Regression with Dependent Observations: Random and Fixed Effects Models

Conrad Ziller, *Universität zu Köln*

Keywords

Random Effects, Fixed Effects, Panel Analysis, Hybrid Models, Multi-level Analysis

Abstract

Many data sets employed in comparative political science are spatially and / or temporally structured. Spatial and temporal structuring typically lead to statistical dependence that need to be taken into account for data analysis. This article introduces random-effects models (RE) and fixed-effects models (FE) as analytical tools dealing with dependent observations. This review discusses the problem of statistical dependence and introduces RE and FE models. In addition, it provides decision heuristics and hints for practical application.

I thank Carl Berning, Hans-Jürgen Andreß, Merlin Schaeffer and the editors for helpful comments.

Ziller, C. (im Erscheinen). Multiple Regressionen mit voneinander abhängigen Beobachtungen: Random-Effects und Fixed-Effects. In: Wagemann/Goerres/Siewert (Hg.): Handbuch Methoden der Politikwissenschaft. Springer.

1. Introduction

An increasing number of political scientific research papers employ international comparative data sets. These are, for example, data from the European Social Survey (ESS), the International Social Survey Program (ISSP) or the Comparative Study of Electoral Systems (CSES). Similarly, panel data and aggregated and comparative time-series data (Time Series Cross-Sectional Data, TSCS[1]) are a popular data types for empirical-comparative social research. If data is spatially and / or temporally structured or grouped, *statistical* dependencies are a likely consequence that needs to be dealt with in empirical models.

More specifically, if data with dependent observations are naively pooled and analyzed using ordinary least squares regression models (OLS), this may lead to biased standard errors and / or regression coefficients. The reason is that the estimation of the regression parameters in the OLS-method is based on specific assumptions regarding statistical independence of observations. If these conditions are met, OLS regression models can be used (CROSS-REFERENCE SENG). If statistical dependence is given, so-called random-effects (RE) or fixed-effects (FE) methods can be used instead. While terms "fixed" and "random" have various meanings in the literature (see Gelman 2005, p. 20-21), they are used here in order to distinguish specific types of statistical modeling. While FE refers to models that use dummy variables to control for group-related differences in the dependent variable, random effects refers to models that take group-related differences through a random variable[2] into account.

The aim of this review is to provide a practical introduction to RE and FE models, to present advantages and disadvantages of the individual procedures and to highlight recent developments and debates. For further technical details of RE and FE models see e.g. Andreß et al. (2013), Wooldridge (2009: chapter 14) and on FE models Brüderl and Ludwig (2015). For the sake of simplicity, this review focuses on continuous dependent variables and linear relationships between variables. RE and FE models may also incorporate other dependent variables (e.g. binary variables or count variables) applying appropriate model specifications (e.g. logistic, ordinal-logistic or Poisson models) (see Snijders and Bosker 2012, Chapter 17, Andreß et al., 2013, chapter 5 for further details).

[1] In comparison to panel data, TSCS data are aggregate data in which the number of time points T is often greater than the number of units N (T> N). Both panel and TSCS data have similar properties with regard to statistical dependence (for specifics of TSCS data see Beck and Katz 1995).

[2] It should be noted that even in RE models, the estimated coefficients are sometimes referred to as "fixed effects" (e.g., in the context of multi-level analysis, CROSS-REFERENCE PÖTSCHKE). This should show that the effects are assumed to be identical or fixed for all groups in the investigated population. By contrast, the terms RE and FE here refer to model classes with distinct properties.

In the remainder of this review, the problem of dependent observations will be discussed. The paper then explains the basics of RE and FE models and shows possible heuristics and decision rules for model selection. The conclusion includes practical hints and insights into current methodological debates.

2. The problem of interdependent observations

2.1 Statistical dependence

A widespread phenomenon in empirical political science is the usage of data that is grouped along certain higher-level units (also referred to as contexts or macro-units), also known as "clustered" or "nested" observations. For example, if we use cross-sectional survey data of individual respondents, spatial structures such as neighborhoods, communities, regions, and countries may be relevant groups.[3] Similarly, data measured over time represents a clustered data structure. For panel data (e.g. repeated measurements of the same individuals), the time-point specific observations are grouped in the respective individuals (i.e. person is the group characteristic here). Using standard regression techniques (such as OLS regression) to model such data structures can lead to biased standard errors and / or regression coefficients. The reason is that these methods assume statistical independence of observations. But what does statistical independence / dependence actually mean? Generally speaking, statistical independence means that observations occur independently of each other and thus each observation contains independent information. Statistical dependence, on the other hand, means that observations (within certain groups) are systematically related, which may violate assumptions of statistical models.

Certain types of data are prone to convey statistically dependent observations. Panel data is structured time-wise and it is likely that current and future observations of characteristics of a person are similarly or correlated. An example is the membership in a political party at time t0, which probably still exists in t1. One reason for this may be that people who are high in political interest are (consistently) party members, while those with little interest do not become members at all. For cross-sectional data from multi-stage random samples (e.g., randomly selected individuals from randomly selected countries), historical or institutional country

[3] Groups typically have no intrinsic information content, are interchangeable, and contain a large number of categories (analogous to the definition of higher levels in the multilevel analysis literature). Characteristics that do not meet these criteria (e.g., gender) are considered as variables, not groups.

differences may cause certain respondent characteristics to be similar within countries. In the case of cross-sectional data that originate from a simple random sample, there is generally no or only limited statistical dependence - looking at the whole sample. However, if the aim is to explain, for example, regional differences of particular characteristics dependencies may also exist due to the belonging of respondents to specific spatial contexts (i.e., federal states).[4] The grouping of the data then follows a theoretically informed interest in explaining group-related differences - a procedure that is common in analyzing context effects in multi-level analyses (CROSS-REFERENCE PÖTSCHKE).

For further illustration prototypical data from a fictional one-time survey of citizens on satisfaction with the current government are used. This survey was conducted in 20 countries and has a sample size of 1,000 respondents per country (20,000 respondents in total). The dependent variable is government satisfaction (y), which was surveyed using an 11-point response scale (from 0 = "extremely dissatisfied" to 10 = "extremely satisfied"). It is also assumed that respondents' characteristics within countries are systematically similar and that there are mean differences in government satisfaction across countries. Formally, the variable government satisfaction consists of the following parts:

$$y_{ij} = \alpha_0 + u_j + e_{ij}. \tag{1}$$

Each observational value y_{ij} (of a respondent *i* within a country context *j*) is determined by a total mean value α_0, a country-specific effect u_j and a residual effect e_{ij}. The country-specific effect u_j represents the deviation of the country mean (of country *j*) from the total mean. The residual effect e_{ij} in turn represents the deviation of a measured value (of respondent *i*) from the country mean (of country *j*). The total variance of the variable government satisfaction Var (y_{ij}) can hence be subdivided into the variance of the group effect σ_u^2 (so-called *between-group variance*) and the variance of the intra-group error or residual σ_e^2 (so-called *within-group variance*). The share of the group-related variance of the total variance is a central measure of statistical dependence, which is referred to as *serial correlation* (in the case of panel data also

[4] In addition to such *spatial-categorical* dependence, there may also exist forms of *spatial-continuous* dependence in which spatially close observations are similar to each other and this similarity decreases with increasing spatial distance. In such cases, special spatial regression methods are used in which surrounding observations are weighted according to their distance from each other.

as autocorrelation). It corresponds to the so-called intra-class correlation coefficient (ICC)[5] (Andreß et al., 2013, pp. 77f, Snijders and Bosker 2012, pp. 17ff.). The larger the ICC, the greater the similarities within the groups, and thus the larger the average differences between groups.

