 4.3.2, we specify the non-zero entries in Λ_D so $D \sim N_v\left(0, \Psi_D\right)$ and $\epsilon^* \sim N_p\left(0, \Theta^*\right)$ where both Ψ_D and Θ^* are diagonal. In this case, with a one-factor model, we get the following working model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} \eta_1 + \begin{bmatrix} 0 & 0 \\ \lambda_{D21} & 0 \\ \lambda_{D31} & 0 \\ 0 & \lambda_{D42} \\ 0 & \lambda_{D52} \end{bmatrix} \begin{bmatrix} D1 \\ D2 \end{bmatrix} + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \epsilon_3^* \\ \epsilon_4^* \\ \epsilon_5^* \end{bmatrix}$$

Aside from the fact that the model by Märien (2011) is a different measure of political trust, the two correlated errors leads to the introduction of two phantom-latent variables instead of one, with only one latent variable. We shall apply the same prior for the intercepts, loadings, latent means and item correlations:

$$
\begin{aligned}
\alpha_p \sim & \quad N(\mu_\alpha, \frac{1}{\tau_\alpha}) \\
\lambda_p \sim & \quad N(\mu_\lambda, \frac{1}{\tau_\lambda}) \\
\nu_p \sim & \quad N(\mu_\nu, \frac{1}{\tau_\nu}) \\
\rho_D \sim & \quad -1 + 2 \cdot \text{Beta}(a, b)
\end{aligned}
$$

4.7.2 Model fit and diagnostics

Running the configural and metric invariance models on the country-level (21 countries) and NUTS-level (265 regions) gives the DIC, WAIC and LOOIC as listed in Table 4.5.

Model	**DIC**	**WAIC**	**LOOIC**
NUTS configural	735845	731349	731537
NUTS metric	735297	731204	731366
NUTS scalar	2401609	744433	744771
Country configural	742236	734521	734522
Country metric	743231	735026	735027
Country scalar	2774776	755820	754838

Table 4.5: DIC, WAIC and LOOIC for country-level and NUTS-level models in 21 countries and 265 regions.

All models were run with 16000 iterations with burnin = 500 and thinning = 5 across two chains, resulting in 4000 saved iterations per model. We ran the model with fewer iterations than previous due to 1) the smaller number of paramters to be estimated and 2) to save computing time. The full JAGS code for the configural invariance model is available in Appendix 8. According to the information criterion, the NUTS-level models provide noticeably better fit compared to the country level model, even taking into consideration the substantially larger number of estimated parameters. We trace 273 parameters in the country-level metric invariance model and 3445 parameters in the NUTS-level metric invariance model, which entails all non-deterministic nodes in the model specification. Moving from the configural to the metric invariance model provides a better fit at the NUTS level, while the country-level metric invariance model is inferior to the configural invariance model. In other words, when we run the model on a larger number of groups, the decrease in the number of parameters outweighs the flexibility of having factor loadings vary between groups. However, the scalar invariance model at both the country and the NUTS level significantly decreases model fit, indicating that it is not possible to obtain scalar invariance.

For illustrative purposes, we are basing the following graphical illustrations on the metric invariance model. However, since we find no evidence of the scalar invariance model to hold, it is necessary to be cautious in the interpretation of the posterior mean values of the latent means.

The comparison of the country and NUTS-level scalar invariance models through the empirical distribution function (ECDF) of the posterior deviance is shown in Figure 4.12 (Aitkin 2010). As seen in Figure 4.12, the NUTS-level and the country-level ECDF look close to identical with the majority of the deviances being in the range of 15-25. However, the NUTS-level distribution is slightly left-skewed, indicating a better fitting model.

We detect potential model issues by comparing trace plots in two low-N groups. As an example, we have present the trace plots and posterior densities for the α_{1-4}

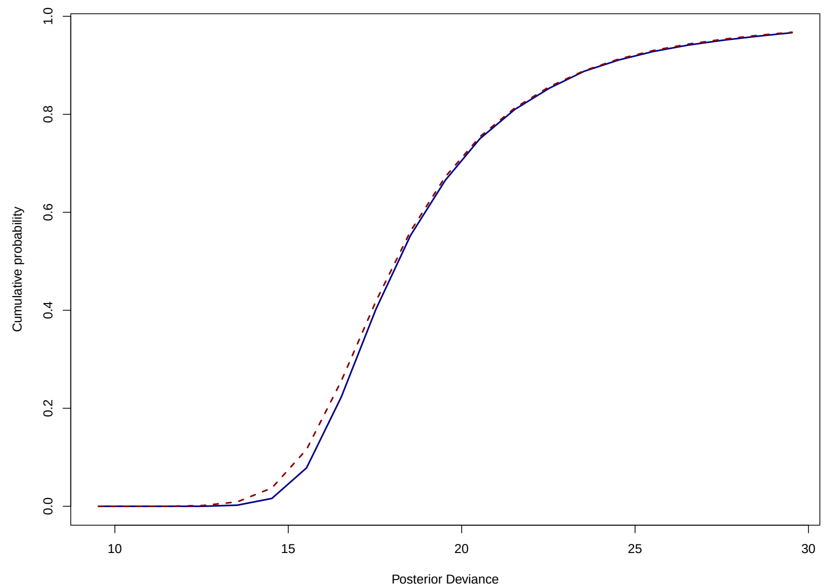


Figure 4.12: ECDF of the posterior deviances for the country-level metric invariance model (blue) and the NUTS-level metric invariance model (red).

and θ_{1-4} parameters in two groups in the NUTS scalar invariance model:

- Group A (163): ITH2 (Trento) with N = 13.
- Group B (68): ES23 (La Rioja) with N = 9.

For group A, the two chains in Figure 4.13 mix quite well. Due to the application of effects-codingα_5 is defined as a function of α_1, α_2 ,$\alpha_3$and α_4. The trace plots for group B shows equally good mixing, with very few posterior draws from the tails of the distribution.

Trace plots and posterior densities for the error variances θ are shown in Figure 4.14. Group A shows good MCMC mixing with the posterior density resembling a normal distribution around the posterior mean variance, with a reasonably short

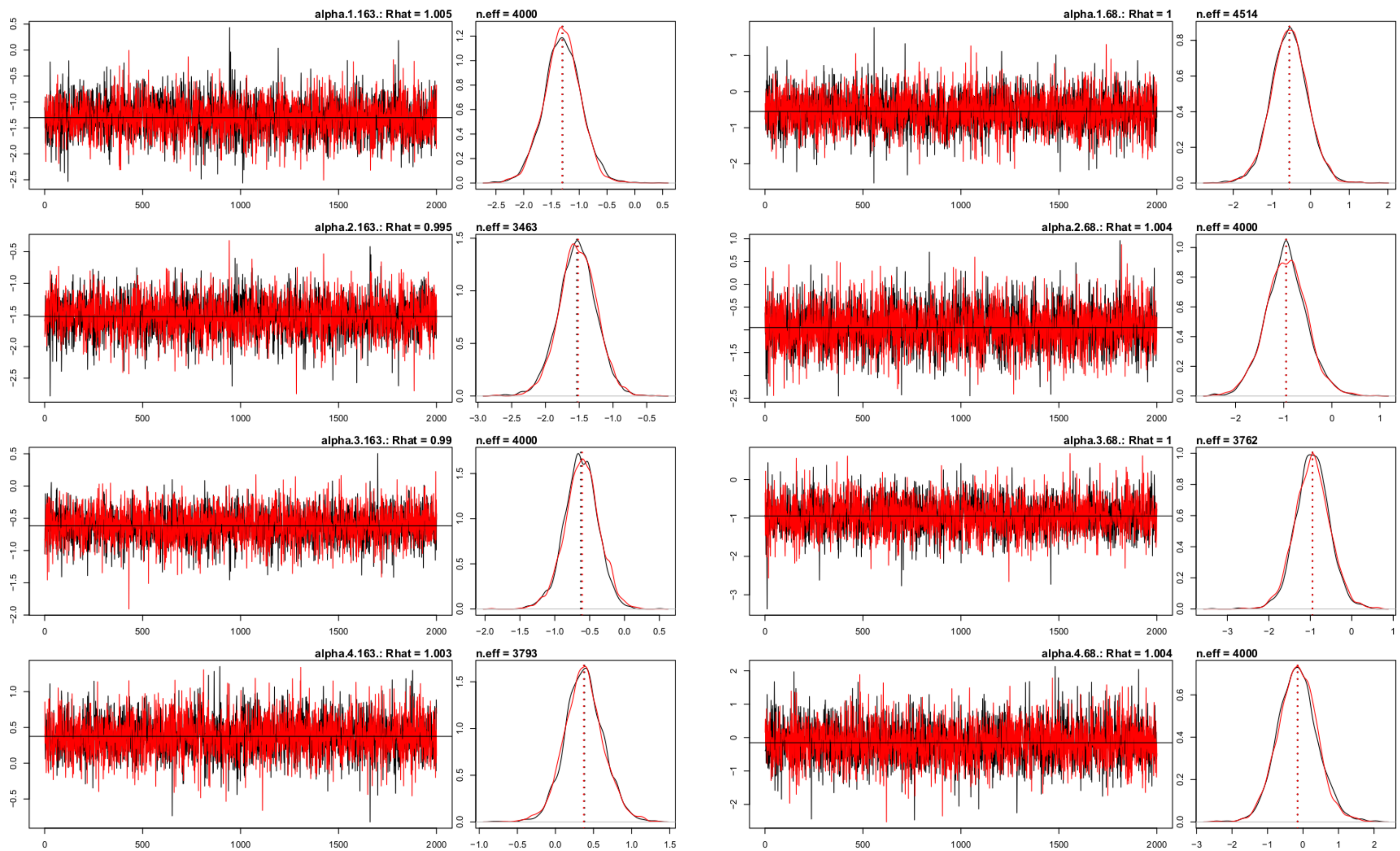


Figure 4.13: Trace plots and posterior densities for α_1-α_4 in Group A (left) and Group B (right).

tail. However, the trace plot for group B has frequent draws from the tail of the distribution and in particular θ_1 does not appear to converge in a satisfying way. This can also be identified by the relatively low effective sample size (under 500 for θ_1). One way of resolving this issue, already mentioned in Section 4.4, is to impe stronger priors on the precision/variance parameters and the precision/covariance matrices.

In Table 4.6, we summarise the mean value of the parameters of interest, their standard deviations and the highest and lowest observed values of the NUTS metric invariance model. The highest mean level of political trust is found in Trøndelag (county in the central part of Norway), while the lowest is in Valle d'Aosta (autonomous region in north-western Italy). Interestingly, Valle d'Aosta was also the region with the lowest level of both NPT and SPT in the model form section 4.2

Parameter	Mean	Std. dev.	Minimum	Min. NUTS	Maximum	Max. NUTS
ν_1	4.69	0.202	2.36	ITC2	6.76	NO06
λ_1	1.25	0.005	-	-	-	-
λ_2	1.09	0.004	-	-	-	-
λ_3	1.02	0.004	-	-	-	-
λ_4	0.98	0.005	-	-	-	-
λ_5	0.66	0.005	-	-	-	-
α_1	-1.37	0.142	-3.07	LT024	-0.18	ES63
α_2	-1.52	0.122	-2.30	ES63	-0.36	ITH1
α_3	-1.23	0.126	-2.34	ES63	-0.37	ITH1
α_4	0.70	0.155	-0.35	ES43	1.91	ITC3
α_5	3.41	0.177	1.77	HU313	4.90	ES43

Table 4.6: Summary of posterior estimates.

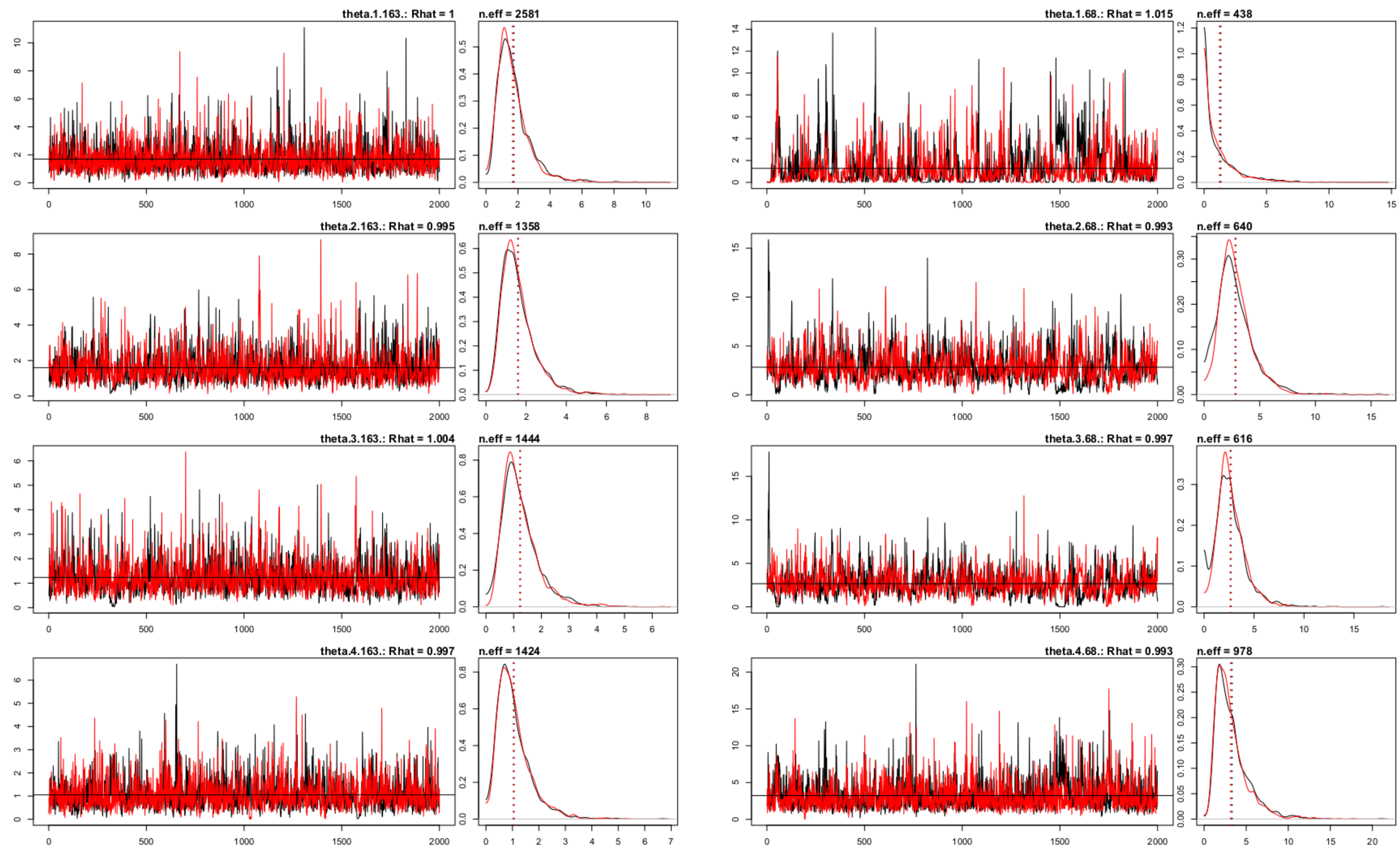


Figure 4.14: Trace plots and posterior densities for θ_1-θ_4 in Group A (left) and Group B (right).

