Marco Giesselmann
Katrin Golsch · Henning Lohmann
Alexander Schmidt-Catran *Hrsg.*

Lebensbedingungen in Deutschland in der Längsschnittperspektive

Springer VS

Lebensbedingungen in Deutschland in der Längsschnittperspektive

Marco Giesselmann
Katrin Golsch · Henning Lohmann
Alexander Schmidt-Catran
(Hrsg.)

Lebensbedingungen in Deutschland in der Längsschnittperspektive

Springer VS

Herausgeber
Marco Giesselmann
Berlin, Deutschland

Henning Lohmann
Hamburg, Deutschland

Katrin Golsch
Osnabrück, Deutschland

Alexander Schmidt-Catran
Köln, Deutschland

Festschrift für Hans-Jürgen Andreß

ISBN 978-3-658-19205-1 ISBN 978-3-658-19206-8 (eBook)
https://doi.org/10.1007/978-3-658-19206-8

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer VS

Lektorat: Katrin Emmerich

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist Teil von Springer Nature
Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Inhalt

D Methoden der längsschnittlichen Forschung

Fixed Effects Regression and Effect Heterogeneity

An Illustration Using a Causal Inference Perspective[1]

Luis Maldonado und Pablo Geraldo

Abstract

Fixed effects regressions are commonly used by social scientists to identify causality. However, several criticisms against the fixed effects estimator emerged in recent years. In addition to confounding factors that are associated with time variant covariates, fixed effects can lead to an improper aggregation of heterogeneous effects. In the present chapter, we discuss the problem that pertains to the fixed effect estimator and show techniques that do not suffer from this source of bias. We also illustrate the problem with empirical analysis of Chilean students for the period time from 2007 to 2013. On the basis of the theoretical framework developed in the chapter and empirical findings, we suggest some implications for research in social sciences.

1 Introduction

In the social sciences, fixed effects regression is a standard method to identify causality by using data with units nested within some group or strata, such as panel data, multilevel research, and experimental block designs. One of the main advantages of this type of estimator is that it accounts for unobserved confounders that do not vary between the groups. An illustration of this advantage is panel data. As Andreß et al. pointed out (2013), analysis of this kind of information with fixed

1 Luis Maldonado acknowledges support provided by FONDECYT REGULAR 1160921 and Centro Nacional de Investigación para la Gestion Integrada de Desastres Naturales (CIGIDEN). We thank Marco Giesselmann for helpful comments.

M. Giesselmann et al. (Hrsg.), *Lebensbedingungen in Deutschland in der Längsschnittperspektive*, https://doi.org/10.1007/978-3-658-19206-8_17

effects regressions makes it possible to control for unmeasured determinants of outcomes that are constant over time. Furthermore, we can assess whether changes of a treatment precede changes of an outcome by examining panel information with fixed effects models.

In spite of the advantages, several criticisms of fixed effects models have emerged in other settings during the last years, especially in the causal inference literature. By focusing on the conditions that are sufficient for drawing causality, recent studies show that this method can generate some source of bias that are not related to either omitted variables or a temporal sequence of treatment and outcome (Middleton et al. 2016). In the field of causal inference, researchers have developed novel methods to break down some of these problems that pertain to a fixed effects model; unfortunately, however, these techniques are not frequently used or are unknown to social scientists.

In such a context, the goals of this chapter are to examine how fixed effects regressions identify causal effects and to discuss some methodological alternatives to the method. We develop our exposition by focusing on blocks designs and linear regression (ordinary least square, or OLS). This kind of data contains units nested within strata that control for several relevant confounders and are frequently used in multilevel research with individuals nested within schools or countries. Importantly, strata can be used as fixed effects in standard linear regressions.

Following the contemporary discussion about fixed effects in causal inference literature (Angrist and Pischke 2008), we conceptualize the fixed effects regression model as a way to average effect heterogeneity.[2] On the basis of this framework, we argue that the averages that result from fixed effects are based on an aggregation procedure of heterogeneous effects that, under certain conditions—for example, data with strongly unbalanced strata—may generate biased results (Aronow and Samii 2013, 2016). We develop this argument by using formal exposition, and, guided by our formalization, we also highlight theoretical conditions under which fixed effects models may be harmful.

In the causal inference setting, most existing research about fixed effects focuses on experimental designs (Middleton et al. 2016), but their extensions to observational designs are lacking. In addition to formal exposition, we consequently evaluate fixed effects OLS regressions with a large N observational study of Chilean students for the period from 2007 to 2013. We processed these data with matching techniques

2 Every regression method can be conceptualized as way to average effect heterogeneity. However, as is said in the following sections, a causal inference framework accounts for the assumptions of the aggregation procedure. We do not think that the standard econometric expositions of regression models account for these assumptions.

to generate blocks that are used to evaluate the performance of fixed effects OLS regression under the conditions highlighted by our theoretical discussion. To qualify the estimates of the fixed effects models, we also present findings based on techniques that do not suffer from the same problems found with this estimator, such as inverse probability weighted regressions. All in all, our formal exposition and empirical findings suggest that scholars should refrain from making an overly optimistic reference to fixed effects estimates as causal effects, insofar as the method generates some additional sources of bias that are not related with the exogeneity of treatments.