The discussion below refers to how statistical methods deal with statistic dependency. Main differences are largely on how the group effect u_j is modelled which is also known as *unobserved heterogeneity at the group level*. Before that, I outline specific consequences of statistical dependence within the OLS regression framework.

2.2 Violation of the assumption of uncorrelated errors

Methods such as OLS regression are based on the assumption that each observation contributes independent information, which implies that there is *no* serial correlation. The assumption thereby refers to serial correlation of errors—that is, the variance of the dependent variable that is not explained by the predictors in the model.[6] If serial correlation is present, the coefficients of an OLS regression might be unbiased,[7] but the standard errors are in any case biased. This in turn affects the validity of statistical testing procedures. Since the errors are usually positively correlated, there is an underestimation of the standard errors and correspondingly overestimated *F*- and *t*-statistics (Moulton 1986). In other words: we detect statistically significant relationships that actually do not exist. As underlying process, serial correlation might be due to (unobserved) group-specific characteristics. For panel and TSCS data, intra-group factors (i.e. time-varying processes) are also possible causes.

To determine serial correlation, an ICC can be calculated. Corresponding routines are implemented in regular data analysis programs[8] and if the variance parameters have been estimated, an ICC can be easily calculated by hand. In addition, the Breusch Pagan test can provide an explicit tests statistic for whether or not systematic group-related differences exist.

---

[5] Formally, this measure indicates the correlation of values of two randomly chosen observation units of the same (randomly selected) group *j*: $Corr(y_{ij}, y_{kj}) = \frac{Cov(y_{ij}, y_{kj})}{\sqrt{Var(y_{ij})} * \sqrt{Var(y_{kj})}} \triangleq \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}, \ i \neq k.$

[6] The error of observation i therefore may not have an effect on the the error of observation k. Formally expressed: $Corr(e_i, e_k) = 0$, für alle $i \neq k$. This means, the expected value of all paired products of errors is zero.

[7] Unbiasedness means that the average of independent replications of an estimate (with repeated sampling) is equal to or at least very close to the actual population value. Consistency means that an estimator becomes more and more accurate with increasing sample size.

[8] For example, in the Stata output of the "xtreg var1 var2, etc., re"-command, "rho" indicates the proportion of unexplained variance of the group level of the particular model.

In addition, the Wooldridge test for serial correlation in panel data can be used to determine if panel or TSCS data have serial correlation in the intra-group error e_{ij} (also called the idiosyncratic error) (Drukker 2003).

In order to obtain correct standard errors despite serial correlation, one can first of all control for factors that are causally underlie the process of serial correlation. In the example on individual party membership, political interest of individuals could be such a factor. Regarding government satisfaction, institutional and economic country characteristics could be relevant. Since suitable indicators are often not available, RE models correct for serial correlation due to group-specific characteristics by explicitly modeling the group effect u_j. In a related vein, FE models eliminate serial correlation by controlling group-related deviations by design. Another possibility is to estimate cluster- respectively panel-robust standard errors that take into account the empirical variances of errors for each group and thus generally correct for forms of serial correlation and heteroscedasticity.[9] Robust standard errors are typically larger than conventional standard errors (see also chapter 5.2). Especially for panel data and serial correlation due to time-varying processes, generalized estimating equations are also considered.[10]

In summary, there are a number of effective counter measures in order to address serial correlation and the related consequences of biased standard errors and test statistics. Furthermore, statistical dependence may also affect the exogeneity assumption, which is another reason for especially using FE models.

2.3 Violation of the exogeneity assumption

Another assumption of the OLS regression framework which may be violated due to statistical dependence is exogeneity of predictor variables. It states that errors and predictor variables must be uncorrelated. Unmeasured characteristics (such as personality traits or genetic influences) likely to influence the explanatory variable must not be correlate with the measured characteristics in the regression equation (Vaisey and Miles 2017, p. 47). In practice and under conditions of imperfect measurement and availability of indicators, this assumption is more

[9] If the homoscedasticity assumption (uniform distribution of the error terms) is violated, this also affects the unbiasedness of standard errors and the validity of test statistics. Violations of this assumption can be checked, for example, using the White-test.

[10] Typically, serial correlation is assumed to be strongest between two adjacent time points and then decrease with increasing time interval. This error structure, which is also referred to as first-order autocorrelation (AR (1)), can be modeled, for example, in Stata with the commands "xtgee var1 var2 etc., corr (ar1)" or "xtregar var1 var2 etc.".

than heroic. The consequences of a violation of the exogeneity assumption can be severe. In contrast to violating the assumption of uncorrelated errors, not only the variance parameters (i.e. standard errors and thus test statistics) are affected, but the unbiasedness the coefficients. A coefficient for which we cannot be sure of the magnitude and direction lacks informational content.[11] The main reason for a violation of the exogeneity assumption is the omission of relevant variables (omitted variables bias).[12] If all relevant variables were included in the model, the dependent variable would be fully explained and no errors would remain that could correlate with the predictors. On the other hand, non-measured or unobserved characteristics that are relevant in this regard can be considered as typical rather than exemption in applied empirical social research.

For further illustration, we refer to the cross-national survey data set on government satisfaction described above. Analogous to theories that explain the evaluation of government performance with economic motives (Lewis-Beck and Stegmaier 2000, Tilley et al., 2017), it is assumed that persons with a positive assessment of their personal income situation are more satisfied with the current government performance than persons with financial problems. The assessment of the personal income situation is measured with an ordinal-scaled variable x (0 = "do not get along / hardly with income", 1 = "get along with income", 2 = "live comfortably with income"). To test the relationship, all 20,000 respondents are pooled together and analyzed using OLS regression. A corresponding linear-additive regression equation has the following functional form:

$$y_i = \beta_0 + \beta_1 x_i + e_i. \tag{2}$$

Here, β_0 is the intercept (also called the constant) that indicates the average value of government satisfaction when x (income-satisfaction) is zero. β_1 is the coefficient[13] indicating the effect of the variable income-satisfaction. e represents the deviations of the individual observations from the regression line respectively the error term. In the case illustrated here, the regression equation refers to individual i ($i = 1, \dots, n$). In order for the coefficient β_1 to be unbiased, a number of assumptions need to be met (Verbeek 2004, p. 16; Wooldridge 2009, pp. 47-52). The most important assumption is that of exogeneity, i.e. that the error term e has an expected