Furthermore, the posterior means for α varies significantly. For example, the posterior mean for α_4 is -0.35 in Extremadura (autonomous community in central Spain) and 1.91 in Liguria (a region in north-western Italy). The high variability in the posterior distributions for the loadings adds to the claim that a scalar invariance model is not preferrable for this particular model.

4.7.3 Regional differences

An overview of the estimated latent means is shown in Figure 4.15, which is based on octiles of the distribution of latent means across all NUTS regions. In line with Märien (2011), we find that the posterior latent mean is high in the Nordic countries, while it is low in Southern Europe. In addition, all regions in the Netherlands, Belgium and Switzerland belong to the upper half of the distribution of NUTS regions. However, with the addition of NUTS-level detail, we observe additional within-country differences.

First, in most countries, except Norway, we observe substantial within-country variability (differences larger than 0.5 on the scale) in the level of political trust. This is most pronounced in Hungary, Lithuania and Germany. Second, some countries are clearly divided into clusters of regions with high/low levels of political trust: Italy (north/south) and Germany (east/west), while some countries contain more randomly scattered regions. Third, it appears that some countries have notably higher levels of political trust in the capital regions, compared to the surrounding regions. This is most obvious in France, Germany, Sweden, Iceland and Ireland. However, the opposite is the case for Austria, the United Kingdom and the Czech Republic, where the mean level of political trust in the capital regions are notably

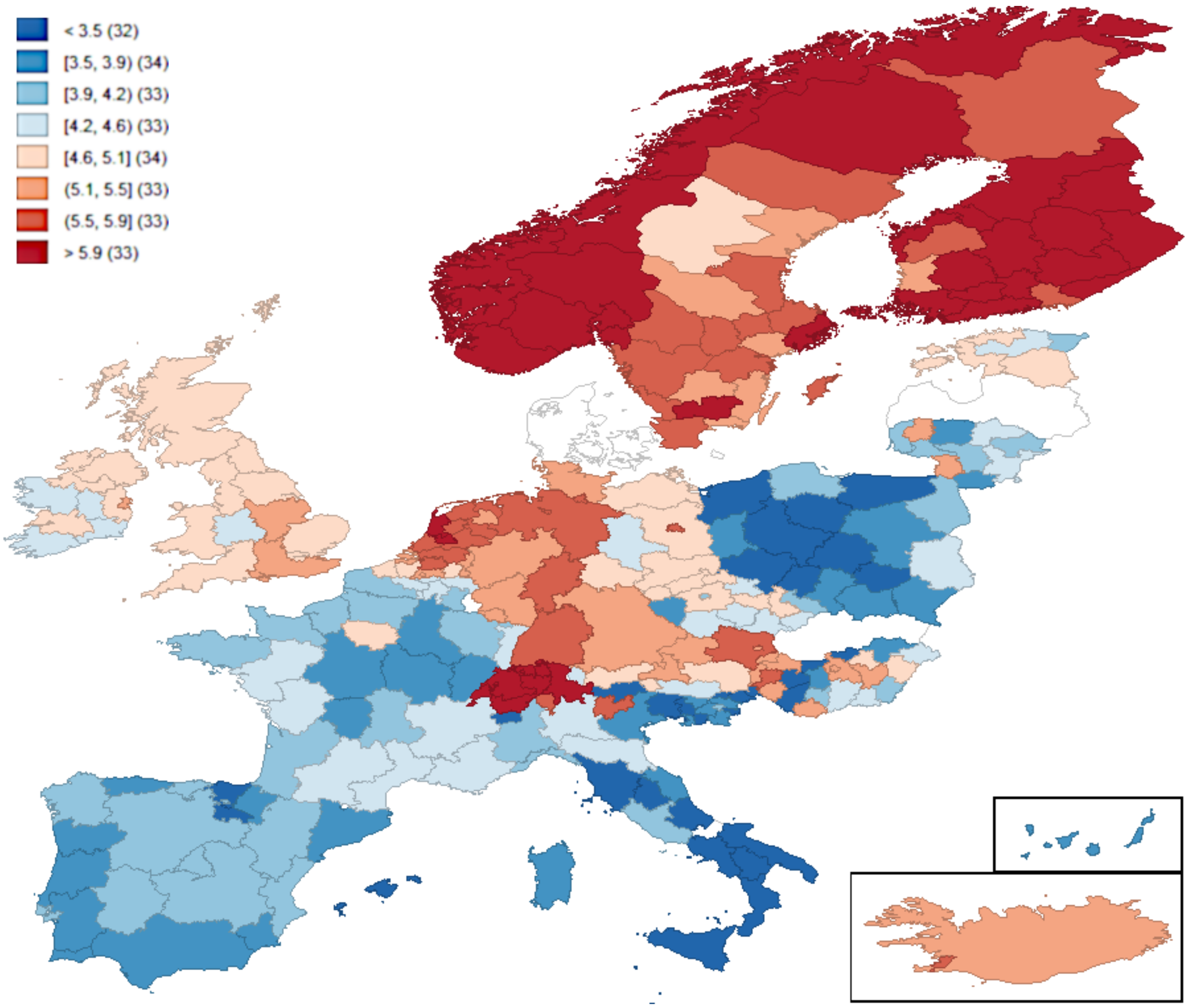


Figure 4.15: Posterior latent mean level of political trust in the NUTS BMGCFA metric invariance model.

lower than the surrounding regions.

The original study by Märien (2011) states that:

> "From the analyses we can conclude that the 'institutional trust' scale is configural equivalent: a similar pattern of factor loadings emerged. Therefore, institutional trust can be meaningfully discussed in the different countries. The factor loadings are not entirely the same in all countries. [...] Therefore, some of the equality constraints had to be removed. After these modifications, the constrained model provides a good fit of the data." (Märien 2011)

She follows a standard approach to establishing measurement invariance, where partial invariance is obtained by relaxing equality constraints as previously discussed in Chapter 2. However, it was not possible to obtain full metric invariance in the original study at the country level. With the BMGCFA approach, however, we not only find that the NUTS level model provides a better fit compared to the country level model, but also that the metric invariance model is preferred over the configural invariance model when analysing the measurement model at the NUTS level. This is not the case on the country level, where the configural invariance model is indeed the best fitting model, which is in line with the findings by Märien (2011). Furthermore, we are able to challenge some of the more substantive statements made in the original study:

> "Therefore, to date the only conclusion that can be drawn is that institutional trust is stable (especially in the more established democracies).

These findings underline the importance of a good operationalisation of political trust and of the study of political trust beyond the United States. [...] Not only is political trust lower in newer democracies, it is also more volatile. Nevertheless, the analyses revealed that institutional trust in new democracies could also be measured in a valid manner. A one-dimensional attitude, institutional trust, exists in all countries under study providing strong support for the claim that institutional trust reflects an assessment of the prevailing political culture in a country." (Märien 2011)

While Märien (2011) does indeed analyse different data and mostly focus on changes to the concept of trust over time, it is important to consider these statements in light of the results shown above. While the majority of the countries may have a stable aggregated level of political trust over time, it does not account for the notable within-country variability, which becomes apparent when analysing the data at the NUTS level. And as pointed out in Chapter 3, it is not uncommon to see NUTS level changes in the level of political trust over time, even though the country mean level may remain relatively stable.

4.7.4 Concluding remarks

The differences in the level of political trust between NUTS regions is potentially highly relevant for the literature on political trust. Not only does it add to the claim that a more detailed analysis is needed to fully understand how political trust as a measure varies across the European landscape, but it also provides the researcher with a better fitting model, which is desirable. However, it requires that one is careful in the modelling, especially when setting the priors for the more sensitive parameters.

But why use MGCFA when alternative methods, such as multilevel models are available and able to better handle a large number of groups? The key argument here is measurement invariance. Although it may be possible to set up a multilevel model which aims to measure political trust across NUTS regions in Europe, it is not possible to determine whether the concept is measurement invariant across said regions, potentially leading to severe misinterpretations.

In this chapter, we have presented two different models: one based on the initial country level model in Chapter 2, and one based on a published model by Märien (2011). We discovered in Chapter 3 that the conventional maximum likelihood estimator in the frequentist framework is not able to estimate a NUTS level model with many small regions, and as a result, we presented a Bayesian modelling framework to overcome this issue. The conclusions drawn from the two models fitted in this chapter are very much similar. We have discovered that a NUTS level model provides a starting point for a more detailed theoretical discussion about regional level differences in political trust. In addition, we have identified a pathway for dealing with problematic parameters, such as the residual variance, which is a direct effect of small-N groups.

Chapter 5

Conclusion and future work

In this thesis, we have been investigating how to measure and analyse political trust across countries and regions in Europe. In the following, we will present chapter-by-chapter conclusions. Next, we will return to the thesis research questions, present key learning and present how this thesis fits into existing research on political trust, measurement invariance, and Bayesian methods. Finally, we will outline potential avenues for future research, focusing on a Bayesian approach to modelling spatial correlation across NUTS regions using CAR priors.

5.1 Chapter conclusions

5.1.1 Chapter 2

In Chapter 2, after having established a theoretical framework, we set up a measurement model for political trust to establish measurement invariance across EU countries in ESS 8. Although it was not possible to obtain full scalar invariance, the partial invariance model had a good fit. Trust in the European Parliament was non-invariant in some countries, indicating that the level of trust in the European

Parliament and the UN is different for invariant and non-invariant countries given the same latent mean level of supranational political trust. However, the non-invariant countries did not seem to share common characteristics. If it was possible to obtain scalar invariance, we would conclude that mean values on both latent variables would correspond to a unique value on each item, which would be the same across all countries. That being said, we did carry out a sensitivity test and found no differences between the metric and the partial invariance models that would change our substantive conclusions. The main findings were 1) national and supranational political trust can be regarded as two different empirical concepts, 2) national and supranational political trust is highly correlated in most countries. This is relevant for future studies on the concept of political trust, be it on the national or supranational level, especially in the light of recent debates and events in EU politics.

5.1.2 Chapter 3

Chapter 3 pursued the idea that countries are not homogenous across regions when it comes to measuring a concept like political trust. However, translating the simple measurement model from the country level to the NUTS level is no simple task. We investigated how the technical challenges of coding NUTS regions may be overcome. Regarding the estimation procedure, we find that it is 1) not possible to estimate the measurement model using the frequentist maximum likelihood estimator with many groups and that 2) we encounter several Heywood cases, even when only using a subset of the original data. Heywood cases are identified when, for example, the residual variance of a given item is estimated to be negative. However, finding cases within larger N groups could point in the direction of, for example, mis-specification. Finally, we test the idea of the items used in the measurement model being non-normal (floor effect), and hence influencing the parameter estimates and substantive conclusions. This would allow us to study phenomena such as those discussed in Section 1.5. Using simple imputation on the 0 values, we find some improvement in model fit, although they may not be large enough to justify further investigation.

5.1.3 Chapter 4

In Chapter 4, we suggest a solution to the estimation issues highlighted in Chapter 3 by setting up a Bayesian MGCFA model with phantom-latent variables on the model from Chapter 2. We discuss and provide a Bayesian model framework, where the focus is on 1) the choice of prior distributions, especially for the residual variance parameters, 2) model comparison and diagnostics and 3) changes to the substantive interpretation of the estimated model. Through successfully fitting a NUTS level model, we rule out that the initial estimation issues were due to model misspecification.

From a substantive point of view, we find that the variability within countries on the two trust measures is notable in most countries, in particular for SPT. This gets to show how level of trust is not only different between countries, but also between regions within countries. Looking at the estimated latent means across NUTS regions, the findings are very much in line with the theoretical arguments put forward by both Mishler and Rose (2001) and Anderson et al. (2005). Although variability exists within the Nordic countries, the level of NPT and SPT is, in general, at the top of the distribution of NUTS regions - something that Delhey and Newton (2005) argue is due to the historical development of the democratic institutions. At the same time, the level of NPT is low in the Southern and Eastern European countries. Overall, NPT appears to be more homogenous within countries than SPT.

One of the key contributions in Chapter 4, is the choice of the half-t distribution as a prior for the residual standard deviations, which we argue is the best choice for the model. This is based on the work by Gelman (2006), who points out how assigning a common InverseGamma $(\varepsilon, \varepsilon)$ prior can lead to issues when the variance is small - something that we confirm in our analysis. In the resulting analysis, we are able to obtain satisfying convergence, even for small-N NUTS regions.

For the NUTS level models, we find that the scalar invariance model is equivalent to the configural invariance model. These conclusions are made based on several model fit indices, as well as through comparing models using the empirical distribution function of the posterior deviances. The latter is inspired by Aitkin (2010) and it is, to our knowledge, the first time it has been applied to the Bayesian measurement model framework.

For the country level models in Chapter 4, we compare the resulting scalar invariance model with the partial scalar invariance model in Chapter 2. For all countries, the latent means and latent variable correlations are equivalent, showing no sign of changes to the substantive conclusions made in Chapter 2. This is to be expected, given that the sample size within each country is sufficiently large and the assigned priors having close to no impact on the posterior distributions. In essence, while we recommend the use of a NUTS level model if possible, fitting a country level model using conventional frequentist methods is less computationally demanding and easier to fit.

At the end of Chapter 4, we fit an equivalent NUTS level model to an already published study by Märien (2011). When comparing the country and NUTS level models, we find that the NUTS level model not only fits better than the country level model but that the metric invariance model fits data better than the configural invariance model. For both the country and the NUTS level model, it is not possible to obtain scalar invariance, however. Furthermore, we find that only a few countries are internally homogenous in terms of political trust (Norway, Netherlands, Finland, Belgium and Switzerland) while most countries appear more heterogeneous. The adds to the claim that political trust is best investigated at the regional level.

5.2 Returning to the research objectives

Throughout the chapters in this thesis, we have touched upon multiple topics to provide the foundation for answering the initial research questions;

1. How can we measure political trust in a large cross-national survey, while ensuring that it is measurement invariant?
2. What are the current issues with the measurement invariance method when extended to many groups or groups with low sample size?
3. How can we develop a framework for testing measurement invariance when conventional methods do not work?

Throughout the chapters outlined above, we have been working towards answering these questions. To summarise, Chapter 2 deals with the definition of the core theoretical and methodological concepts of the thesis, namely political trust and measurement modelling and invariance. The resulting country-level model yields some interesting results, which aligns with existing literature. Based on the illustrative example from Chapter 1, we argue that moving to a NUTS-level model is important for our understanding of political trust. Chapter 3 then identifies core issues when estimating a NUTS-level model, namely that the model cannot be estimated using conventional frequentist methods. Chapter 4 solves this particular issue by introducing a fully Bayesian multiple groups confirmatory factor analysis framework.

In the following, concluding remarks for each of the research questions are provided, alongside considerations on how it fits into existing research that we have been referring to throughout the thesis. It is important to note that theoretical research on political trust, empirical research using measurement modelling and invariance along with the application of Bayesian methods in the social sciences are rapidly developing areas of research. The findings and recommendations in this thesis are aimed to fill the gap that lies in the intersection of these concepts in order to strengthen future research.

5.2.1 Measuring political trust in large cross-national surveys

The first step to answer how political trust should be measured is to settle on a clear definition of political trust. In both the theoretical literature and empirical studies, political trust has been defined and operationalised differently. We started by choosing a definition that aligns with Sztompka (2003), with political trust bring 'the citizens' belief in institutionalised practices and procedures'. This is not a theoretical definition that necessarily collides with alternative definitions (e.g. evaluation of government performance or attitudes towards political institutions and leaders).