The remainder of the chapter is organized as follows: The first section provides a standard presentation of fixed effects OLS models. The second section translates this standard presentation to a counterfactual formulation and highlights the aggregation of the heterogeneous effect that a fixed effects regression uses. The third section offers an empirical illustration, and the final section features conclusions and implications for empirical research.

2 Standard modeling approach of fixed effects regression

Following the work of Andreß et al. (2013), we present the standard modeling approach for fixed effects by assuming that individuals are nested in strata (e.g., people within countries or students within schools). We assume a set of units indexed by i that belongs to stratum j with i=1…N and j=1…J. We are interested in the effect of a treatment D_{ij} on an outcome variable Y_{ij}. Notice that both variables are observed for each unit i in stratum j. The treatment is binary, with 1 indicating the fact that unit i is exposed to the treatment, and 0 denoting units that are exposed to the control condition. The outcome Y is a continuous variable. Formally, the fixed effects model may be written as

$$y_{i,t} = \beta_1 D_{1j} + \mu_j + \varepsilon_{i,j} \tag{1}$$

where μ_j denotes dummy variables for the strata that capture the observed and unobserved characteristics of these blocks. The variable ε_{ij} is the error term that comprises the unobserved characteristics of unit i that belongs to stratum j. A common estimation approach of β_1 is an OLS regression of the outcome Y on the treatment D and dummy variables for the strata. This model estimates the effect of this treatment, controlling for the observed and unobserved characteristics of the blocks. In other words, we estimate the effect of treatment D while accounting for *fixed effects* for the strata and, consequently, (1) is known as the fixed effects model (FE).[3]

Following the econometric literature (Andreß et al. 2013; Wooldridge 2010), the critical assumption of FE to identify the effect of D (β_1) is named the exogeneity of variable ε_{ij} and is expressed in terms of mean independence $E(\varepsilon_{ij}|D_{ij},\mu_j) = 0$. This assumption means that unobserved characteristics of units are not correlated with independent variables of model (1). When this assumption fails, the estimation of the effect of D is biased, insofar as the estimated regression coefficient for D does not capture the effect of D, but also the association between Y and another unobserved variable. Based on the assumption of exogeneity, in the next section we show a counterfactual formulation by which we illustrate some handicaps of the fixed effects models.

3 Fixed effects model and effect heterogeneity

This section presents the fixed effects model as a way to average effect heterogeneity. We develop this topic in two steps. First, the counterfactual formulation of the causal effect of the treatment is illustrated. For simplicity, we begin with a simple experimental design without a nested structure and then we extend this design to block data. On the basis of the counterfactual formulation, we also present alternative procedures with which to analyze a block design. Second, the special type of aggregation of heterogeneous effects that is used by fixed effects is developed. The discussion is focused on OLS linear regression.

3.1 Counterfactual formulation of causal effects

For most experimentalists, causation denotes "something that makes a difference, and the difference it makes must be a difference from what would have happened without it." (Lewis 1974). This idea is the core of the counterfactual approach. In this framework (Imbens and Rubin 2015), each individual has two potential outcomes,

3 For the topic of our chapter, the inclusion of additional independent variables in the models is not critical, and, thus, we prefer to exclude these kinds of variables and to make our exposition as simple as possible. Furthermore, we also do not discuss topics of inference with FE regression, for example cluster standard errors.

but only one of them is realized. Potential outcomes represent individual behavior in the presence and absence of a treatment, and the observed outcome depends on the realized treatment status. As stated above, let D be the treatment binary variable, where 1 indicates the treatment group, 0 indicates the control group, and Y is the outcome variable. Let $Y_i(1)$ be the outcome if unit i is exposed to treatment D and $Y_i(0)$ represents the outcome's value when unit i is exposed to the control condition. For each unit i, the causal effect of treatment is τ_i, and it corresponds to the difference of two outcomes:

$$\tau_i = Y_i(1) - Y(0) \tag{2}$$

This expression is also called the *unit effect*, and it compares the outcome's values in the treatment and control conditions for the same unit i. However, real data only have information for one of these conditions. Put otherwise, a unit i cannot be treated and control at the same time. In the causal inference literature, this characteristic of (2) is denoted as the fundamental problem of causal inference (Holland 1986). One way to solve this problem is to use some kind of procedure that aggregates the unit effects or *effect heterogeneity*. In other words, we cannot estimate the unit effects, but it may be possible to estimate an aggregation of them. Following the work of Imbens and Rubin (2015), an unbiased aggregation's procedure is to use the average information of the unit and, consequently, to define the average treatment effect or ATE:[4]

$$ATE = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} \tau_i \tag{3}$$

An equivalent way to express the average treatment effect is $E[Y_i(1)] - E[Y_i(0)]$. By assuming the exogeneity or random assignment of the treatment[5], an estimator of the ATE is the simple difference between the outcome's average of the treatment and the control groups $ATE = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$.[6] Consequently, this procedure provides estimations of effects as averages and not per unit. Put oth-