[11] It should be stressed that the degree of bias depends on the level of correlation between predictor and error term (see the simulation studies in Clark and Linzer 2015 and Vaisey and Miles 2017).
[12] Other causes of a violation of the exogeneity assumption, also referred to as endogeneity bias, are measurement errors and mutual causality.
[13] Also referred to as estimator, effect, slope parameter, and regression weight.

value of zero independent of the value of the predictor variable x: $E(e|x) = 0$. Put differently, error term and predictor variable are uncorrelated.[14]

In the present example it is quite likely that the exogeneity assumption does not hold. For example, non-measured personality differences (e.g., individual agreeableness) might be related to both personal income satisfaction and government satisfaction. The consequence is that the variable satisfaction with income partially conveys the effect of the personality trait agreeableness. The coefficient of the variable income satisfaction is therefore biased.[15]

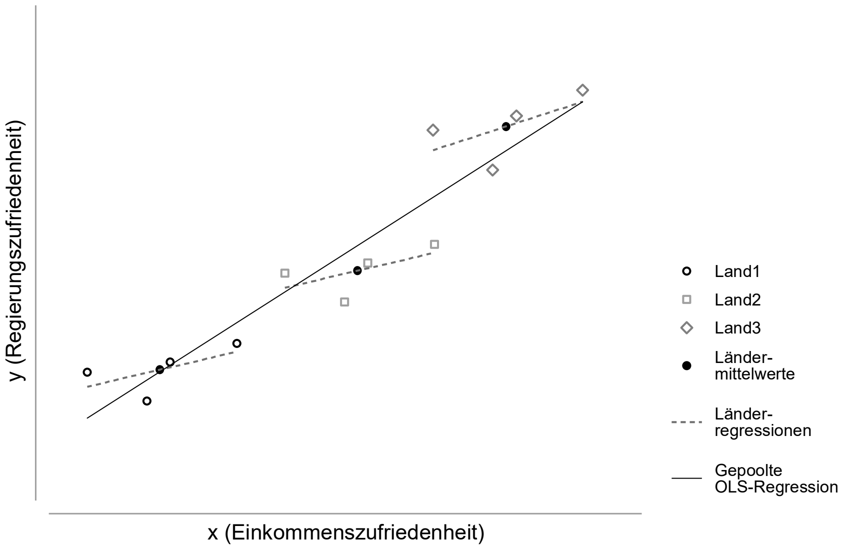
In addition, systematic differences in the dependent variable between groups (unobserved heterogeneity at the group level) may cause a violation of the exogenous assumption. Figure 1 graphically illustrates unexplained group-related differences in the dependent variable. As in the example above, government satisfaction should be explained by the personal income assessment of respondents. Looking at four randomly selected units in country 1, there is a moderate positive correlation between personal income situation and government satisfaction. This is also the case in the other two country contexts. However, there are significant differences between countries regarding the level of dependent variables. Respondents in Country 3 have significantly higher average government satisfaction than those in Country 1. If these differences in levels are neglected in pooled OLS regression, the effect of income satisfaction on government satisfaction β_1 is much stronger (steeper slope of the regression line) than one would expect from the correlations within the particular countries. The reason is that the mean differences represent unobserved country characteristics that are "transported" by the individual variable income satisfaction.

For example, we can assume that the observed country differences in government satisfaction can be attributed to differences in country-specific economic growth. Without the inclusion of this variable, the individual variable conveys the effect of country-specific economic growth, given that in this example personal income satisfaction is also influenced by economic growth (and related increases in incomes). The consequence is a biased estimate of the income satisfaction variable, in this case an overestimation.

[14] In addition to the exogeneity assumption, the linearity of the parameters is assumed, there has to be no perfect multicollinearity between the variables, and x must have varying values. In order that the standard errors (and thus test statistics) are unbiased, the errors must have an expected value of zero, be normally distributed, and the variance of the errors must be constant for all values of x (homoscedasticity). In addition, no serial correlation is assumed. Another assumption is that the observations reflect a random sample from a population, which is necessary for the validity of statistical inference.

[15] If agreeableness was the only omitted variable and there was a suitable indicator that we could include in the equation, the effect of income satisfaction would be partially controlled for agreeableness. The exogeneity assumption would therefore hold and the coefficient for income satisfaction would be unbiased.

Figure 1: Pooled OLS regression with grouped observations



There are no direct tests to detect the violation of the exogenous assumption due to omitted variables in the model.[16] However, group-related mean differences in the dependent variable indicate unobserved heterogeneity, which should either be modeled or controlled. RE- and FE-models are suitable for this.

3. Random-Effects- and Fixed-Effects-models

If there is statistical dependence due to group-related unobserved heterogeneity, the assumption of uncorrelated error terms is in any case violated and - without appropriate countermeasures - standard errors and test statistics from the OLS regression are biased. RE- and FE-models eliminate such forms of serial correlation. Similar to the OLS method, RE-models produce biased coefficient estimators if the exogeneity assumption is violated due to correlated group-related heterogeneity. In contrast, FE-models are not prone to correlated group-related heterogeneity, but have other limitations.

[16] Although tests such as the Ramsy RESET test indicate misspecifications in terms of variables which are already in the model, they do not indicate whether omitted variables are relevant. Which omitted variables are relevant should therefore primarily be based on theoretical considerations.

3.1 Random-Effects-models

The general functional form of the RE-model is given as:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij}. \tag{3}$$

The difference to the OLS-model is that the group level is taken into account by the notation j in the model. The most important extension is that the intercept β_0 can now vary over groups (or countries) j. Modeling a variable intercept takes into account the phenomenon tha, government satisfaction (y) is higher on average in some countries than in others. Furthermore, β_{0j} can be subdivided into an average intercept over all groups β_0 and the respective country-specific deviations u_j:

$$\beta_{0j} = \beta_0 + u_j \tag{3.1}$$

Both equations can be integrated into one:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}. \tag{4}$$

This model is equivalent to a *random intercept* multilevel model (also referred to as hierarchical models or *mixed models*, CROSS-REFERENCE PÖTSCHKE). u_j is the error term of the group level (unobserved group level heterogeneity) and is conceptualized as a normally distributed random variable in which the group effects can be interpreted as random deviations from the overall mean. e in turn represents the deviations of the individual observations in the respective groups. The estimation of the parameter values in the RE-model is based on the generalized least squares (GLS) method, in which a certain proportion of each variable's mean is subtracted from original variable's values (also referred to as quasi-demeaning). The fraction is thereby determined by a transformation parameter (theta), which in turn depends on the extent of group-related heterogeneity as well as the number of periods (for panel data) or the average number of cases per group (grouped cross-sectional data) (see Andreß et. al. 2013, p. 154). As an alternative to GLS, the *maximum likelihood* method (ML) can be used. Beginning with certain starting values, the combination of parameters is sought which iteratively increases the value of the likelihood function until its maximum. This maximum value indicates the combination of parameters for which the realization of the observed data is most likely.