The second step is to operationalise political trust to some measurable outcome using some survey data of choice. We chose the rather popular ESS dataset throughout the thesis, defining two different dimensions of political trust; National Political Trust (trust in politicians, trust in parties and trust in the national parliament) and Supranational Political Trust (trust in the European Parliament and trust in the United Nations). This was very much aligned with André (2013) who arrived at the same two dimensions (alongside a third dimension for order/neutral political trust). Just as in a range of existing research referred to in this thesis (Byrne, Shavelson, and Muthén 1989; Davidov et al. 2014; Schneider 2016; Zercher et al. 2015), also André (2013) is not able to obtain full scalar invariance between countries. As has been pointed out previously, it may have many explanations (Brown et al. 2015; Byrne, Shavelson, and Muthén 1989; Chen 2007). In our case, several of the intercepts for trust in the EP/UN are non-invariant. However, the partial scalar invariance provides a good fit. Most importantly, we find that the actual parameter estimates vary only slightly between the full scalar and partial scalar invariance models, reiterating that measurement invariance methods should be interpreted in context. In other words, it is not sufficient to evaluate models based on model fit indices alone. The final model very much agrees with previous research; we find a positive correlation between national and supranational political trust (Arnold, Sapir, and Zapryanova 2012; Muñoz, Torcal, and Bonet 2011), high levels of political trust in the Nordic and Benelux countries (Delhey and Newton 2005) and low levels of NPT in the Southern European countries (Anderson et al. 2005).

While political trust is a main theoretical component of the thesis, the process of measuring a theoretical concept presented in this thesis using large-scale survey data, such as the ESS, extends to other research areas as well. The standard measurement invariance methods is a useful framework when applied to large sample size surveys in a cross-country setting, but does enter complications when the sample size is small or the number of groups is large.

5.2.2 Measurement invariance: issues and solutions

In cases when we have a large number of groups and/or the sample sizes per group is small, we may encounter issues with the conventional measurement modelling framework. This is the theme of Chapter 3, where we move from a country-level model to a NUTS-level model using the same ESS data as in Chapter 2 (including previous waves to highlight coding issues for specific NUTS regions over time as well).

The main illustrative finding is, that it is not possible to fit a frequentist model using the highest level of granularity of NUTS regions due to the sample size of certain regions being too small. This can be attributed to the so-called Heywood cases (Gagne and Hancock 2006), which are thoroughly discussed in both Chapter 3 and 4. In addition, the analysis on the sensitivity to extreme cases in Section 3.4.1 reveals that the items used from ESS dataset has a feature which complicates the use of fit statistics, namely departure from normality. When moving from a country to a NUTS level model, we only expect this issue to get worse; extreme cases only get more problematic when the group-level sample size is small. Taking into account the effect that the imputation in Section 3.4.1 has on model fit indices, we suggest to further investigate the relationship between deviations from normality and the occurrence of Heywood cases when modelling small-N groups in a frequentist maximum likelihood setting.

Although it is not possible to fit a full frequentist model using all NUTS regions, the ESS subset does indicate potentially interesting differences in both NPT and SPT within countries. This strengthens our initial claim that political is carries useful information for the researcher to investigate the political trust concept at the regional level - even with many countries, the regional differences should not be ignored. However, an expansion of the modelling framework is necessary. The next step, which is introduced in Chapter 4, is to build an MGCFA model to handle the regional-level model equivalent of the country-level frequentist MGCFA model presented in Chapter 2. We do not find many studies that go entirely Bayesian in the measurement modelling/invariance literature within the social sciences. The closest we get is the work of Muthén and Asparouhov (2012), McNeish (2016) and Merkle and Rosseel (2018). The former is part of the development team of the Mplus software, while the latter has developed the 'blavaan' package under R. Both of these software packages provide tools to fit a purely Bayesian structural equation model (which includes measurement models). Hence, the tools are available for the researcher, just rarely used.

In the transition from the frequentist model to the Bayesian model, we dive into the selection of priors, in particular on the residual variance parameters of the model. Many of the software packages use the gamma prior, which we find insufficient for this particular study, resulting in the application of the half-t distribution instead, inspired by Gelman (2006). The additional modelling complexity encountered throughout Chapter 4 is to be expected when moving to a model which is much more detailed than a country-level model. It may also be one reason why these extended models are not encountered too often in the literature, although we do not know the reasoning for the choice of the researcher. The results of the regional-level model add to the claim that political trust is variable across regions, just as we found with the limited model in Chapter 3. Even when estimating an additional model based on the book chapter by Märien (2011), we provide evidence that a regional-level model provides additional insights that were not detected in the first place. This includes large within-country variability in more than three countries, clustering of high/low trust regions within countries and a potential capital effect to be further investigated.

As a result, the outcome of Chapter 4, and the thesis as a whole, not only provides a solution to common methodological issues for measurement invariance methods in special cases like a regional-level model with a small sample size but also adds to our understanding of political trust; both country-level and regional-level variability in national as well as supranational political trust exists. While the findings on the country-level differences are largely equivalent to existing literature, the regional-level variability does contribute to our understanding of why we observe differences as presented in 1.5. We will leave the question of why to be explored in future research.

5.3 Spatial correlation across NUTS regions

Since one of the key outcomes of this thesis is to set up a working model and establish measurement invariance using a Bayesian framework on many groups, it is only natural to consider potential routes for future research within the field. One of the obvious extensions is to incorporate some of the NUTS-level covariates that have already been suggested to have an impact on the individuals' level of political trust. However, this would no longer be a measurement model but open up the possibility of making use of the whole range of structural equation modelling tools. This includes including population density, immigration, and rural/urban areas and socio-economic status as indicators (Schoene 2016; Rustenbach 2010;

Iacono 2019). In addition, we can further extend the model at the NUTS level. Since we have found strong support for NUTS level variation in many countries, we are convinced that this more detailed level of analysis should be the main approach when analysing complex theoretical concepts in a cross-cultural setting. This includes applying a BMGCFA model on the two-factor model proposed in Chapter 2, which would further inform the substantive conclusions.

However, if we acknowledge the importance of spatial factors, we should also account for standard spatial effects. In particular, the measurement model fitted in the present thesis is agnostic to whether two NUTS-regions are bordering each other or on opposite sides of the country. This is problematic for two interrelated issues. Empirically we saw that some regions tend to cluster together in ways that can only partly be explained by region characteristics. Secondly, while NUTS is a more fine-grained spatial resolution than country, the regional borders are to some extent arbitrary. This relates to the modifiable area unit problem (MAUP) (Fotheringham and Wong 1991), namely that through drawing different boundaries we may come up with different conclusions seeing as the aggregations of observations will differ. There may also be substantive reasons to explain why individuals that live close to each other tend to be more similar than people that live far from each other.

Chapter 4 demonstrates that there is variance across NUTS regions. However, in this setting, NUTS regions are treated as independent units. One way to investigate whether there is a spatial correlation between NUTS regions on the measure of interest is to calculate Moran's I (Moran 1950). Let us call the predicted residual of individual i in NUTS region j for ϵ_{ij}. We can then define the NUTS-level residual as

$$\hat{\hat{\epsilon}}_j = \frac{1}{N} \sum_i \hat{\epsilon}_{ij}$$

We define an element $W_{j\ell}^{[t]}$ of the contiguity matrix $\mathbf{W}^{[t]}$ as

$$W_{j\ell}^{[t]} = \begin{cases} 1 & \text{if NUTS } j \text{ is at distance } t \text{ from NUTS } \ell \\ 0 & \text{otherwise} \end{cases}$$

For each NUTS j we can calculate Moran's I (Moran 1950) at lag 1 ($t = 1$) as

$$I = \frac{N}{\sum_{j,\ell} W_{j\ell}^{[t]}} \frac{\sum W_{j\ell}^{[t]} \left(\hat{\epsilon}_j - \hat{\hat{\epsilon}}\right) \left(\hat{\epsilon}_\ell - \hat{\hat{\epsilon}}\right)}{\sum W_{j\ell}^{[t]} \left(\hat{\epsilon}_j - \hat{\hat{\epsilon}}\right)^2}$$

For lag t, the distance t contiguity is used instead. We can think of the NUTS regions as a graph, represented in Figure 5.1. The graph drawn with the NUTS regions used in the Märien (2011) model. The graph is based on a queen contiguity matrix, where region i is considered to neighbour region j if the polygon of the two regions share a border or have one common point. This is in contrast to the more strict 'rook contiguity', where region i and region j are only considered neighbours if their respective polygons share a border (Anselin and Rey 2014).

Using the posterior means from Chapter 4, we can calculate the posthoc Anselin Local Moran's I for the Märien (2011) model and plot it as shown in Figure 5.2 (Anselin 2010). For a given region i, we can calculate a z_{I_i}-score and a pseudo p-value to determine whether region i is bordering regions with significantly high or low value on political trust, compared to the overall mean. For more details on the calculations behind these values, see Anselin (2010). High-High regions are characterised by belonging to a cluster of regions with high values of political trust, where Low-Low regions are clustered with regions primarily surrounded by

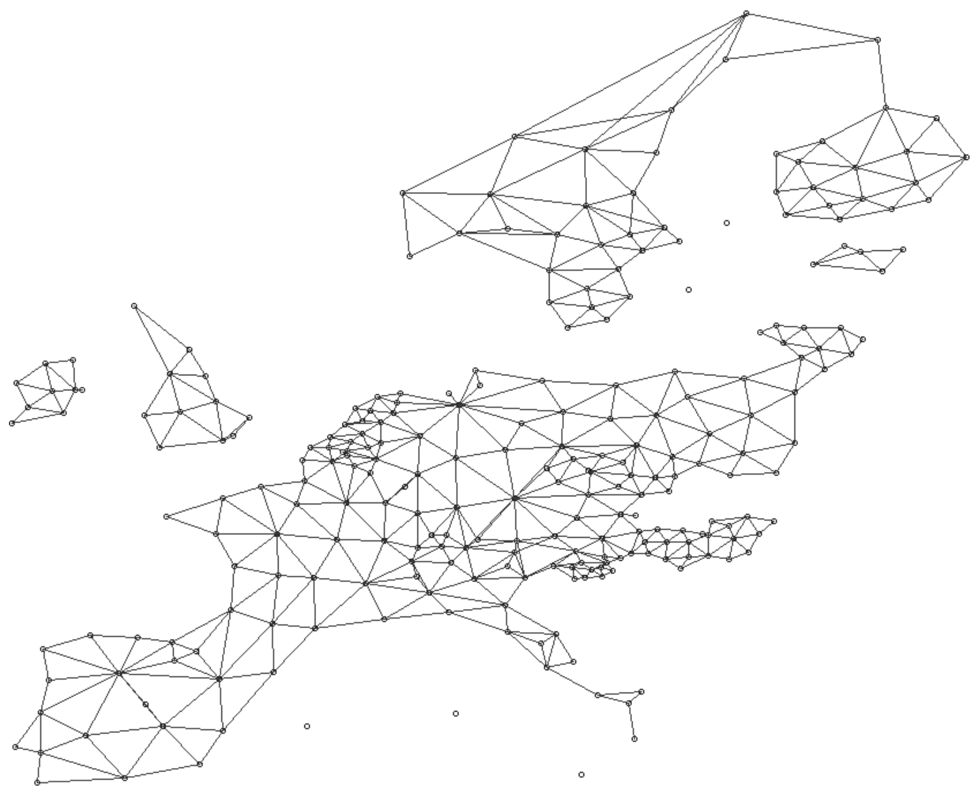


Figure 5.1: Connectivity graph of NUTS regions in ESS 8.

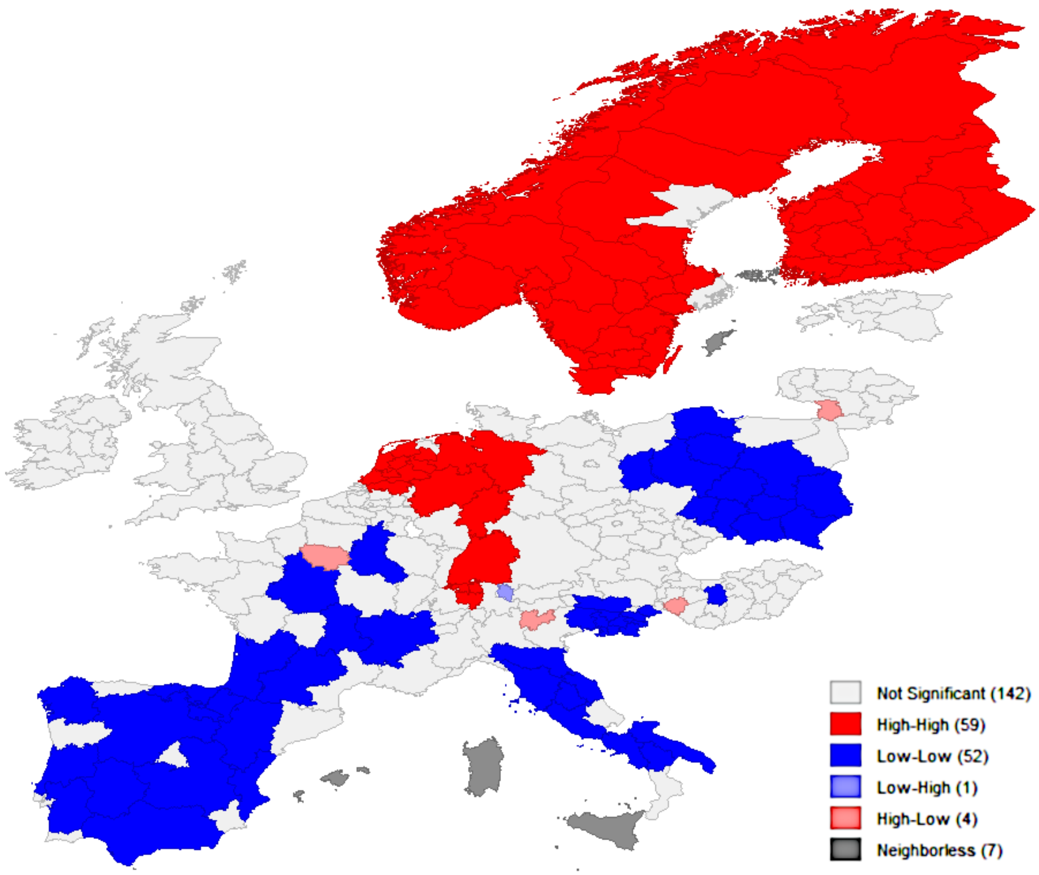


Figure 5.2: Anselin Local Moran's I for posterior means of political trust across NUTS regions.

regions with low levels of political trust.

High-Low regions are outliers with high values of political trust surrounded by regions with low levels of political trust. In this example, those regions are FR1 (Île-de-France, France), ITH2 (Trento, Italy), HU223 (Zala, Hungary) and LT004 (Marijampolė, Lithuania). Equivalently, Low-High regions (here only AT34, Vorarlberg, Austria) are outliers with low values of political trust and surrounded mostly by regions with high levels of political trust. In line with the conclusions from Chapter 4, NUTS regions in the Nordic countries, Benelux countries, Switzerland and western Germany form a cluster characterised by high levels of political trust and high spatial correlation with neighbouring regions at lag 1. Also, Spain, southern France, most of Italy, Slovenia and most of Poland form a cluster of regions characterised by low levels of political trust and high spatial correlation at lag 1. The capital region of France being a High-Low region is an interesting case, since it supports our notion in Section 4.5 that capital regions tend to exhibit higher levels of political trust compared to its neighbouring regions.