4 Standard regressions such as OLS also use this kind of aggregation.

5 The random assignment of units into the control and treatment groups generates the exogeneity of treatment, insofar as the physic mechanism of randomness balances controls and treated groups. Balance means that both groups are equivalent in terms of their observed and unobserved characteristics.

6 Our discussion puts the focus on ATE and not on the average treatment effect for treated (ATT) or for control (ATC).

erwise, while we can estimate average effects, the ATE does not capture how the effect varies across units.[7]

Following the experimental design literature, the extension of the ATE for data with individuals nested within strata is straightforward. This kind of data is classified as being configured in a block design, and the main characteristic of it is that units are partitioned into blocks. As Gerber and Green (2012) pointed out, the block design has three main characteristics. First, blocking ensures the balance of all observed variables that are used to generate blocks. It means that blocking controls for all these variables. Second, blocking improves the precision of ATEs by reducing the standard errors. Third, random assignment occurs within each block. Because of randomization, we can assume the exogeneity of the treatment within each block, but the probabilities of treatment assignment can differ among strata. For example, we might have 50 % of controls and treated elements in block A, but block B could contain 30 % of the treated and 70 % of controls.

To estimate the unbiased causal effect, the main concern of block design is how an estimator can account for the different proportion of treated and control elements among the blocks. A procedure that identifies the ATE and accounts for heterogeneous assignment probabilities among strata is a weighted average of the average treatment effects for each block, and this can be expressed in the following way:

$$ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j \tag{4}$$

where the weight N_j/N refers to the share of all units that belong to block j, the blocks are indexed by j, and J is the number of blocks. Consequently, the weights imply that the larger the size of the block j, the higher the impact of the ATE of this block on the overall ATE. As we will explain in the section that follows, fixed effects regressions can also estimate (4), but only specific conditions. We can estimate (4) by using the following procedure: 1) for each strata, do a stratum-specific simple regression of Y on D and get the stratum-specific coefficients that represent the ATEs; 2) compute the weighted stratum-specific coefficients by using the weight N_j/N; and 3) estimate the sum of the weighted ATEs. In the literature, this method is called *standardization* (Hernan & Robins, 2017). An alternative approach that also accounts for the unequal probabilities of treatment assignment is the *inverse probability weighted (IPW) regression*. The weights are the inverse of the proportion

7 In the setting of causal inference, the exogeneity of treatment is called the independence assumption. On the basis of this assumption, ATE provides a solution to the fundamental problem by imputing the values of the observed conditions to the counterfactual conditions. For details of the imputation procedure, see Imbens and Rubin (2015).

of subjects in a particular block who are assigned to treatment conditions. More specifically, the weights are:

$$g_{ij} = \frac{D_i}{p_{ij}} + \frac{1 - D_i}{1 - p_{ij}} \tag{5}$$

D_i is the binary variable that denotes whether the unit is in the treated or control group, and p_{ij} is the probability that subject i in block j is assigned to treatment—in other words, p_{ij} is the proportion of treated and control elements within each block. To identify the causal effect by using weights (5), we estimate a simple OLS regression of the outcome on the treatment, but we weigh the regression by using inverse probability weights g_{ij}. The results of inverse probability weighting are equivalent to the standardization procedure, but it could be more efficient—it means lower standard errors—insofar as we can include control variables in regression models (Hernan and Robins 2017). In the following section, we relate these unbiased estimators of the ATE to a block design with FE.

3.2 Fixed effect model as a weighting procedure

In the case of block designs, a common estimation approach of ATE is an OLS regression with fixed effects for blocks such as equation (1): The outcome is regressed on the treatment and dummy variables for the blocks. As will become evident, this regression is a type of FE model. Recent studies suggest that this fixed effects model is algebraically equivalent to a weighting procedure that produces an unbiased estimation of the causal effect of the treatment only under certain conditions (Angrist and Pischke 2008; Gerber and Green 2012).