Similarly to OLS-models, RE-models estimate parameters based on both between-group and within-group variance.[17] However, OLS-models place more emphasis on the variance between groups than RE-models. The reason is that for the OLS-model the variance between groups is completely bound in the predictor variables, whereas in the RE-model it is bound in the error term u_j (Greene 2003, p. 295f.). Hence, the results of RE models usually lie between those of an OLS estimation and FE models that are based on a within-variance only (see below). The extent to which the RE estimation is weighted towards the FE estimator largely depends on the average group size. The higher the number of periods (panel data) or the average number of cases per group (cross-sectional data), the more the RE estimation corresponds to a within-variance-based FE estimation (cf. calculation of the transformation parameter theta).

Figure 2 shows the relevant components of a RE-model. Again, government satisfaction is the outcome and the assessment of personal income is the predictor. While the regression coefficient from pooled OLS was very close to the between estimator of the group averages (Figure 1), the RE coefficient β_1 is now more weighted towards the country regressions (i.e., the within estimator). The effect of income satisfaction is now partially decoupled from the group level and thus represents – compared to the OLS estimator – a rather unbiased results as it controls for unobserved heterogeneity.

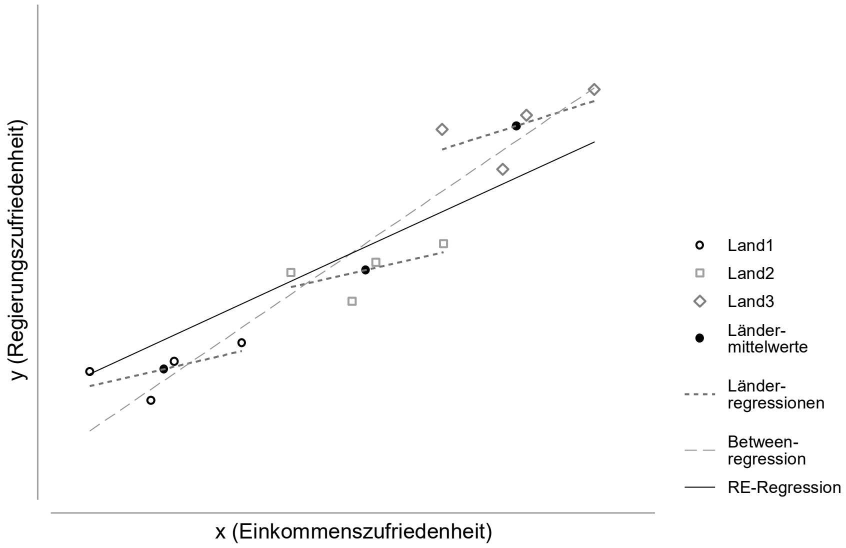
Similar to multi-level analysis, RE-models can include group-level variables (also called macro variables or context variables) in order to explain group-specific deviations from the overall mean u_j.[18] By including country-specific economic growth as a contextual variable z_j, we expect a reduction in group-specific heterogeneity (compared to a model without this variable):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + u_j + e_{ij} \tag{5}$$

[17] Analogously, the so-called between-estimator refers to a regression which includes only the group averages. The within estimator refers to regressions with observations of the respective groups. Within estimators are equivalent to coefficients from FE models (see chapter 3.2).

[18] If variables of the group level (macro variables) are included in an OLS model, as a rule estimators with too low standard errors result, which in turn leads to invalid test statistics (Moulton 1986).

Figure 2: RE model with grouped observations



In addition, it is possible to vary the effects of intra-group variables (here individual variables) across groups or contexts (so-called *random slopes*). These variations can be explained in a further step by a group variable with so-called *cross-level interactions*.

To provide an example from political science: Hakhverdian and Mayne (2012) examine the effect of education on citizens' political trust (in European comparison). In addition, the authors are interested in whether the effect of education varies across country contexts (= random slope) and whether possible differences can be explained by degrees of political corruption (as a country characteristic) (= cross-level interaction). For this, they use survey data from the European Social Survey and find that 27 percent of the variance in political trust is due to country differences (= serial correlation due to group-related heterogeneity). This (as well as the intention to test country characteristics) justifies the use of a RE or multi-level model. Already 10 percent of the group-related variance can be explained by individual variables in the model. Country characteristics and the cross-level interaction explain further group-related variance, so that in the final model only three percent of the remaining total variance can be attributed to non-modeled country differences. As a result, the authors show that the effect of education on political trust varies across countries and interacts with the country characteristic of corruption. Education is negatively associated with political trust in countries with high levels of corruption (the more education, the less trust), while there is a positive link in low corruption countries (the more education, the more trust).

With regard to the model assumptions, RE models assume that error terms are uncorrelated with (intra-group and group-related) predictor variables (exogeneity assumption).[19] If this assumption is not met, the estimators in the RE model will be biased. Again, an omission of relevant variables is a main reason for the exogenous assumption to be violated. In the given example, such an omitted group-related variable could be another economic factor, such as unemployment rates. In summary, RE models take group-related heterogeneity into account with an additional group-level error term. RE models also allow group level variables to be included in the estimation, but just like OLS methods they assume uncorrelatedness between (group-related) unobserved heterogeneity and predictor variables.

3.2 Fixed-Effects-models

In contrast to RE-models, a correlation between predictors and group-related error terms is irrelevant in FE-models, as they eliminate the entire unobserved group-level heterogeneity by design. The basic version of a FE model is estimating an OLS regression with a dummy variable D_j for each group j (and the individual i grouped therein) such that $D_{j[i]} = 1$, if individual i belongs to group j and $D_{j[i]} = 0$ if not.[20]

$$y_{ij} = \beta_0 + \sum_{j=1}^{J-1} \beta_j \, D_{j[i]} + \beta_1 x_{ij} + e_{ij}. \tag{6}$$

The dummy variables absorb the complete variance between groups, and thus the predictor variables can no longer correlate with group differences in the dependent variable. Controlling the variance between groups leads to regression coefficients that are estimated using within-group variance only. For cross-national survey data measured at one time point, level differences in the dependent variable can be absorbed by including country dummies. An estimation of coefficients is therefore based on variances within countries. Unobserved heterogeneity between countries as a result of, for example, historical or institutional

[19] In addition, also in RE models linearity of the parameters, no perfect multicollinearity, error terms (also the error term of the group level) with an expected value of zero, normally distributed error terms, constant variance of the error terms, no serial correlation of the errors, and uncorrelatedness of the errors (u, e) are assumed. Observational units should be drawn from a random sample of the population. In addition, it is assumed that also the group-level units are a random selection from a population (or a universe of groups). If countries represent the group unit, this assumption is not plausible. Therefore, the results in this case should not be interpreted as predictions about underlying populations, but only in terms of country sample. Whether one should even bother performing statistical inference for the parameters at the group level is an open question.