This post-hoc analysis indicates that there is residual variance or correlation to explain stemming from the spatial embedding of the observed units. To account for these spatial effects and to parse out NUTS-specific effects, such as capital regions, from spatial correlation and spill-over, a fully Bayesian, spatial multiple groups confirmatory factor analysis is needed. We outline the main challenges of this modelling framework next.

5.4 A Bayesian approach to spatial modelling using CAR priors

Accounting for spatial effects in the MGCFA could be done in a number of ways and on different levels. It is convenient to allow for spatial correlation for the NUTS-level intercepts. When setting up the Bayesian MGCFA model, we can take into account the spatial autocorrelation by assuming, in addition to spatial prior for the intercepts $\left(\alpha_p \sim N\left(\mu_\alpha, \frac{1}{\tau_\alpha}\right)\right)$, a spatial error that is conditional for each NUTS j

$$S_j \mid S_{i \neq j} \sim N\left(\bar{S}_j, \frac{\sigma^2}{n_j}\right)$$

where $\bar{S}_j$ is the avarage of S_i for the n_j neighbours of j. The un-conditional model for $S_1, \ldots, S_N$ is achieved by assuming the so-called CAR prior (Besag, York, and Mollie 1991):

$$S_j \sim CAR\left(\mathbf{W}^{[t]}, \sigma^2\right)$$

Using this prior, in general, the conditional intercept for a NUTS becomes $\alpha_j + S_j$. Fitting the spatial model requires that you carefully balance the NUTS-level random effects variance and the unconditional variance of the spatial component (Banerjee, Gelfand, and Carlin 2003; Carlin and Pérez 2000). As we saw from Chapter 4, the small-N issues with the ESS requires that we chose non-standard priors for the variance component. By modelling the spatial autocorrelation, we would allow contiguous NUTS to borrow strength from each other, somewhat alleviating some of the small-N issues. A full Bayesian estimation of a MGCFA model with an additional CAR prior would entail a thorough investigation of the joint effects of the prior specification of the two variance components to achieve this balance. An added challenge is that the CAR prior is defined conditionally since there is no closed form expression for the marginal, unconditional prior.

This thesis started out trying to establish measurement invariance in political trust across countries in the standard maximum likelihood framework. For this approach, the definition and testing of different forms of invariance is well known. For a measurement model across NUTS-level regions where you, in addition, afford correlations across regions, defining and establishing invariance is not as straightforward and future research requires that you apply similar considerations to those of Shi et al. (2017).

5.5 Other approaches

In Chapter 3, we presented a brief analysis of a subset of the ESS data. Through eyeballing the trends on the maps in Figure 3.2 to 3.5, we found that neither NPT nor SPT seems to be stable over time. However, the modelling procedure was not rigid in the sense that we used maximum likelihood estimation on a subset of data (since it would otherwise not converge) and each ESS round was modelled independently. If the model were to be extended in future research through the Bayesian approach presented in Chapter 4, the modelling procedure would become increasingly more complicated. First, the number of groups in the BMGCFA model would be a multiplicative of the number of ESS rounds used, since every group would now refer to a NUTS-region-year rather than a NUTS region. In our analysis with 249 NUTS regions, this would extend to 996 NUTS-region-years across four ESS rounds, which becomes a huge computational task. Second, it would be highly relevant to incorporate a longitudinal element directly into the measurement model, something which has not been explored to the same extent as the unit-years approach (Zercher et al. 2015).

Next, we propose to make a comparison between the 'approximate measurement invariance' method (mentioned in Section 4.2) to the full BMGCFA method in Chapter 4. More specifically, we would be interested in investigating whether one or the other method is preferred given specific types of models and data. While the approximate measurement invariance method only applies a zero-mean small-variance prior on the difference between group intercepts/loadings/error terms, this difference is not incorporated in the BMGCFA model, since the intercepts/loadings/error terms themselves are already modelled with a prior and posterior distribution. It remains to be seen how the introduction of a prior for the between-group differences would affect the performance of other prior specifications (Zercher et al. 2015; Shi et al. 2017).

5.6 The future of political trust?

In this thesis, we found that national and supranational political trust can indeed be considered two empirically distinct concepts. However, they are highly correlated across most countries in the EU. With the historical process of EU integration, we might begin questioning the importance of the nation-state in regards to political trust and ask if regional differences are becoming more important. Is the regional variation in the degree of political trust related to the country-level variation in political trust within the EU? Our final model in Chapter 4, which is an extension to an already established measurement model by Märien (2011), shows noticeable differences between regions; differences which surpass those between countries.

But what impact does these regional differences have on our understanding of political trust in times of populism? As described in the motivating example in Section 1.5, the Brexit referendum is an exellent example of how political trust, manifested through a referendum, is an aggregate of multiple within-country differences across regions. Within those regions we might experience what Kriesi (2014) describes as 'protest populism'. In regions with high levels of political trust, populist parties may considered a reliable political alternative, whereas they in regions with low levels of political trust would align more with a demand for changes to the political system (Kriesi 2014). In this regard, researchers within the fields of political trust and populism must take the regional context seriously.

To this extent, we propose that even large-scale surveys, like the ESS, should be conducted with representativity at the regional level in mind by, for example, including region-specific questions to explore the everchaging nature of political trust across both space and time. In other words, we believe that an analysis at the regional level is needed to fully understand the dynamics of political trust - not only to ensure the validity of the measure but also to encourage the use of detailed levels of analysis in theoretical research.

Bibliography

Aitkin, Murray (2010). *Statistical Inference: an Integrated Bayesian/Likelihood Approach*. Chapman and Hall/CRC. DOI: 10.1201/ebk1420093438.

Aitkin, Murray, Duy Vu, and Brian Francis (2016). "Statistical modelling of a terrorist network". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180.3, pp. 751–768. DOI: 10.1111/rssa.12233.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari (2017). "The European Trust Crisis and the Rise of Populism". In: *Brookings Papers on Economic Activity* 2017.2, pp. 309–400. DOI: 10.1353/eca.2017.0015.

Allum, Nick, Sanna Read, and Patrick Sturgis (2018). "Evaluating Change in Social and Political Trust in Europe". In: *Cross-Cultural Analysis*. Routledge, pp. 45–63. DOI: 10.4324/9781315537078-2.

Anderson, Christopher, André Blais, Shaun Bowler, Todd Donovan, and Ola Listhaug (2005). "Winning Isn't Everything: Losers' Consent and Democratic Legitimacy". In: *Losers' Consent*. Oxford University Press, pp. 1–14. DOI: 10.1093/0199276382.003.0001.

André, Stéfanie (2013). "Does Trust Mean the Same for Migrants and Natives? Testing Measurement Models of Political Trust with Multi-group Confirmatory Factor Analysis". In: *Social Indicators Research* 115.3, pp. 963–982. DOI: 10.1007/s11205-013-0246-6.

Andreassen, Tor, Bengt Lorentzen, and Ulf Olsson (2006). “The Impact of Non-Normality and Estimation Methods in SEM on Satisfaction Research in Marketing”. In: *Quality & Quantity* 40.1, pp. 39–58. DOI: 10.1007/s11135-005-4510-y.

Anselin, Luc (2010). “Local Indicators of Spatial Association-LISA”. In: *Geographical Analysis* 27.2, pp. 93–115. DOI: 10.1111/j.1538-4632.1995.tb00338.x.

Anselin, Luc and Sergio Rey (2014). *Modern Spatial Econometrics in Practice: A Guide to Geoda, Geodaspace and Pysal*. Geoda Press LLC.

Ares, Macarena and Enrique Hernández (2017). “The corrosive effect of corruption on trust in politicians: Evidence from a natural experiment”. In: *Research & Politics* 4.2, p. 205316801771418. DOI: 10.1177/2053168017714185.

Ariely, Gal and Eldad Davidov (2010). “Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey”. In: *Social Indicators Research* 104.2, pp. 271–286. DOI: 10.1007/s11205-010-9693-5.

Arnold, Christine, Eliyahu Sapir, and Galina Zapryanova (2012). “Trust in the Institutions of the European Union: A Cross-Country Examination”. In: *EIOP European Integration Online Papers* 16. DOI: 10.1695/2012008.

Bache, Ian and Matthew Flinders (2004). “Themes and Issues in Multi-level Governance”. In: *Multi-level Governance*. Oxford University Press, pp. 1–12. DOI: 10.1093/0199259259.003.0001.

Bäck, Maria, Peter Soderlund, Josefina Sipinen, and Elina Kestila-Kekkonen (2018). “Political alienation, generalized trust and anti-immigrant perceptions: A multi-level assessment”. In: *ECPR General Conference*.

Banerjee, Sudipto, Alan Gelfand, and Bradley Carlin (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC. DOI: 10.1201/9780203487808.

Barthelmé, Simon (2012). *Why an inverse-Wishart prior may not be such a good idea*. https://dahtah.wordpress.com/2012/03/07/why-an-inverse-wishart-prior-may-not-be-such-a-good-idea/. Accessed: 2020-01-04.

Bennike, Christian (2019). "New data reveals serious problems with the EU's official public opinion polls." In: *Dagbladet Information*. URL: https://www.information.dk/udland/2019/12/new-data-reveals-serious-problems-with-the-eus-official-public-opinion-polls.

Bentler, Peter and Chih-Ping Chou (1992). "Some New Covariance Structure Model Improvement Statistics". In: *Sociological Methods & Research* 21.2, pp. 259–282. DOI: 10.1177/0049124192021002006.

Bernard, John, Robert McCulloch, and Xiao-Li Meng (2010). "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage". In: *Statistica Sinica* 10, pp. 1281–1311. DOI: 10.5705/ss.2010.253a.

Besag, Julian, Jeremy York, and Annie Mollie (1991). "Bayesian image restoration, with two applications in spatial statistics". In: *Annals of the Institute of Statistical Mathematics* 43.1, pp. 1–20. DOI: 10.1007/bf00116466.

Brown, Gavin, Lois Harris, Chrissie O'Quin, and Kenneth Lane (2015). "Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance". In: *International Journal of Research & Method in Education* 40.1, pp. 66–90. DOI: 10.1080/1743727x.2015.1070823.

Byrne, Barbara, Richard Shavelson, and Bengt Muthén (1989). "Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance." In: *Psychological Bulletin* 105.3, pp. 456–466. DOI: 10.1037/0033-2909.105.3.456.

Carlin, Bradley and María-Eglée Pérez (2000). "Robust Bayesian Analysis in Medical and Epidemiological Settings". In: *Robust Bayesian Analysis*. Springer New York, pp. 351–372. DOI: 10.1007/978-1-4612-1306-2_19.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017). "Stan: A Probabilistic Programming Language". In: *Journal of Statistical Software* 76.1. DOI: 10.18637/jss.v076.i01.

Charron, Nicholas and Bo Rothstein (2014). "Social Trust, Quality of Government and Ethnic Diversity. An Empirical Analysis of 206 Regions in Europe". In: *QoG Working Paper Series 2014:20*. The Quality of Government Institute, Department of Political Science, University of Gotheburg.

Chen, Fainian, Kenneth Bollen, Pamela Paxton, Patrick Curran, and James Kirby (2001). "Improper Solutions in Structural Equation Models". In: *Sociological Methods & Research* 29.4, pp. 468–508. DOI: 10.1177/0049124101029004003.

Chen, Fang Fang (2008). "What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research." In: *Journal of Personality and Social Psychology* 95.5, pp. 1005–1018. DOI: 10.1037/a0013193.

Chen, Fang (2007). "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance". In: *Structural Equation Modeling: A Multidisciplinary Journal* 14.3, pp. 464–504. DOI: 10.1080/10705510701301834.

Clarke, Harold, Matthew Goodwin, and Paul Whiteley (2017). *Brexit*. Cambridge University Press. DOI: 10.1017/9781316584408.

Colatone, Italo and Piero Stanig (2018). "Global Competition and Brexit". In: *American Political Science Review* 112.2, pp. 201–218. DOI: 10.1017/s0003055417000685.

Commission, European (2020). *Public opinion in the European Union: report. Standard Eurobarometer 92. "Europeans opinions about the EU's priorities"*. Tech. rep. European Commission. DOI: 10.2775/033750.

Cordero, Guillermo and Pablo Simón (2015). "Economic Crisis and Support for Democracy in Europe". In: *West European Politics* 39.2, pp. 305–325. DOI: 10.1080/01402382.2015.1075767.

Curran, Patrick, Stephen West, and John Finch (1996). "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis." In: *Psychological Methods* 1.1, pp. 16–29. DOI: 10.1037/1082-989x.1.1.16.

Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt, and Jaak Billiet (2014). "Measurement Equivalence in Cross-National Research". In: *Annual Review of Sociology* 40.1, pp. 55–75. DOI: 10.1146/annurev-soc-071913-043137.

Delhey, Jan and Kenneth Newton (2005). "Predicting Cross-National Levels of Social Trust: Global Pattern or Nordic Exceptionalism?" In: *European Sociological Review* 21.4, pp. 311–327. DOI: 10.1093/esr/jci022.

Depaoli, Sarah and James Clifton (2015). "A Bayesian Approach to Multilevel Structural Equation Modeling With Continuous and Dichotomous Outcomes". In: *Structural Equation Modeling: A Multidisciplinary Journal* 22.3, pp. 327–351. DOI: 10.1080/10705511.2014.937849.

Dillon, William, Ajith Kumar, and Narendra Mulani (1987). "Offending estimates in covariance Structural analysis: Comments on the causes and solutions to Haywood cases". In: *Psychological Bulletin* 101.1, pp. 126–135.

European Social Survey (2017). *European Social Survey (ESS), Round 8 - 2016*. DOI: 10.21338/NSD-ESS8-2016.

Fontanella, Lara, Annalina Sarra, Simone Di Zio, and Pasquale Valentini (2016). "A hierarchical generalised Bayesian SEM to assess quality of democracy in Europe". In: *METRON* 74.1, pp. 117–138. DOI: 10.1007/s40300-016-0081-z.

Foster, Chase and Jeffry Frieden (2017). "Crisis of trust: Socio-economic determinants of Europeans' confidence in government". In: *European Union Politics* 18.4, pp. 511–535. DOI: 10.1177/1465116517723499.

Fotheringham, Stewart and David Wong (1991). "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis". In: *Environment and Planning A: Economy and Space* 23.7, pp. 1025–1044. DOI: 10.1068/a231025.

Fukuyama, Francis (1995). *Trust: The social virtues and the creation of prosperity*. Vol. 99. Free press New York, NY.

Gabry, Jonah (2019). "Package 'loo'". In: *R journal*.

Gagne, Phill and Gregory Hancock (2006). "Measurement Model Quality, Sample Size, and Solution Propriety in Confirmatory Factor Models". In: *Multivariate Behavioral Research* 41.1, pp. 65–83. DOI: 10.1207/s15327906mbr4101_5.

Gao, Shengyi, Patricia Mokhtarian, and Robert Johnston (2008). "Nonnormality of Data in Structural Equation Models". In: *Transportation Research Record: Journal of the Transportation Research Board* 2082.1, pp. 116–124. DOI: 10.3141/2082-14.