To identify these conditions, we have to relate FE regressions with equation (4). As Humphreys (2009) points out, an OLS regression of outcomes on the treatment and dummies for blocks is similar to a procedure that is analogous to the *standardization method*, but the weights do not only consider the share of all units that belong to a block; they also include the variance of treatment for each block. More specifically, following the notation of Humphreys and assuming random assignment of treatment within blocks, FE regression estimates an ATE that is equivalent to

$$ATE = \sum_{j=1}^{J} \frac{w_j}{\sum_{j=1}^{J} w_j} ATE_j \tag{6}$$

where $w_j = (N_j/N)P_j(1 - P_j)$ and P_j is the proportion of observations in block j that are assigned to the treatment condition—or the probability of being treated.

The core characteristic of (6)—and the central point of our argument—is that P_j $(1 - P_j)$ is the variance of a dichotomous variable; for example, the treatment D. In comparison with the estimator of the ATE for block design expressed in equation (4), FE as expressed in (6) consequently generates a causal effect that is reweighted by the conditional variance of the treatment. Therefore, notice that the condition of LM-FE to produce an unbiased estimation of the causal effect of the treatment: *P_j is constant across blocks*. It is only under this condition that FE and the standardization method produce identical findings. For example, a design that complies with this condition is a pair block design in which each block has one treated and one control element. When this condition does not hold, the FE is not free of bias and the reason is not the presence of omitted variable bias—recall that we assume random assignment of treatment—but the variance of treatment for each block that is considered by the weighting procedure of FE. Put otherwise, the reason is the improper aggregation of heterogeneous effects that is used by fixed effects regressions. In the causal inference literature, this problem of FE is conceptualized as *aggregation bias* (Aronow and Samii 2013).

In the setting of observational studies, aggregation bias can be a problem. Even though the exogeneity of the treatment is credible, the estimates of FE are still biased if what we really want is a proper estimator of the ATE that accounts for unequal probabilities of treatment assignment (i.e., standardization or IPW regression). When the probabilities of treatment assignment are very unequal—for example, as is common in unbalanced panel data—FE may be harmful. However, it is not clear in practice under which conditions this kind of bias is a major concern. According to our own knowledge, most existing research on aggregation bias focuses on experimental designs such as, for example, the work on Middleton et al. (2016), but explorations of this problem in the setting of observational studies are lacking. In the following section, we attempt to fill in this gap by examining aggregation bias associated with FE by using real data.

4 An Empirical Illustration

As previously explained, using fixed effects models when analyzing grouped observations (v.gr. experimental block designs, panel data, or matching designs) could produce the so-called aggregation bias when the probabilities of receiving the treatment of interest differ across groups. For example, this is the case when we have experimental block designs with different shares of treatment and control units across strata, or unbalanced panel data. In this section, we illustrate the importance of taking into account aggregation bias in FE models by using observational data coming from Chilean educational system.

We develop our example by comparing the performance on a mathematics test of students attending vocational and general schools.[8] In Chile, due to a reform that took place in 1998, the official duration of the vocational track was shortened from 4 to 2 years, currently covering grades 11 and 12. The purpose of the reform was to provide the same academic opportunities to students attending general and vocational schools for a longer period. Despite this, most secondary schools exclusively offer one educational track, being specialized institutions irrespective of the grade. In consequence, although previous research on vocational education has focused on the official two-year duration of vocational training, there is a broad consensus that, due to structural differences with comprehensive schools, vocational schools expose their students to more limited learning opportunities from the very beginning of their secondary education, in grades 9 and 10 (Larrañaga et al. 2013).

4.1 Data and Methods

The empirical analysis of this paper is based on a subset of a Chilean census of students named "Measurement System of Educational Quality" (SIMCE, for the Spanish acronym). SIMCE is applied in variable years to students in grades 4, 6, 8, and 10. Our data set included IRT scores for mathematics, Spanish, and the social and natural sciences. Additionally, during SIMCE's application, students, parents, and teachers were surveyed to collect information on factors associated with the academic performance of the students. These surveys provide background information that accounts for an extensive set of confounders.[9] We use data from the application in grades 4 and 8 (years 2007 and 2011) for the baseline students' characteristics and from the application in grade 10 (year 2013) for the outcome variable. The analytic sample is comprised of 7,525 students, which only includes urban youth from the Metropolitan Area in Santiago, Chile.

On the basis of this data, our empirical illustration looks to identify the causal effect on students' academic performance of attending a secondary school that

8 The data and empirical analysis used in this chapter are based on the M.A. thesis of Pablo Geraldo (Geraldo, 2015).

9 In our analysis, we consider confounders that include gender, preschool attendance, family income, and previous academic performance in mathematics and Spanish tests. To avoid some post-treatment bias, the covariates are measured in 2007 and 2011, the treatment corresponds to the school track attended through years 2012 and 2013, and the outcome is measured in 2013.

specializes in vocational education during grades 9 and 10. Due to the non-random assignment of students in the educational track, we use a *propensity score matching* method to resemble a block experimental design (Gerber and Green 2012; Rosenbaum 2010). In other words, we grouped students in many strata, based on their similarity in observable characteristics, constructing an as-if-random scenario: Within each stratum, treated (vocational) and control (general) students have the same observable characteristics, so comparing them represents an unbiased approximation to the causal effect of interest under the assumption of the exogeneity of the treatment that is conditional on observable characteristics. Put otherwise, matching techniques transform the data into a block design accounting only for observable variables, but not for unobservable features. Therefore, matching is not insensitive to omitted variable bias.