[20] J-1 dummies are estimated because a group is omitted as a reference category.

differences is controlled for by taking into account the country-specific mean differences (using the dummy variable). As an alternative to regression with dummy variables, FE models can also be specified by centering the variables on the respective group average.[21] This procedure is also referred to as *mean differencing* or *demeaning* and leads to equivalent results as with the use of dummy variables (Andreß et al., 2013, pp. 133):

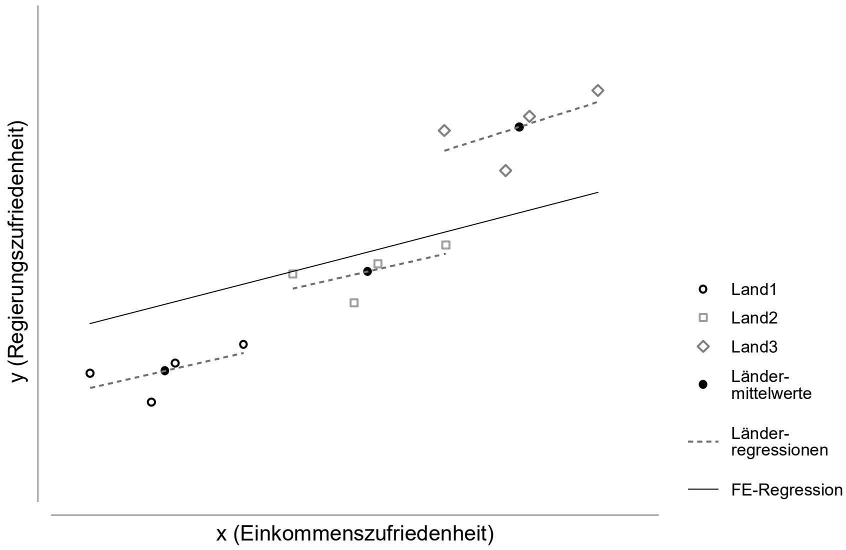
$$(y_{ij} - \bar{y}_j) = \beta_1(x_{ij} - \bar{x}_j) + (e_{ij} - \bar{e}_j). \quad (7)$$

Figure 3 refers to the example developed above and illustrates the estimate of a separate intercept per country, which absorbs country differences in the dependent variable government satisfaction. The estimated coefficient of income satisfaction β_1 is the average of the slope parameters from the three country regressions. Compared to the OLS regression and RE-models, the estimator β_1 is smaller, but it corresponds to the unbiased effect, since group differences in the dependent variable, which in our example are associated with economic growth and were transported by the predictor variables, are controlled for. What remains is an estimated association at the individual level free of unobserved heterogeneity of the group level. It should be emphasized that although unobserved heterogeneity at the group level no longer plays a role, omitted individual characteristics (or unobserved heterogeneity within groups) may nevertheless confound the relationship under study and possible bias the results. Accordingly, relevant control variables within the groups should be taken into account as well.[22] An example from comparative politics is the study by Ziller and Schübel (2015), which examines the relationship between personal corruption experience, political trust, and voting for right-wing populist parties. For the analysis of the individual relationships, the authors analyze survey data from 12 European countries (European Social Survey). 21 percent of the variance of the dependent variable right-wing populist voting are due to country differences. In order to control for unobserved heterogeneity at the country level, which may bias the individual relationships, country dummies are included in the analyses. One result of the study is that individual corruption experience is systematically related to lower political trust, which in turn increases the probability of choosing a right-wing populist party.

[21] Programmed FE routines in statistical programs (for example "xtreg var1 var2 etc., fe" in Stata) usually use this variant as a specification.

[22] In addition to the exogeneity of the predictor variables, linearity, no perfect multicollinearity and certain properties of the errors (expected value of zero, normal distribution, constant variance, no serial correlation) are assumed in the FE-model. Observational units should be drawn from a random sample of the population (for panel data at least at the beginning of the observation period) in order for statistical testing to be valid.

Figure 3: FE-model with grouped observations



In summary, FE-models, compared to RE-models, have the advantage of avoiding biased estimators due to unobserved group-level heterogeneity. In the case of panel data, the fixed effects absorb the variance between persons, and for cross-country data the variance between countries. As a result, group-level variables cannot be included. These and other limitations are therefore also taken into account in the following discussion of criteria of model selection.

4 Criteria for model selection

4.1 Selection heuristics and Hausman Test

If group-related unobserved heterogeneity exists, the exogeneity assumption may be violated due to omitted variables. Whether group-related heterogeneity can be determined by calculating the ICC. One scenario is that variables included in the model (almost) completely explain group-related heterogeneity and the ICC goes (near) zero. In such a case, an OLS model with adaptation of the standard errors for serial correlation could be estimated (see Van der Brug et al., 2007 as an example). Often, however, not all relevant variables are available. RE-models model group-related variance with an additional error term, but assume its uncorrelatedness with variables in the model. Omitting relevant group level predictor variables will violate this assumption and biased regression coefficients are the result. FE-models control group

heterogeneity by design and if differences between RE and FE models occur, these differences indicate correlated group-related heterogeneity and that the exogeneity assumption is violated.[23]
The Hausman test (Hausman 1978) formalizes a comparison of the RE- and FE-models by calculating the standard error of the difference of the parameter estimates of both models. The null hypothesis is on the independence of group-related error term and predictor variable(s). A statistically significant test result thus rejects this assumption and gives reason to estimate a FE-model. If the test is not significant, a RE-model can be estimated. Nevertheless, even with a non-significant result, it cannot be ruled out that there are small correlations between group-related error terms and predictor variables.[24]
The question then arises under what circumstances it makes sense to estimate RE-models. Two main reasons exist: A lower efficiency of FE-models and the impossibility to estimate group variables in the FE-model.

4.2 Low efficiency of FE-models

Assuming the rather untypical case of correlated unobserved heterogeneity and the Hausman test shows no systematic differences between the two models, the use of RE-models is recommended. The reason is that FE-models only use variance within groups and estimates are therefore less efficient, i.e. its standard error is larger. In addition, an estimation of one dummy variable per group leads to a considerable loss in degrees of freedom. Also in the case of mean differentiation, degrees of freedom are lost for each mean value estimated from the data. The consequence of low degrees of freedom is low statistical power in detecting significant effects. Depending on the data structure, low efficiency and power are more or less relevant. When analyzing comparative survey data with countries as a group level and large-N samples, there is usually enough variance within countries, and the inclusion of 25 or 30 dummy variables will not lead to a significant loss in power. For panel data with many respondents (i.e., a high number

[23] As emphasized in the remarks on RE-models, the extent to which a RE-estimate is weighted toward the FE estimator depends on the number of periods (panel data) and the average number of cases per group. For cross-sectional data with a high number of cases per group (country-comparative survey data), therefore, the estimators of RE- and FE-models will hardly deviate from each other. In the case of panel data (especially with a low number of waves), RE-models are more based on between-group variance.
[24] Clark and Linzer (2015, p. 405) found in a simulation study that even with the smallest correlations between group-related error terms and predictor variables, FE-models are preferable to the use of OLS or RE-models. Only in a (practically unlikely) scenario where correlated unobserved heterogeneity was not very large, the number of cases within the groups was low (<20), and the variance within the groups was very low (also referred to as *slow moving* or *sluggish data*), RE-models had a lower bias than FE estimates.

of fixed effects) and low variability in measured characteristics over time, the inefficiency is more prevalent.