Gelman, Andrew (2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)". In: *Bayesian Analysis* 1.3, pp. 515–534. DOI: 10.1214/06-ba117a.

Gelman, Andrew, John Carlin, Hal Stern, David Dunson, and Aki Vehtari (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. ISBN: 1439840954.

Giddens, Anthony (1991). *The Consequences of Modernity*. Stanford University Press. ISBN: 0804718911.

Habermas, Jürgen (2013). *The Crisis of the European Union*. Polity Press. ISBN: 0745662439.

Hambleton, Ronald (1985). *Item response theory : principles and applications*. Boston Hingham, MA, U.S.A: Kluwer-Nijhoff Pub. ISBN: 9789401719889.

Hardin, Russell (1998). "Trust and Governance". In: ed. by Margaret Levi and Valerie Braithwaite. Russell Sage. ISBN: 0871541351. URL: https://www.ebook.de/de/product/3816077/trust_and_governance.html.

Heitjan, Daniel (1989). "Inference from Grouped Continuous Data: A Review". In: *Statistical Science* 4.2, pp. 164–179. DOI: 10.1214/ss/1177012601.

Hetherington, Marc (1998). "The Political Relevance of Political Trust". In: *American Political Science Review* 92.4, pp. 791–808. DOI: 10.2307/2586304.

Heywood, H. B. (1931). "On Finite Sequences of Real Numbers". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 134.824, pp. 486–501. DOI: 10.1098/rspa.1931.0209.

Hobolt, Sara (2016). "The Brexit vote: a divided nation, a divided continent". In: *Journal of European Public Policy* 23.9, pp. 1259–1277. DOI: 10.1080/13501763.2016.1225785.

Hooper, Daire, Joseph Coughlan, and Michael Mullen (2008). "Structural equation modelling: Guidelines for determining model fit". In: *Electronic journal of business research methods* 6.1, pp. 53–60.

Hox, Joop, Rens van de Schoot, and Suzette Matthijsse (2012). "How few countries will do? Comparative survey analysis from a Bayesian perspective". In: *Survey Research Methods; Vol 6, No 2 (2012)* 6.2, pp. 87–93. ISSN: 1864-3361. DOI: 10.18148/srm/2012.v6i2.5033.

Hyman, Michael and Jeremy Sierra (2012). "Adjusting Self-Reported Attitudinal Data for Mischievous Respondents". In: *International Journal of Market Research* 54.1, pp. 129–145. DOI: 10.2501/ijmr-54-1-129-145.

Iacono, Sergio Lo (2019). "Law-breaking, fairness, and generalized trust: The mediating role of trust in institutions". In: *PLOS ONE* 14.8. Ed. by Jonathan Jackson. DOI: 10.1371/journal.pone.0220160.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen (2014). *World Values Survey: Round Six - Country-Pooled Datafile Version: https://www.worldvaluessurvey.org/WVSDocumentationWV6.j*

Inglehart, Ronald and Pippa Norris (2016). "Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2818659.

Jöreskog, Karl Gustav (1971). "Simultaneous factor analysis in several populations". In: *Psychometrika* 36.4, pp. 409–426. DOI: 10.1007/bf02291366.

Kestilä-Kekkonen, Elina and Peter Söderlund (2015). "Political Trust, Individual-level Characteristics and Institutional Performance: Evidence from Finland, 2004-13". In: *Scandinavian Political Studies* 39.2, pp. 138–160. DOI: 10.1111/1467-9477.12052.

Klingemann, Hans-Dieter and Steven Weldon (2012). "A crisis of integration? The development of transnational dyadic trust in the European Union, 1954-2004". In: *European Journal of Political Research* 52.4, pp. 457–482. DOI: 10.1111/1475-6765.12005.

Kolenikov, Stanislav and Kenneth Bollen (2012). "Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification?" In: *Sociological Methods & Research* 41.1, pp. 124–167. DOI: 10.1177/0049124112442138.

Kriesi, Hanspeter (2014). “The Populist Challenge”. In: *West European Politics* 37.2, pp. 361–378. DOI: 10.1080/01402382.2014.887879.

Lavrakas, Paul (2008). *Encyclopedia of Survey Research Methods*. Sage Publications, Inc. DOI: 10.4135/9781412963947.

L’ecuyer, Pierre and Richard Simard (2006). “Inverting the symmetrical beta distribution”. In: *ACM Transactions on Mathematical Software* 32.4. DOI: 10.1145/1186785.1186786.

Lee, Lung-Fei (1979). “On the first and second moments of the truncated multi-normal distribution and a simple estimator”. In: *Economics Letters* 3.2, pp. 165–169. DOI: 10.1016/0165-1765(79)90111-3.

Lei, Ming and Richard Lomax (2005). “The Effect of Varying Degrees of Nonnormality in Structural Equation Modeling”. In: *Structural Equation Modeling: A Multidisciplinary Journal* 12.1, pp. 1–27. DOI: 10.1207/s15328007sem1201_1.

Little, Todd, David Slegers, and Noel Card (2006). “A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models”. In: *Structural Equation Modeling: A Multidisciplinary Journal* 13.1, pp. 59–72. DOI: 10.1207/s15328007sem1301_3.

Lu, Yi and Daniel Bolt (2015). “Examining the attitude-achievement paradox in PISA using a multilevel multidimensional IRT model for extreme response style”. In: *Large-scale Assessments in Education* 3.1. DOI: 10.1186/s40536-015-0012-0.

Lubke, Gitta and Bengt Muthén (2004). “Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons”. In: *Structural Equation Modeling: A Multidisciplinary Journal* 11.4, pp. 514–534. DOI: 10.1207/s15328007sem1104_2.

Lunn, David, Andrew Thomas, Nicky Best, and David Spiegelhalter (2000). "WinBUGS - A Bayesian modeling framework: Concepts, structure and extensibility". In: *Statistics and Computing* 10, pp. 325–337. DOI: 10.1023/A:1008929526011.

MacCallum, Robert, Michael Browne, and Hazuki Sugawara (1996). "Power analysis and determination of sample size for covariance structure modeling." In: *Psychological Methods* 1.2, pp. 130–149. DOI: 10.1037/1082-989x.1.2.130.

Märien, Sofie (2011). "Measuring political trust across time and space". In: *Political trust: Why context matters*. ECPR Press. Chap. 2, p. 13.

Märien, Sofie and Marc Hooghe (2011). "Does political trust matter? An empirical investigation into the relation between political trust and support for law compliance". In: *European Journal of Political Research* 50.2, pp. 267–291. DOI: 10.1111/j.1475-6765.2010.01930.x.

Marozzi, Marco (2014). "Measuring Trust in European Public Institutions". In: *Social Indicators Research* 123.3, pp. 879–895. DOI: 10.1007/s11205-014-0765-9.

McNeish, Daniel (2016). "On Using Bayesian Methods to Address Small Sample Problems". In: *Structural Equation Modeling: A Multidisciplinary Journal* 23.5, pp. 750–773. DOI: 10.1080/10705511.2016.1186549.

Merkle, Edgar and Yves Rosseel (2018). "blavaan: Bayesian Structural Equation Models via Parameter Expansion". In: *Journal of Statistical Software* 85.4. DOI: 10.18637/jss.v085.i04.

Miller, Arthur and Ola Listhaug (1999). "Political performance and institutional trust". In: *Critical citizens: Global support for democratic government*. Oxford University Press Oxford, pp. 204–216.

Mishler, William and Richard Rose (2001). "What Are the Origins of Political Trust?" In: *Comparative Political Studies* 34.1, pp. 30–62. DOI: 10.1177/0010414001034001002.

Moran, Patrick (1950). "Notes on Continuous Stochastic Phenomena". In: *Biometrika* 37.1/2, p. 17. DOI: 10.2307/2332142.

Muñoz, Jordi, Mariano Torcal, and Eduard Bonet (2011). "Institutional trust and multilevel government in the European Union: Congruence or compensation?" In: *European Union Politics* 12.4, pp. 551–574. DOI: 10.1177/1465116511419250.

Muthén, Bengt and Tihomir Asparouhov (2012). "Bayesian structural equation modeling: A more flexible representation of substantive theory." In: *Psychological Methods* 17.3, pp. 313–335. DOI: 10.1037/a0026802.

Muthén, Bengt and David Kaplan (1985). "A comparison of some methodologies for the factor analysis of non-normal Likert variables". In: *British Journal of Mathematical and Statistical Psychology* 38.2, pp. 171–189. DOI: 10.1111/j.2044-8317.1985.tb00832.x.

Newton, Kenneth (1999). "Social and Political Trust in Established Democracies". In: *Critical Citizens*. Oxford University Press, pp. 169–187. DOI: 10.1093/0198295685.003.0008.

– (2001). "Trust, Social Capital, Civil Society, and Democracy". In: *International Political Science Review* 22.2, pp. 201–214. DOI: 10.1177/0192512101222004.

Pinker, Robert (2017). "On the post-Brexit prospects for social policy in the UK". In: *Social policy and welfare pluralism*. Bristol University Press, pp. 295–310. DOI: 10.2307/j.ctt22p7jvf.22.

Plummer, Martyn (2008). "Penalized loss functions for Bayesian model comparison". In: *Biostatistics* 9.3, pp. 523–539. DOI: 10.1093/biostatistics/kxm049.

Plummer, Martyn, Alexey Stukalov, and Matt Denwood (2016). "Package 'rjags'". In: *R journal*.

Putnam, Robert (1995). “Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America”. In: *Political Science and Politics* 28.4, p. 664. DOI: 10.2307/420517.

Putnick, Diane and Marc Bornstein (2016). “Measurement invariance conventions and reporting: The state of the art and future directions for psychological research”. In: *Developmental Review* 41, pp. 71–90. DOI: 10.1016/j.dr.2016.06.004.

Rahn, Wendy and Thomas Rudolph (2005). “A Tale of Political Trust in American Cities”. In: *Public Opinion Quarterly* 69.4, pp. 530–560. DOI: 10.1093/poq/nfi056.

Rasch, Georg (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press. ISBN: 0941938050.

Rosseel, Yves (2012). “lavaan: An R Package for Structural Equation Modeling”. In: *Journal of Statistical Software* 48.2. DOI: 10.18637/jss.v048.i02.

Rothstein, Bo and Dietlind Stolle (2008). “The state and social capital: An institutional theory of generalized trust”. In: *Comparative politics* 40.4, pp. 441–459.

Rubin, Donald, ed. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc. DOI: 10.1002/9780470316696.

Rustenbach, Elisa (2010). “Sources of Negative Attitudes toward Immigrants in Europe: A Multi-Level Analysis”. In: *International Migration Review* 44.1, pp. 53–77. DOI: 10.1111/j.1747-7379.2009.00798.x.

Ryan, Cillian (2016). “Where does one start to make sense of Brexit?” In: *International Economics and Economic Policy* 13.4, pp. 531–537. DOI: 10.1007/s10368-016-0363-1.

Saris, Willem and Irmtraud Gallhofer (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.

Satorra, Albert and Peter Bentler (1994). “Corrections to test statistics and standard errors in covariance structure analysis.” In: *Latent variables analysis: Applications for developmental research*. Ed. by A. von Eye and C. C. Clogg. Sage Publications, Inc.

Schmitt, Hermann, Daniela Braun, Sebastian A. Popa, Slava Mikhaylov, and Felix Dwinger (2016). *European Parliament Election Study 2014, Euromanifesto Study*. de. DOI: 10.4232/1.5162.

Schneider, Irena (2016). “Can We Trust Measures of Political Trust? Assessing Measurement Equivalence in Diverse Regime Types”. In: *Social Indicators Research* 133.3, pp. 963–984. DOI: 10.1007/s11205-016-1400-8.

Schoene, Matthew (2016). “Urban Continent, Urban Activism? European Cities and Social Movement Activism”. In: *Global Society* 31.3, pp. 370–391. DOI: 10.1080/13600826.2016.1203295.

Shi, Dexin, Hairong Song, Xiaolan Liao, Robert Terry, and Lori Snyder (2017). “Bayesian SEM for Specification Search Problems in Testing Factorial Invariance”. In: *Multivariate Behavioral Research* 52.4, pp. 430–444. DOI: 10.1080/00273171.2017.1306432.

Skrondal, Anders and Sophia Rabe-Hesketh (2004). *Generalized Latent Variable Modeling*. Chapman and Hall/CRC. DOI: 10.1201/9780203489437.

Song, Xin-Yuan and Sik-Yum Lee (2012). *Basic and Advanced Bayesian Structural Equation Modeling*. Wiley. DOI: 10.1002/9781118358887.

Spiegelhalter, David, Nicola Best, Bradley Carlin, and Angelika van der Linde (2002). “Bayesian measures of model complexity and fit”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.4, pp. 583–639. DOI: 10.1111/1467-9868.00353.

Steenkamp, Jan-Benedict and Hans Baumgartner (1998). "Assessing Measurement Invariance in Cross-National Consumer Research". In: *Journal of Consumer Research* 25.1, pp. 78–107. DOI: 10.1086/209528.

Sztompka, Piotr (2003). "Trust: A cultural resource". In: *The Moral Fabric in Contemporary Societies*. Vol. 9. Brill, pp. 47–66.

Torcal, Mariano and José Ramón Montero (1999). "Facets of social capital in new democracies". In: *Social capital and European democracy*, pp. 167–91.

Traub, Ross E. (2005). "Classical Test Theory in Historical Perspective". In: *Educational Measurement: Issues and Practice* 16.4, pp. 8–14. DOI: 10.1111/j.1745-3992.1997.tb00603.x.

Tze Hu, Li and Peter Bentler (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives". In: *Structural Equation Modeling: A Multidisciplinary Journal* 6.1, pp. 1–55. DOI: 10.1080/10705519909540118.

Van de Schoot, Rens, Peter Lugtig, and Joop Hox (2012). "A checklist for testing measurement invariance". In: *European Journal of Developmental Psychology* 9.4, pp. 486–492. DOI: 10.1080/17405629.2012.686740.

Van den Bos, Adriaan (2007). "Appendix C: Positive Semidefinite and Positive Definite Matrices". In: *Parameter Estimation for Scientists and Engineers*. John Wiley & Sons, Inc., pp. 259–263. DOI: 10.1002/9780470173862.app3.

Van der Linde, Angelika (2005). "DIC in variable selection". In: *Statistica Neerlandica* 59.1, pp. 45–56. DOI: 10.1111/j.1467-9574.2005.00278.x.

Van der Meer, Tom (2010). "In what we trust? A multi-level study into trust in parliament as an evaluation of state characteristics". In: *International Review of Administrative Sciences* 76.3, pp. 517–536. DOI: 10.1177/0020852310372450.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2016). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* 27.5, pp. 1433–1433. DOI: 10.1007/s11222-016-9709-3.

Verhagen, Josine and Jean-Paul Fox (2012). "Bayesian tests of measurement invariance". In: *British Journal of Mathematical and Statistical Psychology*. DOI: 10.1111/j.2044-8317.2012.02059.x.

Wilhelm, Stefan (2015). "Package 'tmvtnorm'". In: *R journal*.

Zercher, Florian, Peter Schmidt, Jan Cieciuch, and Eldad Davidov (2015). "The comparability of the universalism value over time and across countries in the European Social Survey: exact vs. approximate measurement invariance". In: *Frontiers in Psychology* 6. DOI: 10.3389/fpsyg.2015.00733.