As previously indicated, one could aggregate such strata-specific effects by taking the weighted average of them in which the weights correspond to the size of each strata over the sample size. Equivalently, we can use an IPW regression that yields the same results. On the other hand, when the number of strata grows, it is a common practice to use FE regression to aggregate the strata-specific effects: one may run a regression on the pooled data and include a dummy variable for each stratum. However, as we will show, that way of aggregating the overall effect does not always correspond to the weighting process we seek. In fact, when the share of treatment and control units varies across strata, FE regression produces results that depart from the straightforward aggregation of strata-specific effects because of the (improper) variance-weighting aggregation described above (equation (6)).

In our matching example, one would observe the FE and IPW regression yielding the same result when we have observations stratified in groups of the same size with the same share of treatment and control units within: the simpler case corresponds to a *pair matching* design, i.e. comparing one vocational student with the most similar general student, and then taking the average across all the pairwise comparisons (Rosenbaum 2010).

On the contrary, one would observe the FE and IPW regression yielding different results when we have a more complex matching schema with observations stratified in different-sized groups with varying shares of treatment and control units within—that case corresponds to a *full matching* design, i.e. comparing many-to-many vocational and general students). In that scenario, the probability of receiving the treatment varies across strata, and the aggregation of strata-specific effects through FE regression (i.e., including a dummy for each strata) is biased, whereas the IPW regression offers unbiased results if we assume exogeneity of treatment when observed covariates are included in the model (Gerber and Green 2012).

5 Results

In what follows, we will compare the results of FE regression and IPW regression in the two scenarios described above: one based on pair matching and one based on full matching. Based on those matched samples, Tab. 1 presents the estimations of the effect of attending vocational schools in comparison to attending general schools on mathematics test performance, obtained through FE and IPW regressions.

Tab. 1 Effect of Vocational Education estimated through FE and IPW regressions

	FE	IPW
OLS Regression	-0.23	
Pair Matching	-0.60	-0.60
Full Matching	-0.26	-0.29

Source: Author's own calculations based on a subsample of Chilean System of Measurement of Quality in Education (SIMCE), Agency for Quality in Education.

Notes: The treatment variable is vocational vs. general school, and the outcome variable is mathematics test performance in 10th grade. Linear models were estimated in three different samples: 1) a multivariate OLS regression, 2) bivariate OLS regressions in data pre-processed with pair matching, with FE identifying pairs, and then IPW by pairs, and 3) bivariate regression in data pre-processed with full matching, with FE identifying matching blocks, and then IPW by blocks.

First, we estimate a multivariate OLS regression as a baseline, which indicates a detrimental impact of attending vocational schools of 23 % of a standard deviation on the mathematics test performance, which is a little more than the gap observed in the Chilean educational system between for-profit and non-profit high schools.

Then, the second row presents the results of estimation by using FE and IPW regressions with data pre-processed through pair matching. In both cases, the estimates correspond to bivariate regressions. The FE accounts for the block design by including a fixed effect that indicates the pair to which each observation belongs, while the IPW accounts for the block design by weighting each case by the inverse of the probability of being treated within its block. As expected, the results obtained through both methods are exactly the same. The estimates are consistent with a detrimental effect of attending a vocational school, but the magnitude of the effect is much bigger: 60 % of a standard deviation on a mathematics test.[10] As explained above, in aggregating the treatment effect across blocks, the FE and IPW deliver identical results because the size of the blocks and the share of treatment subjects within each block are the same. In other words, this occurs because the treatment probability is homogeneous across strata.

However, if we look at the estimates obtained in the data pre-processed with full matching, the scenario is different. Again, the results correspond to bivariate regressions, but the data was pre-processed by generating blocks of variable sizes and variable shares of treatment units within them. Now, accounting for the blocking structure of the data through the inclusion of fixed effects or weighting the observations by the inverse of the treatment probability within each block leads to different results. The FE model estimated the effect of attending to vocational education as a detrimental effect of 26 % of a standard deviation in the mathematics score, while the IPW estimated such effect as a decrease of 29 % of a standard deviation on the same test.