In relation to the risk of biased coefficients, decreased efficiency of FE-models is a low price to pay. At the same time, in applying FE-model one should be aware of specifics of the data structure, such as low average number of cases or low variability within groups. Therefore, if a Hausman test implies that FE- and RE-models produce equivalent results, then a RE-model should be estimated. In practical application, however, this will rarely be the case (especially with panel data) and the use of FE-models remains the benchmark for the analysis of grouped data.

4.3 Inclusion of group characteristics and Hybrid-models

Since all group differences are absorbed in FE-models, the effects of group characteristics cannot be estimated even if they are of theoretical interest. For cross-country survey data, this could be the influence of country characteristics such as democratization. For panel data, the influence of personal characteristics such as gender or ethnic background (group time constant properties) cannot be estimated in FE-models.[25] Depending on the research question, this can be an enormous limitation. A circumvention of this potential drawback is the so-called hybrid-model, which combine RE-alike and FE estimates (see Allison 2009, pp. 23ff and Andreß et al., 2013, pp. 157ff.).

Hybrid-models are basically RE or multilevel models, which use centered or disaggregated predictor variables to simultaneously estimate *between* components, which corresponds to the results of a regression by group averages (*between*-regression), and *within* components, which corresponds to the results of FE regression. This allows group-level variables to be included in the model, while preserving the FE benefits for variables with variance within groups. With regard to cross-national survey data, in which respondents are grouped in countries, individual-level effects can be estimated free of unobserved heterogeneity at the group level (equivalent to the FE-model) while at the same time group characteristics can be included (e.g. degree of democratization). For panel data, relationships over time (e.g. whether a change in the use of social media affects a change in political engagement) are modeled as FE estimates, and at the

[25] In fact, group properties can be interacted with intra-group variables. However, this does not reflect the average influence of a group characteristic, but rather whether and how much the effect of an intra-group variable varies across different values of the respective group characteristic.

same time, time-constant personal characteristics, such as the gender of a person, can be included.

The technical illustration is based on panel data for which between-variance refers to time-constant differences between respondents j and within-variance refers to time-varying characteristics measured at time i (i.e., "within" respondents over time). A hybrid-model is given as:

$$y_{ij} = \beta_0 + \beta_1(x_{ij} - \bar{x}_j) + \beta_2\bar{x}_j + \beta_3 z_j + u_j + e_{ij}. \tag{8}$$

Where x_{ij} is a time-varying variable and $\bar{x}_j$ is the average of this variable for person j. Before analysis, the average of the variable is demeaned from the respective original values of the variable $(\mathrm{x_{ij}} - \bar{\mathrm{x}}_\mathrm{j})$ is carried out in advance.[26] Thus, β_1 represents the within estimator, which has the same properties as an estimator of a FE-model (no unobserved group level heterogeneity). β_2 represents the between estimator, which is not identical to estimators of a RE-model, but corresponds to a regression with group averages (see Figure 2). For the dependent variable, no transformation is necessary because the transformed variables address the corresponding variance structure in the dependent variable (Wang and Maxwell 2015). Genuinely time constant influences like the variable z_j can be included in the model. β_3 represents the between estimator of such a time-constant characteristic, e.g. gender. Hybrid-models can be used to model panel data with 3 or more observational periods. For data structures with only 2 waves, RE- or FE-models can be estimated.

If the within and between estimators of a given variable in the model differ, how can this be interpreted? From a statistical point of view, it seems likely that correlated unobserved heterogeneity causes the deviation of the between estimator. On the other hand, between effects have sometimes been interpreted as long-term, temporally constant relationships and within effects as short-term, time-varying relationships.[27] Strictly speaking, this argument would only be correct if we could be sure that the between effect is not biased by unobserved heterogeneity. This can only be assumed with certainty if the between-group variance is absorbed, which

[26] This corresponds to centering at the group mean. In addition, other variants of the disaggregation of the predictor variables have been suggested, especially when predictor variables are trending (see Curran and Bauer 2011 and Wang and Maxwell 2015).

[27] For example, Schmidt-Catran and Spies (2016) investigate whether immigration in Germany has an impact on support for welfare state measures. The authors disaggregate the variable immigration into a within and a between component. The results of the hybrid model show that only the within estimator has a significant (negative) impact on welfare state support, which is interpreted as evidence for the relevance of temporal changes (compared to the long-term overall level).

brings us back to the within or FE estimator. It therefore makes sense to interpret the between and within effects of a particular time-varying variable not as different but as an uncontrolled *versus* controlled effect and to interpret primarily the within effect.

Overall, it should be noted that hybrid-models represent a very flexible model class, which allows estimating the influence of group characteristics. Random Slopes and cross-level interactions can also be included in the model. Depending on the question, it should be noted, however, that between effects do not represent so-called context effects (Snijders and Bosker 2012, p. 56ff.). Context effects are effects of group characteristics under the control of relevant individual relationships and compositional differences. For example, the context effect of an aggregated variable (i.e., mean values per group) is the difference between the between and within effect—that is, the group differences that go beyond individual relationships. As between- and within-components are estimated separately in the hybrid-model, if you are interested in the context effect it can be calculate afterwards. Alternatively, uncentered intra-group variables can be included which leads the group variable to be estimated as context effect (see also Enders and Tolfighi 2007).

5. Hints for practical application

5.1 Mutual causality and "parallel trends"

All model specifications presented in this article assume that the predictor variable of the dependent variable is causally preceding, which can be derived from the exogeneity assumption. For the presence of causality, in addition to correlation, there must be an exclusion of alternative explanations (via control of a confounding variable) and changes in predictor variable must (temporally) precede the outcome (Bollen 1989). Causal ordering is quite obvious for some variable relationships (for example, political confidence does not affect a person's age), while for others it is less clear (for example, whether political trust influences social trust). Since FE-models with panel data model temporal deviations from the group average, an estimation of causal relationships was sometimes assumed here. However, in the FE-model with panel data, causality is not modeled, but is assumed *a priori* (Morgan and Winship 2007, Vaisey and Miles 2017). If there are mutual causal influences between the predictor and the outcome, the exogeneity assumption is violated and biased coefficients are the result. Here, the use of lagged dependent variables can be considered, which effectively controls for the proportion of endogeneity. However, these models are similarly to OLS- and RE-models prone to a violation of the exogenous assumption due to correlated unobserved heterogeneity. One solution is the combination of lagged dependent variables and fixed effects in structural equation models (CROSS-REFERENCE BERNING), in particular so-called *cross-lagged* panel models with fixed effects (Allison 2017, Hamaker et al., 2015).