Zielonka, Jan (2014). *Is the EU Doomed?* Polity Press. ISBN: 0745683975.

Ziller, Conrad (2014). "Ethnic Diversity, Economic and Cultural Contexts, and Social Trust: Cross-Sectional and Longitudinal Evidence from European Regions, 2002-2010". In: *Social Forces* 93.3, pp. 1211–1240. DOI: 10.1093/sf/sou088.

Appendix 1 - Random intercepts for NUTS regions in UK MLM

Region	**Total**
East Midlands	-0.1017739
East of England	-0.0561702
London	0.2628417
North East	-0.0018723
North West	-0.1200933
Northern Ireland	-0.067964
Scotland	0.6557528
South East	-0.1419616
South West	-0.1860316
Wales	0.0662259
West Midlands	-0.2574747
Yorkshire and The Humber	-0.0514788

Appendix 2 - Sample changes to NUTS classification

Country	Year	NUTS level	Previous NUTS code	New NUTS code
Finland	2010	2	FI131	FI1D1
Finland	2010	2	FI132	FI1D2
Finland	2010	2	FI133	FI1D3
Finland	2010	2	FI134	FI1D4
Finland	2010	2	FI181	FI1B1
Finland	2010	2	FI182	FI1B1
Finland	2010	2	FI183	FI1C1
Finland	2010	2	FI184	FI1C2
Finland	2010	2	FI185	FI1C3
Finland	2010	2	FI186	FI1C4
Finland	2010	2	FI187	FI1C5
Finland	2010	2	FI1A1	FI1D5
Finland	2010	2	FI1A2	FI1D6
Finland	2010	2	FI1A3	FI1D7
Slovenia	2013	2	SI011	SI031
Slovenia	2013	2	SI012	SI032
Slovenia	2013	2	SI013	SI033
Slovenia	2013	2	SI014	SI034
Slovenia	2013	2	SI015	SI035
Slovenia	2013	2	SI016	SI036
Slovenia	2013	2	SI017	SI037
Slovenia	2013	2	SI018	SI038
Slovenia	2013	2	SI021	SI041
Slovenia	2013	2	SI022	SI042
Slovenia	2013	2	SI023	SI043
Slovenia	2013	2	SI024	SI044
Ireland	2016	2	IE011	IE041
Ireland	2016	2	IE012	IE063
Ireland	2016	2	IE013	IE042
Ireland	2016	2	IE021	IE061
Ireland	2016	2	IE022	IE062
Ireland	2016	2	IE023	IE051
Ireland	2016	2	IE024	IE052
Ireland	2016	2	IE025	IE053
France	2016	2	FR21	FRF2
France	2016	2	FR22	FRE2

Country	Year	NUTS level	Previous NUTS code	New NUTS code
France	2016	2	FR23	FRD2
France	2016	2	FR24	FRB0
France	2016	2	FR25	FRD1
France	2016	2	FR26	FRC1
France	2016	2	FR30	FRE1
France	2016	2	FR41	FRF3
France	2016	2	FR42	FRF1
France	2016	2	FR43	FRC2
France	2016	2	FR51	FRG0
France	2016	2	FR52	FRH0
France	2016	2	FR53	FRI3
France	2016	2	FR61	FRI1
France	2016	2	FR62	FRJ2
France	2016	2	FR63	FRI2
France	2016	2	FR71	FRK2
France	2016	2	FR72	FRK1
France	2016	2	FR81	FRJ1
France	2016	2	FR82	FRL0
Lithuania	2016	2	LT00A	LT011
Lithuania	2016	2	LT001	LT021
Lithuania	2016	2	LT002	LT022
Lithuania	2016	2	LT003	LT023
Lithuania	2016	2	LT004	LT024
Lithuania	2016	2	LT005	LT025
Lithuania	2016	2	LT006	LT026
Lithuania	2016	2	LT007	LT027
Lithuania	2016	2	LT008	LT028
Lithuania	2016	2	LT009	LT029
Hungary	2016	2	HU101	HU110
Hungary	2016	2	HU102	HU120
Poland	2016	2	PL11	PL71
Poland	2016	2	PL12	PL91
Poland	2016	2	PL31	PL81
Poland	2016	2	PL32	PL82
Poland	2016	2	PL34	PL84

Appendix 3 - NUTS regions and sample sizes in Chapter 3

Region	ESS round				Total
	5	6	7	8	
BE10	106	164	139	186	**595**
BE21	302	299	249	275	**1,125**
BE22	165	160	148	148	**621**
BE23	214	219	203	249	**885**
BE24	166	180	156	159	**661**
BE25	199	217	218	206	**840**
BE31	54	60	60	61	**235**
BE32	170	184	231	160	**745**
BE33	143	208	181	133	**665**
BE34	47	51	52	47	**197**
BE35	73	72	76	86	**307**
CZ010	257	236	244	267	**1,004**
CZ020	217	186	228	247	**878**
CZ031	137	114	103	121	**475**
CZ032	99	102	117	126	**444**
CZ041	61	59	71	57	**248**
CZ042	169	142	176	193	**680**
CZ051	98	62	94	87	**341**
CZ052	109	97	119	116	**441**
CZ053	107	90	111	102	**410**
CZ063	109	96	89	100	**394**
CZ064	234	205	186	236	**861**
CZ071	153	114	95	134	**496**
CZ072	140	111	87	100	**438**
CZ080	283	217	276	234	**1,010**
DE1	236	280	302	289	**1,107**
DE2	385	358	398	364	**1,505**
DE3	91	98	150	114	**453**
DE4	197	159	166	154	**676**
DE7	179	151	185	177	**692**
DE8	110	121	115	112	**458**
DE9	155	246	216	193	**810**
DEA	499	440	467	470	**1,876**

Region	ESS round				Total
	5	6	7	8	
DEB	117	125	122	109	**473**
DED	219	286	250	261	**1,016**
DEE	161	142	146	135	**584**
DEF	71	84	79	87	**321**
DEG	177	122	148	140	**587**
EE001	468	738	823	776	**2,805**
EE004	188	217	177	208	**790**
EE006	169	196	155	171	**691**
EE007	137	263	320	169	**889**
EE008	471	542	344	462	**1,819**
ES11	85	98	114	115	**412**
ES21	74	68	68	71	**281**
ES24	54	48	45	45	**192**
ES30	284	284	225	221	**1,014**
ES41	99	78	83	94	**354**
ES42	79	80	79	61	**299**
ES51	192	239	232	250	**913**
ES52	162	164	159	172	**657**
ES61	373	316	289	311	**1,289**
ES62	43	52	42	42	**179**
ES70	61	65	61	71	**258**
FI193	97	116	87	98	**398**
FI194	57	67	85	72	**281**
FI195	68	65	66	69	**268**
FI196	88	100	100	95	**383**
FI197	156	196	188	152	**692**
FI1B1	432	562	536	521	**2,051**
FI1C1	169	188	166	149	**672**
FI1C2	69	70	65	52	**256**
FI1C3	63	72	66	70	**271**
FI1C4	56	70	81	66	**273**
FI1C5	49	54	44	46	**193**
FI1D1	58	60	52	53	**223**
FI1D2	94	106	91	97	**388**
FI1D3	53	60	54	56	**223**
FI1D6	124	154	141	128	**547**
FI1D7	64	80	71	69	**284**
FR10	250	285	272	242	**1,049**
FRB0	56	87	80	77	**300**

Region	ESS round				Total
	5	**6**	**7**	**8**	
FRD1	45	44	45	60	**194**
FRD2	49	60	50	56	**215**
FRE1	117	114	85	125	**441**
FRF1	45	69	63	67	**244**
FRF3	89	108	41	87	**325**
FRG0	107	98	110	122	**437**
FRH0	91	93	126	113	**423**
FRI1	95	93	116	109	**413**
FRJ1	54	54	80	93	**281**
FRJ2	68	101	129	107	**405**
FRK2	172	228	138	233	**771**
FRL0	101	127	169	147	**544**
HU110	168	279	283	275	**1,005**
HU120	151	217	208	135	**711**
HU211	48	64	55	64	**231**
HU221	55	68	52	57	**232**
HU231	56	61	55	70	**242**
HU232	53	77	44	42	**216**
HU311	111	107	80	114	**412**
HU321	50	94	79	46	**269**
HU322	80	85	63	80	**308**
HU323	76	89	61	67	**293**
HU331	81	124	70	112	**387**
IE041	279	312	247	303	**1,141**
IE042	303	236	211	245	**995**
IE051	186	163	167	166	**682**
IE052	267	189	266	232	**954**
IE053	278	310	290	343	**1,221**
IE061	401	514	468	596	**1,979**
IE062	208	311	196	227	**942**
IE063	111	138	130	162	**541**
LT011	268	430	463	466	**1,627**
LT021	71	72	83	82	**308**
LT022	255	382	430	429	**1,496**
LT023	101	173	190	193	**657**
LT024	81	81	92	101	**355**
LT025	159	157	185	165	**666**
LT026	120	141	181	162	**604**
LT027	43	76	91	79	**289**

Region	ESS round				Total
	5	6	7	8	
LT028	65	70	90	92	**317**
LT029	55	83	81	93	**312**
NL11	54	59	43	43	**199**
NL12	78	66	76	81	**301**
NL13	47	57	69	51	**224**
NL21	134	121	123	105	**483**
NL22	193	210	226	214	**843**
NL31	133	152	126	105	**516**
NL32	287	296	254	224	**1,061**
NL33	324	290	363	278	**1,255**
NL41	262	265	255	242	**1,024**
NL42	117	109	148	116	**490**
PL21	111	155	147	140	**553**
PL22	165	187	147	172	**671**
PL41	129	118	121	112	**480**
PL42	57	76	51	56	**240**
PL51	93	109	78	113	**393**
PL61	88	83	84	63	**318**
PL62	54	50	44	57	**205**
PL63	94	81	78	97	**350**
PL71	94	112	105	90	**401**
PL81	98	84	81	92	**355**
PL82	88	90	97	100	**375**
PL84	42	43	51	51	**187**
PL91	191	216	182	183	**772**
PT11	726	638	414	419	**2,197**
PT16	294	357	269	255	**1,175**
PT17	594	746	284	324	**1,948**
PT18	70	76	105	108	**359**
SE110	218	394	371	263	**1,246**
SE123	62	61	51	71	**245**
SE224	161	229	181	218	**789**
SE232	212	272	248	44	**776**
SE312	41	41	44	50	**176**
SI031	74	71	81	86	**312**
SI032	242	203	144	185	**774**
SI034	161	125	135	170	**591**
SI037	100	78	74	100	**352**
SI041	230	245	242	250	**967**

Region	ESS round				Total
	5	**6**	**7**	**8**	
SI042	110	99	98	106	**413**
SI043	76	54	68	58	**256**
UKC	83	87	94	68	**332**
UKD	196	184	224	235	**839**
UKE	183	148	147	142	**620**
UKF	143	140	159	148	**590**
UKG	161	161	164	157	**643**
UKH	183	164	187	166	**700**
UKI	173	185	181	157	**696**
UKJ	273	235	290	227	**1,025**
UKK	156	161	189	152	**658**
UKL	106	99	120	112	**437**
UKM	199	174	186	148	**707**
UKN	43	65	53	61	**222**
Total	**24,767**	**26,910**	**25,729**	**25,568**	**102,974**

Appendix 4 - R code for multiple imputation

```r
#Required packages
library(readstata13)
library(lavaan)
library(tmvtnorm)
library(haven)
library(ggplot2)
library(mitools)
library(survey)
library(lavaan.survey)
library(mice)
library(plyr)
library(matrixStats)
library(semTools)

#Disable scientific notation
options(scipen = 999)

#Reading in ESS7 data
ess8 <- read.dta13("ESS8.dta", nonint.factors = TRUE)
head(ess8)
warnings()

#Descriptives
aggregate(ess8[, 2:6], list(ess8$country), mean)

mean(ess8$trstprl)
mean(ess8$trstplt)
mean(ess8$trstprt)
mean(ess8$trstep)
```

```
mean(ess8$trstun)

#Metric invariance model definition

trust.mg.metric1 <- 'NPT =~ c(l1, l1, l1, l1, l1, l1, l1, l1, l1, l1,
    l1, l1, l1, l1, l1, l1, l1, l1)*trstprl +
c(l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2, l2,
    l2)*trstplt +
c(l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3, l3,
    l3)*trstprt
SPT =~ c(l4, l4, l4, l4, l4, l4, l4, l4, l4, l4, l4, l4, l4, l4, l4,
    l4, l4, l4)*trstun +
c(l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5, l5,
    l5)*trstep

3 == l1+l2+l3
2 == l4+l5

trstprl ~ c(i11, i12, i13, i14, i15, i16, i17, i18, i19, i110, i111,
    i112, i113, i114, i115, i116, i117, i118)*1
trstplt ~ c(i21, i22, i23, i24, i25, i26, i27, i28, i29, i210, i211,
    i212, i213, i214, i215, i216, i217, i218)*1
trstprt ~ c(i31, i32, i33, i34, i35, i36, i37, i38, i39, i310, i311,
    i312, i313, i314, i315, i316, i317, i318)*1
trstep ~ c(i41, i42, i43, i44, i45, i46, i47, i48, i49, i410, i411,
    i412, i413, i414, i415, i416, i417, i418)*1
trstun ~ c(i51, i52, i53, i54, i55, i56, i57, i58, i59, i510, i511,
    i512, i513, i514, i515, i516, i517, i518)*1

0 == i11+i21+i31
0 == i12+i22+i32
```