The difference between both estimates leads us to an improper aggregation of heterogeneous effects that characterizes FE regressions. The difference of approximately 3 % denotes the presence of aggregation bias in the FE model. However, this finding suggests that the aggregation bias is weak in our data. The reason seems to depend on the fact that the contribution of blocks with extremely unequal probabilities of treatment assignment to the estimation of the ATE is compensated by blocks with similar numbers of treated and control elements. We think that this pattern is typical of studies with a large N sample and a large number of blocks. Furthermore, these estimates could be biased for other reasons; for example, as a result of the imbalance in the distribution of students' or schools' covariates within each block. In the following section, we discuss implications of the findings of our empirical illustration.

10 Although this is not the main focus of our analysis, it is worth explaining the size of the effect estimated. Pair matching does not minimize the overall distance between observations. The practical result of this is that such matching is not very effective in controlling for the confounder variables, which is expressed through this likely biased effect estimate.

6 Discussion

In this chapter, we show that fixed effects regression can lead to an improper aggregation of heterogeneous effects. We frame the presentation of the problem by assuming individuals nested within strata. On the basis of this kind of research design, we offer a formal discussion of the problem by focusing on a specific nested structure: an experimental block design. Furthermore, we show an empirical illustration by discussing matching data as a block design. Our formal discussion and the findings of the empirical illustration suggest important implications.

The formal exposition and findings indicate that omitted variable is not the unique source of bias that can affect the estimates of FE regressions. Even though the exogeneity of the treatment is credible and simultaneous causality is ruled out, our chapter shows that an additional source of bias can persist in estimates of FE models. It implies some ideas as a guide to interpret FE regressions. First, we recommend considering the proportions of treated and control elements nested within the group of interest (e.g., schools or countries). When the proportion is unbalanced, one must be aware that some aggregation bias may emerge. Second, blocks do not have the same contribution to the average causal effect. Some stratum may have higher weights and, thus, some external validity problems can emerge (Aronow and Samii 2016). We consequently recommend examining the relative size of the block with respect to the population of interest and to recall that FE is not impartial. We think that this recommendation is particularly important for a multilevel design that examines individuals nested within countries, insofar as the number of nations is small and, consequently, the estimates can be affected for the weights of some nations.

How important is aggregation bias in FE regressions? To clearly answer this question, we need to go beyond the empirical illustration of this chapter; we need a systematic evaluation. In order to provide this kind of test, future research should examine the role of aggregation bias under different patterns of balances of covariates by using, for example, simulations. Furthermore, more research is necessary in order to evaluate aggregation bias with observational longitudinal data. It is not clear how the weighting procedure used by FE assesses the aggregation of unit causal effects in this kind of setting. Developing formal and empirical evaluations of this problem with large N micro panel data with FE or random effects models, as in the work of Blackwell (2013), would be valuable.

References

Andreß, Hans-Jürgen, K. Golsch, and A.W. Schmidt, A. W. 2013. *Applied panel data analysis for economic and social surveys*. Springer Science & Business Media.

Angrist, Josh D., and J.-S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Aronow, P. M., and C. Samii. 2013. Estimating average causal effects under interference between units. *arXiv Preprint arXiv:1305.6156*.

Aronow, P. M., and C. Samii. 2016. Does Regression Produce Representative Estimates of Causal Effects? *American Journal of Political Science* 60(1): 250-267.

Blackwell, M. 2013. A framework for dynamic causal inference in political science. *American Journal of Political Science* 57(2): 504-520.

Geraldo, P. 2015. *El rol de la ensenanza media tecnico profesional en la reproduccion de la desigualdad educativa. Un estudio cuasi-experimental basado en el modelo de efectos primarios y secundarios del origen social.* (Tesis de Magister). Pontificia Universidad Catolica de Chile, Santiago de Chile.

Gerber, Alan S., and D.P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Hernan, Miguel A., and J.M. Robins. 2017. *Causal inference*. Chapman and Hall/CRC.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81(396): 945-960.

Humphreys, M. 2009. Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. *Manuscript, Columbia University*.

Imbens, Guido W., and D.B. Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Larrañaga, O., G. Cabezas, and F. Dussaillant. 2013. *Informe completo del Estudio de la Educacion Tecnico Profesional*. Programa de Naciones Unidas para el Desarrollo.

Lewis, D. (1974). Causation. *The Journal of Philosophy* 70(17): 556-567.

Middleton, J. A., M.A. Scott, R. Diakow, and J.L. Hill. 2016. Bias amplification and bias unmasking. *Political Analysis* 24(3): 307-323.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Verzeichnis der Autorinnen und Autoren

Anja-Kristin Abendroth ist akademische Rätin an der Fakultät für Soziologie der Universität Bielefeld. Ihre Forschungsschwerpunkte sind soziale Ungleichheiten und Interdependenzen von Arbeit und Privatem sowie vergleichende Organisations- und Wohlfahrtsstaatsforschung.

Gerrit Bauer ist akademischer Rat an der Ludwig-Maximilians-Universität (LMU) München. Seine Arbeitsschwerpunkte liegen in der Familien- und Ungleichheitsforschung.