Another assumption that plays a role in the identification of causal and unbiased effects is that of "parallel trends". For illustration, we examine the effect of individual unemployment on attitudes to the welfare state. For this purpose, a panel individual data set with five survey waves is assumed, in which some of the respondents report the event unemployment during the survey period (a re-entry into professional life is neglected here). We estimate a FE-model and discard possible unobserved heterogeneity associated with time-constant characteristics (such as gender). The results show a positive statistically significant relationship: if people become unemployed, they advocate more generous social benefits. For this effect to be unbiased, the temporal trends of the variables "advocacy of social benefits" for those who have become unemployed may not deviate from those of those who did not become unemployed. If there were different trajectories, those who in our example became unemployed might have developed a stronger support for social benefits because of the divergent trends (and the associated unobserved traits). Thus, even in the counterfactual case of non-unemployment, they would have developed a stronger preference for social benefits and the observed differences in preferences are therefore not caused by the event of unemployment.

Heterogeneous trends can be modeled in the FE-model using individualized slope lines (Brüderl and 2015, p. 336ff.) or in the hybrid-model using an interaction between group average and time dummies (Vaisey and Miles 2017, p. 56f.).

5.2 Robust standard errors

As explained earlier, cluster-robust standard errors correct for serial correlation (as well as heteroscedasticity) and therefore are often larger than conventional standard errors. Because of its ready-to-use and general applicableness, cluster-robust standard errors are widely used. At the same time, there are a number of potential complications that should be considered. First, robust standard errors are asymptotically correct and should not be used in samples with a low number of cases. In addition, the method refers exclusively to the estimation of standard errors. Correlated unobserved heterogeneity at the group level due to omitted variables is not taken into account as it would be the case in FE-models. For this reason, King and Roberts (2015) argue that deviations from conventional and robust standard errors indicates the presence of serial correlation and / or heteroskedasticity, which should primarily give rise to a better model specification (e.g., inclusion of additional relevant predictors or interactions). Another problem is that for certain correlation structures robust standard errors can also be smaller than conventional standard errors. Angrist and Pischke (2009) therefore suggest using the highest standard errors to report conservative significance tests. In addition, it was pointed out that the application of cluster-robust standard errors only works optimally when the number of groups is high (60 and more) and the number of observations within groups is relatively similar (Cameron and Miller 2015).

In summary, cluster-robust standard errors should preferably be used as an addition to RE- and FE-models, especially for panel data. RE- and FE-models correct for serial correlation due to group effects and the robust standard errors additionally correct for serial correlation due to time-varying factors and / or heteroscedasticity.

5.3 Multiple statistical dependencies

A classic example of multiple dependencies are observations that are hierarchically grouped into multiple spatial contexts, such as students in classes, which are grouped into schools, or interviewees in neighborhoods, which are grouped into communities and regions or countries.[28] How much variance can be attributed to the grouping structures can be determined via the ICC. Such multiple dependencies can be modeled with RE- or multilevel models that contain a separate error term for each group.

In addition, spatial and temporal dependencies can exist simultaneously. Again, RE-models and a combination of RE-models and dummy variables can be used for individual groupings (see Schmidt-Catran and Fairbrother 2016). For example, Ziller and Helbling (2017) analyze how temporal changes in anti-discrimination policies and citizens' knowledge of gender equality affect political support. The data used for this are structured as follows: Respondents from repeated European comparative cross-sectional surveys are grouped into country time points, which in turn are grouped into countries. The authors now estimated a RE-model in which interviewees are grouped at country time points. In order to avoid possible bias of the regression

[28] When embedded in several non-hierarchical contexts, so-called cross-classified multi-level models can be considered (Snijders and Bosker 2012, pp. 205).

coefficients due to unobserved heterogeneity between countries, dummy variables for countries are included (corresponds to the FE approach).

For panel data, in addition to controlling group-related heterogeneity using FE models, other forms of time-related or spatial serial correlation may be relevant for which can be corrected with cluster-robust or Driscoll-Kraay standard errors.

5.4 Number of Groups in RE-models

If there is a substantive interest in modeling group-level variables, RE-models can be estimated. This raises the question of whether the number of groups plays a role in the reliability of the estimation results. Literature on multilevel analysis has suggested that estimates of group-level variables produce reliable results only in models with 30 or more groups (for example, Bryan and Jenkins 2016, Stegmueller 2013). In models with a low number of groups (<20) and with more complex designs (e.g., multiple context effects or interactions), the coefficients are biased, in particular the standard errors. Since an increase in the number of groups is often not feasible (for example, analysis of countries without feasible data extension), some authors have proposed the use of Bayesian estimation methods (rather than maximum likelihood) (Stegmueller 2013).

In fact, with increasing model complexity, the number of degrees of freedom decreases and the risk of outliers[29] increases. Nonetheless, Elff et al. (2016) show that the coefficient estimators for the influence of group variables from the maximum likelihood (ML) method are generally unbiased, even for a small number of groups (e.g. 15). However, in a small number of groups, the *restricted maximum likelihood* method (REML) should be used, which estimates variance parameters by taking into account the degrees of freedom consumed by the coefficient estimation of group-level variables. The estimation of the standard errors and confidence intervals is therefore more reliable than in the classical ML procedure.[30] In addition, the calculation of p-values should be based on a t-distribution with appropriate adaptation of the degrees of freedom present in the model. This means in practice that with a small number of groups (e.g. 15), the standard errors of the group level coefficients must be substantially lower

[29] Outliers are influential cases which significantly influence coefficient estimators and whose exclusion would alter the interpretation of the results.

[30] One reason for the frequent use of ML is that model comparisons with deviance tests such as the *likelihood ratio* test must be based on ML estimates.

to match the same p-values (and corresponding statistical inference) as those from models with a higher number of groups (e.g. 40).

As a recommendation, it can be summarized that RE-models with a small number of groups should be estimated using REML methods. In addition, p-values of such context effects with small numbers of groups (<20) should no longer be taken directly from the data analysis program, but derived manually with the t-distribution corresponding to the degrees of freedom present in the model. The number of degrees of freedom refers to the group level and can be approximated by the formula $n - l - 1$ (number of groups - number of coefficients at the group level - 1). In addition, greater attention should be paid to possible influential cases by graphically presenting results, performing residual diagnostics, and / or calculating explicit outlier tests (Snijders and Bosker 2012, p. 161ff.).