0 == i13+i23+i33

0 == i14+i24+i34

0 == i15+i25+i35

0 == i16+i26+i36

0 == i17+i27+i37

0 == i18+i28+i38

0 == i19+i29+i39

0 == i110+i210+i310

0 == i111+i211+i311

0 == i112+i212+i312

0 == i113+i213+i313

0 == i114+i214+i314

0 == i115+i215+i315

0 == i116+i216+i316

0 == i117+i217+i317

0 == i118+i218+i318

0 == i41+i51

0 == i42+i52

0 == i43+i53

0 == i44+i54

0 == i45+i55

0 == i46+i56

0 == i47+i57

0 == i48+i58

0 == i49+i59

0 == i410+i510

0 == i411+i511

0 == i412+i512

0 == i413+i513

0 == i414+i514

```
0 == i415+i515
0 == i416+i516
0 == i417+i517
0 == i418+i518'

model.metric1.mg <- lavaan(trust.mg.metric1,
                           data = ess8,
                           group = "country",
                           group.equal = c("loadings"),
                           estimator = "MLM",
                           int.ov.free = TRUE,
                           int.lv.free = TRUE,
                           auto.fix.first = FALSE,
                           auto.var = TRUE,
                           auto.cov.lv.x = TRUE)

#summary(model.metric1.mg, standardized = TRUE, fit.measures = TRUE)
fitMeasures(model.metric1.mg, c("chisq.scaled", "df.scaled", "cfi.
    scaled", "rmsea.scaled", "srmr", "aic", "bic"))

####Imputation

#Define imputation data and create output data
imputation_data <- ess8[c(2,3,4,5,6,1)]
imputation_data$country =  as.factor(imputation_data$country)
imputation_output <- matrix(0,dim(imputation_data)[1],dim(imputation_
    data)[2])
impreps <- 30
impnames <- sprintf("imp_%d",seq(1:impreps))
impoutnames <- sprintf("impout_%d",seq(1:impreps))
```

```
#Create imprep duplicates of imputation data
for(j in 1:impreps) {
  assign(paste("imp",j,sep="_"),imputation_data)
}
imp.list <- mget(paste0("imp_", 1:impreps))
for (j in 1:length(impnames)) {
  imp.list[[j]] <- get(impnames[j])
}

#Create impreps duplicates of output matrix
for(h in 1:impreps) {
  assign(paste("impout",h,sep="_"),imputation_output)
}

impout.list <- mget(paste0("impout_", 1:impreps))

for (h in 1:length(impoutnames)) {
  impout.list[[h]] <- get(impoutnames[h])
}

#Define the bounds of the distribution
lower <- c(0)
upper <- c(Inf)

#De-truncating across all items and countries
for (j in 1:length(impnames))
{
  for (i in levels(imp.list[[j]]$country))
  {
    data = subset (imp.list[[j]], imp.list[[j]]$country == i)
```

```
    for (k in c(1:5))
    {
      gmm.fit <- gmm.tmvnorm(matrix(imp.list[[j]][,k],length(imp.list[[
          j]][,k]),1), lower=lower, upper=upper)

      useMu <- matrix(gmm.fit$coefficients[1],1,1)
      useSigma <- matrix(gmm.fit$coefficients[2],1,1)
      replaceThese <- imp.list[[j]][,k]<=0
      impout.list[[j]][,k] <- imp.list[[j]][,k]

      impout.list[[j]][replaceThese,k]  <- rtmvnorm(n=sum(replaceThese)
          , c(useMu), c(useSigma), c(-Inf), c(0))
      impout.list[[j]] <- subset(impout.list[[j]], select = c
          (1,2,3,4,5))
      impout.list[[j]] <- cbind(impout.list[[j]], imputation_data$
          country)
      colnames(impout.list[[j]]) <- c("trstprl","trstplt","trstprt","
          trstep","trstun", "country")
    }
  }
}

#Distributions before/after for one imputation
ess8_impute <- as.matrix(impout.list$impout_15)

par(mfrow=c(2,5))

for (k in c(1:5))
{
  hist(imputation_data[,k],
       main='',
```

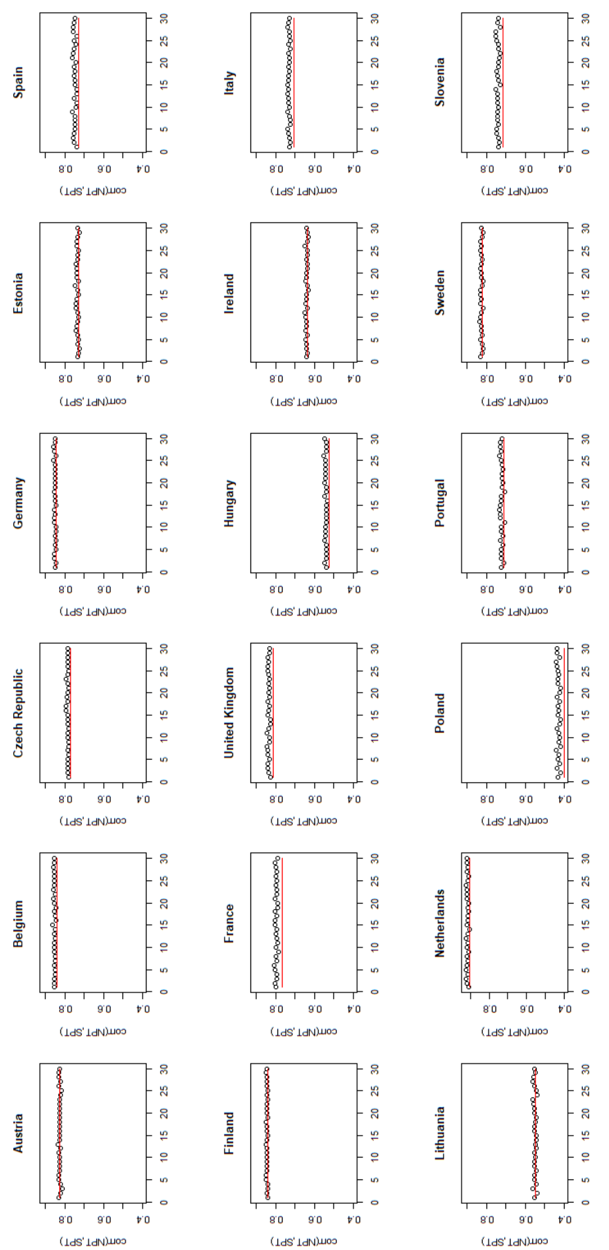
```
        xlab = 'Response value'[k],
        xlim=c(-11.5,10.5),
        col="grey",
        breaks = c(seq(-11.5, 10.5, 1))
  )
  hist(ess8_impute[,k],
        main='',
        xlab = 'Imputed value',
        xlim=c(-11.5,10.5),
        col="blue",
        breaks = c(seq(-11.5, 10.5, 1))
  )
}

##Define the number of reps used and create a results vector
impute_est <- vector('list',impreps)


##Run the model for all impreps and save results in impute_est[j]
for (j in c(1:impreps)) {
  impute_test <- as.data.frame(impout.list[[j]])

  model_metric1.impute <- lavaan(trust.mg.metric1,
                                  data = impute_test,
                                  group = "country",
                                  group.equal = c("loadings"),
                                  estimator = "MLM",
                                  int.ov.free = TRUE,
                                  int.lv.free = TRUE,
                                  auto.fix.first = FALSE,
                                  auto.var = TRUE,
```

```
                                        auto.cov.lv.x = TRUE)
  impute_est[[j]] <- model_metric1.impute # here is where you store the
      results
}

#Create matrix for storing estimates (fit, NPT~~SPT est, se and
    standardised)
fit_impute <- matrix(0,impreps,7)
corr_imp_est <- matrix(0,impreps,18)
corr_imp_se <- matrix(0,impreps,18)
corr_imp_std <- matrix(0,impreps,18)

#Define the cells to extract
dummy <- parameterEstimates(impute_est[[j]], standardized = TRUE)
userows <- (dummy$lhs == 'NPT')* (dummy$op == '~~')* (dummy$rhs == 'SPT
    ')

#Extract all the relevant estimates and store in seperate lists
for (i in c(1:impreps))
{
  fit_impute[i,] <- fitMeasures(impute_est[[i]], c("chisq.scaled", "df.
      scaled", "cfi.scaled", "rmsea.scaled", "srmr", "aic", "bic"))
  corr_imp_est[i,]<- parameterEstimates(impute_est[[i]], standardized =
      TRUE)[userows==1,]$est
  corr_imp_se[i,]<- parameterEstimates(impute_est[[i]], standardized =
      TRUE)[userows==1,]$se
  corr_imp_std[i,]<- parameterEstimates(impute_est[[i]], standardized =
      TRUE)[userows==1,]$std.all
}

#### Observed model fit results model_scalar2.mg vs imputed data model_
```

```
  scalar2.impute

obs_est  <- parameterEstimates(model.metric1.mg, standardized = TRUE)[
  userows==1,]$std.all
fit_est <- as.matrix(fitMeasures(model.metric1.mg, c("chisq.scaled", "
  df.scaled", "cfi.scaled", "rmsea.scaled", "srmr", "aic", "bic")))

#CFI
par(mfrow=c(1,1))

plot(fit_impute[,3],ylim=c(fit_est[3,1]-0.005,fit_est[3,1]+0.005),
     xlab = "Imputation data",
     main = "CFI",
     ylab = "CFI",
     cex.main=2.0, cex.lab=1.5, cex.axis=1.5)
lines(c(1,impreps),c(fit_est[3,1],fit_est[3,1]),col='red')

#RMSEA
plot(fit_impute[,4],ylim=c(fit_est[4,1]-0.05,fit_est[4,1]+0.05),
     xlab = "Imputation data",
     main = "RMSEA",
     ylab = "RMSEA",
     cex.main=2.0, cex.lab=1.5, cex.axis=1.5)
lines(c(1,impreps),c(fit_est[4,1],fit_est[4,1]),col='red')

#SRMR
plot(fit_impute[,5],ylim=c(fit_est[5,1]-0.005,fit_est[5,1]+0.005),
     xlab = "Imputation data",
     main = "SRMR",
     ylab = "SRMR",
     cex.main=2.0, cex.lab=1.5, cex.axis=1.5)
```

```
lines(c(1,impreps),c(fit_est[5,1],fit_est[5,1]),col='red')

#Difference across countries in corr(NPT,SPT)
colnames(corr_imp_std) <- c("Austria","Belgium","Czech Republic","
    Germany","Estonia","Spain","Finland","France",
                                   "United Kingdom","Hungary","Ireland","Italy
                                       ","Lithuania","Netherlands","Poland",
                                   "Portugal","Sweden","Slovenia")

par(mfrow=c(3,6))
for (country in c(1:18))
{
  plot(corr_imp_std[,country], ylim=range(c(corr_imp_std[,country],obs_
      est)),
        main=colnames(corr_imp_std)[country],
        ylab="corr(NPT,SPT)",
        xlab="")
  lines(c(1,impreps),c(obs_est[country], obs_est[country]),col='red')
}
```

Appendix 5 - Correlation between NPT and SPT for 30 imputed datasets



Correlation between NPT and SPT for 30 imputed datasets.

Appendix 6 - JAGS model code for BMGCFA configural invariance model

```
model {

###DATA AND MODEL

for(i in 1:N) {
        trstprl[i] ~ dnorm(mu[i,1], 1/thetastar[1,g[i]])
        trstplt[i] ~ dnorm(mu[i,2], 1/thetastar[2,g[i]])
        trstprt[i] ~ dnorm(mu[i,3], 1/thetastar[3,g[i]])
        trstep[i] ~ dnorm(mu[i,4], 1/thetastar[4,g[i]])
        trstun[i] ~ dnorm(mu[i,5], 1/thetastar[5,g[i]])

#LATENT VARIABLES
        eta[i,1:2] ~ dmnorm(mu_eta[i,1:2], ibpsi[1:2,1:2,g[i]])
        eta[i,3] ~ dnorm(mu_eta[i,3], ibdpsi[1,g[i]])

#LATENT MEANS
        mu_eta[i,1] <- nu[1,g[i]]
        mu_eta[i,2] <- nu[2,g[i]]
        mu_eta[i,3] <- dnu[3,g[i]]

#MU
        mu[i,1] <- alpha[1,g[i]] + lambda[1,g[i]]*eta[i,1]
        mu[i,2] <- alpha[2,g[i]] + lambda[2,g[i]]*eta[i,1] + dlam[2,g[i
            ]]*eta[i,3]
        mu[i,3] <- alpha[3,g[i]] + lambda[3,g[i]]*eta[i,1] + dlam[3,g[i
            ]]*eta[i,3]
        mu[i,4] <- alpha[4,g[i]] + lambda[4,g[i]]*eta[i,2]
        mu[i,5] <- alpha[5,g[i]] + lambda[5,g[i]]*eta[i,2]
```

```
}

###PRIORS
#ITEM INTERCEPTS
for(k in 1:K){
        alpha[1,k] ~ dnorm(0,0.01)
        alpha[2,k] ~ dnorm(0,0.01)
        alpha[3,k] <- dsum(-alpha[1,k],-alpha[2,k])
        alpha[4,k] ~ dnorm(0,0.01)
        alpha[5,k] <- dsum(-alpha[4,k])
}
#ITEM LOADINGS
for(k in 1:K){
        lambda[1,k] ~ dnorm(1,0.01)
        lambda[2,k] ~ dnorm(1,0.01)
        lambda[3,k] ~ dsum(3,-lambda[1,k],-lambda[2,k])
        lambda[4,k] ~ dnorm(1,0.01)
        lambda[5,k] ~ dsum(2,-lambda[4,k])
}

#LATENT MEANS
for(k in 1:K){
        nu[1,k] ~ dnorm(5,0.01)
        nu[2,k] ~ dnorm(5,0.01)
        dnu[3,k] <- 0
}
#LATENT COVARIANCE AND CORRELATION
I[1,1] <- 1
I[1,2] <- 0
I[2,1] <- 0
I[2,2] <- 1
```

```
for(k in 1:K){
        ibpsi[1:2,1:2,k] ~ dwish(I,3)
        psi[1:2,1:2,k] <- inverse(ibpsi[1:2,1:2,k])
        latcor[1,k] <- psi[1,2,k]/sqrt(psi[1,1,k]*psi[2,2,k])
        ibdpsi[1,k] <- 1
}
#ERROR VARIANCES
for(k in 1:K){
        theta[1,k] <- ersd[1,k]^2
        theta[2,k] <- ersd[2,k]^2
        theta[3,k] <- ersd[3,k]^2
        theta[4,k] <- ersd[4,k]^2
        theta[5,k] <- ersd[5,k]^2
        ersd[1,k] ~ dt(0,0.5,20) T(0,)
        ersd[2,k] ~ dt(0,0.5,20) T(0,)
        ersd[3,k] ~ dt(0,0.5,20) T(0,)
        ersd[4,k] ~ dt(0,0.5,20) T(0,)
        ersd[5,k] ~ dt(0,0.5,20) T(0,)
}

#ERROR VARIANCES STAR
for(k in 1:K){
        thetastar[1,k] <- theta[1,k]
        thetastar[2,k] <- theta[2,k] - (sqrt(abs(rho[2,3,k])*theta[2,k
            ]))^2
        thetastar[3,k] <- theta[3,k] - ((-1 + 2*step(rho[2,3,k]))*sqrt(
            abs(rho[2,3,k])*theta[3,k]))^2
        thetastar[4,k] <- theta[4,k]
        thetastar[5,k] <- theta[5,k]
}
```

```
#CORRELATIONS
for(k in 1:K){
        rho[2,3,k] <- -1 + 2*rhodist[1,k]
        rhodist[1,k] ~ dbeta(1,1)
}

#PHANTOM LOADINGS
for(k in 1:K){
        dlam[2,k] <- sqrt(abs(rho[2,3,k])*theta[2,k])
        dlam[3,k] <- (-1 + 2*step(rho[2,3,k]))*sqrt(abs(rho[2,3,k])*
            theta[3,k])
}

#INFERENTIAL COVARIANCES
for(k in 1:K){
        thetacov[2,3,k] <- dlam[2,k]*dlam[3,k]*ibdpsi[1,k]
}

#LOG DENSITY
for(k in 1:K){
        cov.mat[1,1,k] <- lambda[1,k]^2 * psi[1,1,k] + theta[1,k]
        cov.mat[1,2,k] <- lambda[1,k] * lambda[2,k] * psi[1,1,k]
        cov.mat[1,3,k] <- lambda[1,k] * lambda[3,k] * psi[1,1,k]
        cov.mat[1,4,k] <- 0
        cov.mat[1,5,k] <- 0
        cov.mat[2,1,k] <- lambda[2,k] * lambda[1,k] * psi[1,1,k]
        cov.mat[2,2,k] <- lambda[2,k]^2 * psi[1,1,k] + theta[2,k] +
            dlam[2,k]^2
        cov.mat[2,3,k] <- lambda[2,k] * lambda[3,k] * psi[1,1,k] +
            thetacov[2,3,k] + dlam[2,k]*dlam[3,k]
```

```
        cov.mat[2,4,k] <- 0
        cov.mat[2,5,k] <- 0
        cov.mat[3,1,k] <- lambda[3,k] * lambda[1,k] * psi[1,1,k]
        cov.mat[3,2,k] <- lambda[3,k] * lambda[2,k] * psi[1,1,k] +
            thetacov[2,3,k] + dlam[2,k]*dlam[3,k]
        cov.mat[3,3,k] <- lambda[3,k]^2 * psi[1,1,k] + theta[3,k] +
            dlam[3,k]^2
        cov.mat[3,4,k] <- 0
        cov.mat[3,5,k] <- 0
        cov.mat[4,1,k] <- 0
        cov.mat[4,2,k] <- 0
        cov.mat[4,3,k] <- 0
        cov.mat[4,4,k] <- lambda[4,k]^2 * psi[2,2,k] + theta[4,k]
        cov.mat[4,5,k] <- lambda[4,k] * lambda[5,k] * psi[2,2,k]
        cov.mat[5,1,k] <- 0
        cov.mat[5,2,k] <- 0
        cov.mat[5,3,k] <- 0
        cov.mat[5,4,k] <- lambda[5,k] * lambda[4,k] * psi[2,2,k]
        cov.mat[5,5,k] <- lambda[5,k]^2 * psi[2,2,k] + theta[5,k]
        sigma[1:5,1:5,k] <- inverse(cov.mat[1:5,1:5,k])
}
for(i in 1:N){
        data.mv[i,1:5] <- c(trstprl[i],trstplt[i],trstprt[i],trstep[i],
            trstun[i])
        mean.trstprl[i] <- alpha[1,g[i]] + nu[1,g[i]] * lambda[1,g[i]]
        mean.trstplt[i] <- alpha[2,g[i]] + nu[1,g[i]] * lambda[2,g[i]]
        mean.trstprt[i] <- alpha[3,g[i]] + nu[1,g[i]] * lambda[3,g[i]]
        mean.trstep[i] <- alpha[4,g[i]] + nu[2,g[i]] * lambda[4,g[i]]
        mean.trstun[i] <- alpha[5,g[i]] + nu[2,g[i]] * lambda[5,g[i]]
        mean.mv[i,1:5] <- c(mean.trstprl[i],mean.trstplt[i],mean.
            trstprt[i],mean.trstep[i],mean.trstun[i])
```

```
        log_lik[i] <- logdensity.mnorm(data.mv[i,], mean.mv[i,], sigma
            [,,g[i]])
    }
}
```