Klaus Birkelbach ist apl. Professor für Soziologie im Institut für Soziale Arbeit und Sozialpolitik der Universität Duisburg-Essen. Seine inhaltlichen Arbeitsschwerpunkte sind Bildungsforschung, Religionssoziologie, Soziale Ungleichheit, Soziologie des Lebenslaufs.

Miriam Bröckel ist wissenschaftliche Mitarbeiterin am Institut für Sozialwissenschaften in Osnabrück. Ihre Arbeitsschwerpunkte sind soziale Ungleichheit, Partnerschaft und Karriere, soziale Unterstützung und Quantitative Methoden.

Josef Brüderl ist Professor für Soziologie an der LMU in München. Seine Arbeitsschwerpunkte sind Quantitative Methoden der Sozialforschung (insbesondere Längsschnittmethoden), Familien- und Arbeitsmarktforschung.

Anne Busch-Heizmann ist Professorin für Soziologie mit dem Schwerpunkt Soziale Ungleichheit und Genderforschung an der Universität Duisburg-Essen. Ihre Forschungsschwerpunkte umfassen Arbeitsmarkt, Berufe und Organisationen, Soziale Ungleichheit (Gender, Migrationshintergrund), Arbeitsinhalte, -bedingungen, -belastungen, Partnerschaft und Karriere und Quantitative Methoden.

M. Giesselmann et al. (Hrsg.), *Lebensbedingungen in Deutschland in der Längsschnittperspektive*, https://doi.org/10.1007/978-3-658-19206-8

Romana Careja is Associate Professor at the Political Science Department, University of Southern Denmark. Her main publications are in the area of immigration policies and attitudes towards immigrants in Western European countries.

Frederike Esche ist wissenschaftliche Mitarbeiterin am Fachbereich Sozialökonomie an der Universität Hamburg. Ihre Arbeitsschwerpunkte sind soziale Ungleichheiten, Lebenszufriedenheit und Lebenslauf.

Pablo Geraldo is instructor in the Department of Sociology at the Pontificia Universidad Catolica of Chile, and researcher in the Center of Public Policy UC, Chile.

Marco Giesselmann vertritt die Professur für empirische Sozialforschung an der TU Chemnitz und ist Mitarbeiter des DIW Berlin. Seine inhaltlichen Arbeitsschwerpunkte sind Armut, Arbeitsmarktsoziologie und Methoden der quantitativen Sozialforschung.

Achim Goerres ist Professor für Empirische Politikwissenschaft und Leiter des Profilschwerpunkts „Wandel von Gegenwartsgesellschaften" an der Universität Duisburg-Essen. Er arbeitet an der Schnittstelle zwischen politischer Verhaltens- und Wohlfahrtsstaatenforschung mit dem Fokus auf Migration, gesellschaftliche Alterung und Politik.

Katrin Golsch ist Professorin für Quantitative Methoden der empirischen Sozialforschung und Sozialstrukturanalyse am Institut für Sozialwissenschaften an der Universität Osnabrück. Ihre Arbeitsschwerpunkte sind Methoden der empirischen Sozialforschung, Familien- und Arbeitsmarktforschung.

Olaf Groh-Samberg ist Professor für Soziologie an der Universität Bremen. Er leitet dort am Forschungszentrum „Ungleichheit und Sozialpolitik" (Socium) in der Abteilung „Ungleichheitsdynamiken in Wohlfahrtsgesellschaften" die Arbeitsgruppe „Soziale, kulturelle und ökonomische Ungleichheiten" und ist Dean der „Bremen International Graduate School of Social Sciences" (BIGSSSS). Seine Arbeitsschwerpunkte gelten der Soziologie sozialer Ungleichheiten.

Jacques Hagenaars is emeritus professor of Methodology of the Social Sciences at Tilburg University, Faculty of Social and Behavioral Sciences. His main publications are in the area of categorical data analysis, longitudinal research, loglinear mondeling and latent class analysis.

Karsten Hank ist Professor für Soziologie an der Universität zu Köln, Research Fellow am DIW Berlin und Co-PI des Beziehungs- und Familienpanels pairfam. Seine Forschungsschwerpunkte liegen im Bereich der Familien- und Alternsforschung.

Moritz Heß ist Wissenschaftlicher Mitarbeiter am Institut für Gerontologie an der TU Dortmund. Seine Arbeitsschwerpunkte sind Wohlfahrtstaat und soziale Gerontologie.

Dirk Hofäcker ist Professor für Methoden der quantitativen Sozialforschung am Institut für Soziale Arbeit und Sozialpolitik der Universität Duisburg-Essen. Seine Arbeitsschwerpunkte liegen im Bereich der international vergleichenden Lebenslaufforschung sowie der Arbeitsmarktsoziologie.