6. Conclusion

In empirical-comparative political science, the analysis of grouped or clustered data structures is widespread. RE-, FE- and hybrid-models allow the modeling of data structures with statistically dependent observations. The present article focused on two consequences of statistical dependence. On the one hand, serial correlation can lead to biased standard errors and invalid test statistics. Controlling relevant variables and using RE- and FE-models can help here. On the other hand, statistical dependence leads to group-related differences, which may violate the exogeneity assumption and leads to biased coefficients. While RE-models model unobserved group level heterogeneity with an additional error term and can estimate the influence of group characteristics, they continue to assume that there is no correlation between group-related error term and the model predictor variables. FE-models do not make this assumption because they control all group-related heterogeneity by design.

Finally, the most important steps toward a suitable model specification are the following: In practice, one should first obtain a detailed overview of the data structure. Are the data spatially and / or temporally grouped? What are the relevant groups (e.g. countries, regions, cities, or individuals in panel data)? Is there a multiple grouping?

An ICC should then be calculated to estimate the level of statistical dependence or serial correlation in the dependent variable. Even with small proportions of group-related differences of total variance (e.g. 2 percent), the use of RE- or-FE models is recommended. A Hausman test reveals whether or not RE- and FE-estimation differs. If the test is not significant, a RE-model can be estimated in terms of higher efficiency. If it is significant, an FE-model should be estimated. When using FE-models, there should be an awareness of whether the relevant variables have sufficient variance within groups and whether the causal ordering of the variables is plausible. For panel and TSCS data, serial correlation tests should be done based on time-varying factors and, if necessary, robust standard errors should also be used. Moreover, in models with time-varying data, dummy variables for time points should be included to control for unobserved time related factors common to all observations (e.g., exogenous shocks or trends).

If the central research interest concerns the influence of group characteristics, such as in the context of multi-level analysis (CROSS REFERENCE Pötschke), a RE-model or hybrid-model can be estimated. In order to avoid possible bias due to unobserved heterogeneity, relevant control variables (also at the group level) should be included. The aim here is to reduce as much group-related variance as possible. Moreover, researchers should be aware of the consequences of a small number of groups, existing degrees of freedom, possible collinearities between predictor variables, and influential cases at the group level. In addition, random slopes and cross-level interactions can improve the model fit to the data and test theoretically relevant hypotheses.

7. Bibliography

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Los Angeles u.a.: SAGE publications.

Allison, Paul D., Richard Williams, und Enrique Moral-Benito. 2017. Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. *Socius* 3: 1–17.

Andreß, Hans-Jürgen, Katrin Golsch, und Alexander W. Schmidt. 2013. *Applied panel data analysis for economic and social surveys*. Heidelberg, u.a.: Springer Science & Business Media.

Angrist, Joshua D. und Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Beck, Nathaniel und Jonathan N. Katz. 1995. What to do (and not to do) with Time-Series Cross-Section Data. *American Political Science Review* 89 (3): 634–647.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Brüderl, Joself und Volker Ludwig. 2015. Fixed-Effects Panel Regression. S. 327-357 in: H. Best und C. Wolf (Hrsg.) *The SAGE handbook of regression analysis and causal inference*. Los Angeles, u.a.: Sage reference.

Bryan, Mark L. und Stephen P. Jenkins. 2015. Multilevel modelling of country effects: a cautionary tale. *European Sociological Review* 32 (1): 3–22.

Cameron, A. Colin und Douglas L. Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50 (2): 317–372.

Clark, Tom S. und Drew A. Linzer. 2015. Should I use fixed or random effects? *Political Science Research and Methods* 3 (2): 399–408.

Curran, Patrick J. und Daniel J. Bauer. 2011. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology* 62: 583–619.

Drukker, David M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3 (2): 168-177.

Elff, Martin, Jand P. Heisig, Merlin Schaeffer, und Susumu Shikano. 2016. No Need to Turn Bayesian in Multilevel Analysis with Few Clusters: How Frequentist Methods Provide Unbiased Estimates and Accurate Inference. *Working Paper*. Retrieved from osf.io/preprints/socarxiv/z65s4 (08.01.2018).

Enders, Craig K. und Davood Tofighi. 2007. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods* 12 (2): 121.

Gelman, Andrew. 2005. Analysis of Variance—Why It is More Important than Ever. *The Annals of Statistics* 33(1):1–53.

Greene, William H. 2003. *Econometric Analysis*. 5. Auflage. New Jersey: Prentice Hall.

Hakhverdian, Armen, und Quinton Mayne. 2012. Institutional trust, education, and corruption: A micro-macro interactive approach." *The Journal of Politics* 74.3 (2012): 739-750.

Hamaker, Ellen L., Rebecca M. Kuiper, und Raoul PPP Grasman. 2015. A critique of the cross-lagged panel model. *Psychological Methods* 20 (1): 102.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society:* 1251–1271.

King, Gary, und Margaret E. Roberts. 2015. How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis* 23 (2): 159–179.

Lewis-Beck, Michael S. und Mary Stegmaier. 2000. Economic determinants of electoral outcomes. *Annual Review of Political Science* 3 (1): 183–219.

Morgan, Stephen L., und Christopher Winship. 2007. *Counterfactuals and causal analysis: Methods and principles for social research*. Cambridge, UK: Cambridge: Harvard University Press.

Moulton, Brent R. 1986. Random group effects and the precision of regression estimates. *Journal of Econometrics* 32 (3): 385-397.

Schmidt-Catran, Alexander W., und Dennis C. Spies. 2016. Immigration and welfare support in Germany. *American Sociological Review* 81 (2): 242–261.

Schmidt-Catran, Alexander W., und Malcolm Fairbrother. 2015. The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review* 32.1: 23-38.

Snijders, Tom A. B., und Roel J. Bosker. 2012. *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. 2. Auflage. Los Angeles, u.a.: Sage.

Stegmueller, Daniel. 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57 (3): 748–761.

Tilley, James, Anja Neundorf, und Sara Hobolt. 2017. When the pound in people's pocket matters: how changes to personal financial circumstances affect party choice. *Journal of Politics*.

Vaisey, Stephen und Andrew Miles. 2017. What you can—and can't—do with three-wave panel data. *Sociological Methods & Research* 46 (1): 44–67.

Van der Brug, Wouter, Cees Van der Eijk, and Mark Franklin. 2007. *The economy and the vote: Economic conditions and elections in fifteen countries*. Cambridge, MA.: Cambridge University Press.

Verbeek, Marno. 2004. *A Guide to Modern Econometrics.* Chichester: John Wiley & Sons.

Wang, Lijuan Peggy und Scott E. Maxwell. 2015. On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods* 20 (1): 63.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A modern approach*. Boston: Cengage Learning.

Ziller, Conrad, und Thomas Schübel. 2015. "The Pure People" versus "the Corrupt Elite"? Political Corruption, Political Trust and the Success of Radical Right Parties in Europe. *Journal of Elections, Public Opinion and Parties* 25 (3): 368–386.

Ziller, Conrad, und Marc Helbling 2017. Antidiscrimination Laws, Policy Knowledge, and Political Support. *British Journal of Political Science*. First View doi:10.1017/S0007123417000163. S. 1-18.