Appendix 7 - R code for BMGCFA model

```
#Required packages
library(saveJAGS)
library(loo)
library(wiqid)

#Set number of cores for loo package
options(mc.cores = 10)

#Disable scientific notation and max print
options(scipen = 999)
options(max.print=999999)

#Read model data
nuts.data <- read.bugsdata('nuts_full.txt')

####NUTS CONFIGURAL MODEL
#Define monitored parameters
jags.params <- c("alpha", "lambda", "theta", "nu", "rho", "log_lik")
n.nuts <- nuts.data$K

#Define initial values for non-deterministic nodes
nuts.inits <- function(chain) list(alpha = matrix(rep(c(1,1,1,1,NA),n.
   nuts), nrow=5, ncol=n.nuts), lambda = matrix(rep(c(1,1,1,1,NA),n.
   nuts), nrow=5, ncol=n.nuts))

#Run model
ptm <- proc.time()
nuts.config.model <- saveJAGS(data = nuts.data,
                              params = jags.params,
```

```
                              modelFile = "model_nuts_config.txt",
                              inits = nuts.inits,
                              chains = 2,
                              sample2save = 400,
                              nSaves = 5,
                              burnin = 500,
                              thin = 5,
                              fileStub = "nuts_config_files/nutsconfig")
proc.time()-ptm

#Quick summary
str(nuts.config.model)
summary(nuts.config.model)

#Extract parameter of interest
mcmc.nuts.config.nu <- combineSaves(nuts.config.model, params=c("nu"))
str(mcmc.nuts.config.nu)
format(object.size(mcmc.nuts.config.nu), units = "Mb")

#Quick diagnostics
nuts.config.bw.nu <- as.Bwiqid(mcmc.nuts.config.nu)
nuts.config.bw.nu

diagPlot(nuts.config.bw.nu)
tracePlot(nuts.config.bw.nu)
densityPlot(nuts.config.bw.nu)
acfPlot(nuts.config.bw.nu)

#Extract log_likelihood
mcmc.nuts.config.loglik <- combineSaves(nuts.config.model, params=c("
   log_lik"))
```

```
str(mcmc.nuts.config.loglik)
format(object.size(mcmc.nuts.config.loglik), units = "Mb")

#Calculate DIC manually (ineffective at the pD stage)
ln.paramlist.nuts.config <- simsList(mcmc.nuts.config.loglik)
ln.loglik.nuts.config <- ln.paramlist.nuts.config$log_lik
est.ln.dev.nuts.config <- -2 * rowSums(ln.loglik.nuts.config)
deviance.nuts.config <- mean(est.ln.dev.nuts.config)
pd.nuts.config <- sum(var(ln.loglik.nuts.config))
dic.nuts.config <- deviance.nuts.config + pd.nuts.config

#Calculate WAIC
ln.waic.nuts.config <- waic(ln.loglik.nuts.config)
ln.waic.nuts.config

#Calculate LOOIC
ln.loo.nuts.config <- loo(ln.loglik.nuts.config)
ln.loo.nuts.config
plot(ln.loo.nuts.config)
```

Appendix 8 - JAGS model code for Marien (2011) model

```
model {
    ###DATA AND MODEL
    for(i in 1:N){
        trstprl[i] ~ dnorm(mu[i,1], 1/thetastar[1,g[i]])
        trstplt[i] ~ dnorm(mu[i,2], 1/thetastar[2,g[i]])
        trstprt[i] ~ dnorm(mu[i,3], 1/thetastar[3,g[i]])
        trstlgl[i] ~ dnorm(mu[i,4], 1/thetastar[4,g[i]])
        trstplc[i] ~ dnorm(mu[i,5], 1/thetastar[5,g[i]])
    #LATENT VARIABLES
        eta[i,1] ~ dnorm(mu_eta[i,1], 1/psi[1,g[i]])
        eta[i,2] ~ dnorm(mu_eta[i,2], 1/psi[2,g[i]])
        eta[i,3] ~ dnorm(mu_eta[i,3], 1/psi[3,g[i]])
    #LATENT MEANS
        mu_eta[i,1] <- nu[1,g[i]]
        mu_eta[i,2] <- dnu[2,g[i]]
        mu_eta[i,3] <- dnu[3,g[i]]
    #MU
        mu[i,1] <- alpha[1,g[i]] + lambda[1,g[i]]*eta[i,1]
        mu[i,2] <- alpha[2,g[i]] + lambda[2,g[i]]*eta[i,1] +
            dlam[2,g[i]]*eta[i,2]
        mu[i,3] <- alpha[3,g[i]] + lambda[3,g[i]]*eta[i,1] +
            dlam[3,g[i]]*eta[i,2]
        mu[i,4] <- alpha[4,g[i]] + lambda[4,g[i]]*eta[i,1] +
            dlam[4,g[i]]*eta[i,3]
        mu[i,5] <- alpha[5,g[i]] + lambda[5,g[i]]*eta[i,1] +
            dlam[5,g[i]]*eta[i,3]
    }
    ###PRIORS
    #ITEM INTERCEPTS
```

```
for(k in 1:K){
        alpha[1,k] ~ dnorm(0,0.01)
        alpha[2,k] ~ dnorm(0,0.01)
        alpha[3,k] ~ dnorm(0,0.01)
        alpha[4,k] ~ dnorm(0,0.01)
        alpha[5,k] ~ dsum(-alpha[1,k],-alpha[2,k],-alpha[3,k],-
            alpha[4,k])
        }
#ITEM LOADINGS
for(k in 1:K){
        lambda[1,k] ~ dnorm(1,0.01)
        lambda[2,k] ~ dnorm(1,0.01)
        lambda[3,k] ~ dnorm(1,0.01)
        lambda[4,k] ~ dnorm(1,0.01)
        lambda[5,k] ~ dsum(5,-lambda[1,k],-lambda[2,k],-lambda
            [3,k],-lambda[4,k])
}
#LATENT MEANS
for(k in 1:K){
        nu[1,k] ~ dnorm(5,0.01)
        dnu[2,k] <- 0
        dnu[3,k] <- 0
}
#LATENT VARIANCES
for(k in 1:K){
        psi[1,k] <- psisd[1,k]^2
        psisd[1,k] ~ dt(0,0.5,20) T(0,)
        psi[2,k] <- 1    psi[3,k] <- 1
}
#ERROR VARIANCES
for(k in 1:K){
```

```
        theta[1,k] <- ersd[1,k]^2
        theta[2,k] <- ersd[2,k]^2
        theta[3,k] <- ersd[3,k]^2
        theta[4,k] <- ersd[4,k]^2
        theta[5,k] <- ersd[5,k]^2
        ersd[1,k] ~ dt(0,0.5,20) T(0,)
        ersd[2,k] ~ dt(0,0.5,20) T(0,)
        ersd[3,k] ~ dt(0,0.5,20) T(0,)
        ersd[4,k] ~ dt(0,0.5,20) T(0,)
        ersd[5,k] ~ dt(0,0.5,20) T(0,)
}
#ERROR VARIANCES STAR
for(k in 1:K){
        thetastar[1,k] <- theta[1,k]
        thetastar[2,k] <- theta[2,k] - (sqrt(abs(rho[2,3,k])*
            theta[2,k]))^2
        thetastar[3,k] <- theta[3,k] - ((-1 + 2*step(rho[2,3,k
            ]))*sqrt(abs(rho[2,3,k])*theta[3,k]))^2
        thetastar[4,k] <- theta[4,k] - (sqrt(abs(rho[4,5,k])*
            theta[4,k]))^2
        thetastar[5,k] <- theta[5,k] - ((-1 + 2*step(rho[4,5,k
            ]))*sqrt(abs(rho[4,5,k])*theta[5,k]))^2
}
#CORRELATIONS
for(k in 1:K){
        rho[2,3,k] <- -1 + 2*rhodist[1,k]
        rho[4,5,k] <- -1 + 2*rhodist[2,k]
        rhodist[1,k] ~ dbeta(1,1)
        rhodist[2,k] ~ dbeta(1,1)
}
#PHANTOM LOADINGS
```

```
for(k in 1:K){
        dlam[2,k] <- sqrt(abs(rho[2,3,k])*theta[2,k])
        dlam[3,k] <- (-1 + 2*step(rho[2,3,k]))*sqrt(abs(rho
            [2,3,k])*theta[3,k])
        dlam[4,k] <- sqrt(abs(rho[4,5,k])*theta[4,k])
        dlam[5,k] <- (-1 + 2*step(rho[4,5,k]))*sqrt(abs(rho
            [4,5,k])*theta[5,k])
}
#INFERENTIAL COVARIANCES
for(k in 1:K){
        thetacov[2,3,k] <- dlam[2,k]*dlam[3,k]*psi[2,k]

        thetacov[4,5,k] <- dlam[4,k]*dlam[5,k]*psi[3,k]
}
#LOG DENSITY
for(k in 1:K){
cov.mat[1,1,k] <- lambda[1,k]^2 * psi[1,k] + theta[1,k]

cov.mat[1,2,k] <- lambda[1,k] * lambda[2,k] * psi[1,k]
cov.mat[1,3,k] <- lambda[1,k] * lambda[3,k] * psi[1,k]
cov.mat[1,4,k] <- lambda[1,k] * lambda[4,k] * psi[1,k]
cov.mat[1,5,k] <- lambda[1,k] * lambda[5,k] * psi[1,k]
cov.mat[2,1,k] <- lambda[2,k] * lambda[1,k] * psi[1,k]
cov.mat[2,2,k] <- lambda[2,k]^2 * psi[1,k] + theta[2,k] + dlam
    [2,k]^2
cov.mat[2,3,k] <- lambda[2,k] * lambda[3,k] * psi[1,k] +
    thetacov[2,3,k] + dlam[2,k]*dlam[3,k]
cov.mat[2,4,k] <- lambda[2,k] * lambda[4,k] * psi[1,k]
cov.mat[2,5,k] <- lambda[2,k] * lambda[5,k] * psi[1,k]
cov.mat[3,1,k] <- lambda[3,k] * lambda[1,k] * psi[1,k]
cov.mat[3,2,k] <- lambda[3,k] * lambda[2,k] * psi[1,k] +
```

```
    thetacov[2,3,k] + dlam[2,k]*dlam[3,k]
cov.mat[3,3,k] <- lambda[3,k]^2 * psi[1,k] + theta[3,k] + dlam
    [3,k]^2
cov.mat[3,4,k] <- lambda[3,k] * lambda[4,k] * psi[1,k]
cov.mat[3,5,k] <- lambda[3,k] * lambda[5,k] * psi[1,k]
cov.mat[4,1,k] <- lambda[4,k] * lambda[1,k] * psi[1,k]
cov.mat[4,2,k] <- lambda[4,k] * lambda[2,k] * psi[1,k]
cov.mat[4,3,k] <- lambda[4,k] * lambda[3,k] * psi[1,k]
cov.mat[4,4,k] <- lambda[4,k]^2 * psi[1,k] + theta[4,k] + dlam
    [4,k]^2
cov.mat[4,5,k] <- lambda[4,k] * lambda[5,k] * psi[1,k] +
    thetacov[4,5,k] + dlam[4,k]*dlam[5,k]
cov.mat[5,1,k] <- lambda[5,k] * lambda[1,k] * psi[1,k]
cov.mat[5,2,k] <- lambda[5,k] * lambda[2,k] * psi[1,k]
cov.mat[5,3,k] <- lambda[5,k] * lambda[3,k] * psi[1,k]
cov.mat[5,4,k] <- lambda[5,k] * lambda[4,k] * psi[1,k] +
    thetacov[4,5,k] + dlam[4,k]*dlam[5,k]
cov.mat[5,5,k] <- lambda[5,k]^2 * psi[1,k] + theta[5,k] + dlam
    [5,k]^2
sigma[1:5,1:5,k] <- inverse(cov.mat[1:5,1:5,k])
}
for(i in 1:N){
data.mv[i,1:5] <- c(trstprl[i],trstplt[i],trstprt[i],trstlgl[i
    ],trstplc[i])
mean.trstprl[i] <- alpha[1,g[i]] + nu[1,g[i]] * lambda[1,g[i]]

mean.trstplt[i] <- alpha[2,g[i]] + nu[1,g[i]] * lambda[2,g[i]]

mean.trstprt[i] <- alpha[3,g[i]] + nu[1,g[i]] * lambda[3,g[i]]

mean.trstlgl[i] <- alpha[4,g[i]] + nu[1,g[i]] * lambda[4,g[i]]
```

```
        mean.trstplc[i] <- alpha[5,g[i]] + nu[1,g[i]] * lambda[5,g[i]]
        mean.mv[i,1:5] <- c(mean.trstprl[i],mean.trstplt[i],mean.
            trstprt[i],mean.trstlgl[i],mean.trstplc[i])
        log_lik[i] <- logdensity.mnorm(data.mv[i,], mean.mv[i,], sigma
            [,,g[i]])
        }
}
```