Julie Korbmacher ist wissenschaftliche Mitarbeiterin am Munich Center for the Economics of Aging am Max-Planck-Institut für Sozialrecht und Sozialpolitik. Ihre Forschungsschwerpunkte sind Interviewereffekte und Datenqualität.

Fabian Kratz ist Post-Doc am Institut für Soziologie der Ludwig-Maximilians Universität München. Seine inhaltlichen Arbeitsschwerpunkte umfassen soziale Ungleichheit, Migration und Methoden der quantitativen Sozialforschung.

Steffen Kühnel ist Professor für quantitative Methoden in den Sozialwissenschaften. Seine Arbeitsschwerpunkte sind multivariate statistische Modelle sowie die politischer Einstellungs- und Verhaltensforschung.

Henning Lohmann ist Professor für Soziologie, insbesondere Methoden der empirischen Sozialforschung an der Universität Hamburg. Seine Arbeitsschwerpunkte sind Armut und Ungleichheit in wohlfahrtsstaatlich vergleichender Perspektive.

Luis Maldonado is a professor in the Department of Sociology at the Pontificia Universidad Catolica of Chile, with affiliation in the The National Research Center for Integrated Natural Disaster Management (CIGIDEN), Chile.

Anja Mays ist wissenschaftliche Mitarbeiterin am Methodenzentrum Sozialwissenschaften der Universität Göttingen und leitete das DFG-Forschungsprojekt „Politische Sozialisation im Erwachsenenalter". Ihre Arbeitsschwerpunkte sind Einstellungs- und Werteforschung, politische Psychologie, Wahl- und Partizipationsforschung.

Heiner Meulemann war bis 2013 Professor für Soziologie an der Universität zu Köln. Arbeitsgebiete: Allgemeine Soziologie, Bildungssoziologie, Sozialer Wandel, Internationaler Vergleich. 2000 – 2010 im Projektleiterteam des deutschen Teils des European Social Survey.

Timo-Kolja Pförtner ist wissenschaftlicher Mitarbeiter am Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaften an der Universität zu Köln. Seine Arbeitsschwerpunkte sind Armut, soziale Determinanten von Gesundheit und Methoden der empirischen Sozialforschung.

Andreas Pöge ist akademischer Oberrat im Arbeitsbereich 2, Methoden der empirischen Sozialforschung mit dem Schwerpunkt Quantitative Methoden an der Fakultät für Soziologie der Universität Bielefeld.

Jost Reinecke ist Professor für Quantitative Methoden der empirischen Sozialforschung an der Universität Bielefeld. Seine Arbeits- und Forschungsschwerpunkte sind Rational-Choice Theorien in den Sozialwissenschaften, die Methodologie und Anwendung von Klassifikations- und Strukturgleichungsmodellen, Verfahren zur mehrfachen Ersetzung von fehlenden Werten und die Entwicklung der Jugendkriminalität im Altersverlauf.

Alexander Schmidt-Catran ist akademischer Rat am Institut für Soziologie und Sozialpsychologie an der Universität zu Köln. Seine Arbeitsschwerpunkte sind Migration und Wohlfahrtsstaat sowie Methoden der empirischen Sozialforschung.

Jürgen Schupp ist Direktor des Sozio-oekonomischen Panels (SOEP) am Deutschen Institut für Wirtschaftsforschung und Professor für Soziologie, insb. Methoden der empirischen Sozialforschung am Institut für Soziologie der Freien Universität Berlin. Seine inhaltlichen Arbeitsschwerpunkte sind Soziale Indikatoren, Subjektives Wohlbefinden, Sozialstrukturanalyse und Methoden empirischer Sozialforschung.

Hawal Shamon ist Post-Doktorand am Institut für Energie- und Klimaforschung Systemforschung und Technologische Entwicklung (IEK-STE) am Forschungszentrum Jülich. Seine Arbeitsschwerpunkte sind die Akzeptanzforschung von Energietechnologien, methodische Aspekte der Umfrageforschung und Zukunftsforschungsmethoden.

David Stich ist ehemalige wissenschaftliche Hilfskraft des Mannheimer Zentrums für Europäische Sozialforschung. Sein Arbeitsschwerpunkt war empirische Arbeitsmarktforschung.

Wolfgang Strengmann-Kuhn ist Mitglied des Deutschen Bundestages und sozialpolitischer Sprecher der Grünen Bundestagsfraktion. Er ist Volkswirt und hat u. a. am Fachbereich Wirtschaftswissenschaften der Goethe-Universität Frankfurt gelehrt und geforscht, wo er 2002 promovierte und sich 2006 habilitierte.

Michael Windzio ist Professor für Soziologie mit dem Schwerpunkt Migration und Stadtforschung an der Universität Bremen (SOCIUM). Arbeitsgebiete: Migration und Integration, Netzwerke.