MICRO-LEVEL DETERMINANTS OF MACRO-LEVEL OUTCOMES

THE MICRO-MACRO LINK WITH EMPIRICAL METHODS.

F. BENJAMIN ROSCHE

Explaining how macro-level outcomes emerge from their constituting parts at the micro-level is a complex undertaking. In empirical research, however, statistical methods that feature trivial aggregation functions dominate because methods to study more complex aggregation processes are underdeveloped. In this thesis, I contribute to the development of empirical-statistical methods for the study of micro-macro links. The developed methods complement existing analytical-theoretical approaches, such as agent-based modeling and game theory. An advantage of empirical-statistical methods is that they confront modeled mechanisms with empirical data, which facilitates assessing their relevance, validity, and generalizability.

CHAPTER 1

Treatment effects on within-group and between-group inequality.
An explanatory variance decomposition approach.

CHAPTER 2

A multilevel model for coalition governments.
Uncovering dependencies within and between governments due to parties.

CHAPTER 3

Socioeconomic segregation in adolescent friendship networks.
A network analysis of social closure in US high schools.

MICRO-LEVEL DETERMINANTS OF MACRO-LEVEL OUTCOMES.

THE MICRO-MACRO LINK WITH EMPIRICAL METHODS.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

F. Benjamin Rosche

May 2024

MICRO-LEVEL DETERMINANTS OF MACRO-LEVEL OUTCOMES. THE MICRO-MACRO LINK WITH EMPIRICIAL METHODS.

F. Benjamin Rosche, Ph.D.
Cornell University 2024

Explaining how macro-level outcomes emerge from their constituting parts at the micro-level is a complex undertaking. In empirical research, however, statistical methods that feature trivial aggregation functions dominate because methods to study more complex aggregation processes remain underdeveloped. In this thesis, I contribute to the development of empirical-statistical methods for the study of micro-macro links.

In Chapter 1, I develop a method to analyze the distributional consequences of heterogeneous treatment effects in a population that can be separated into subgroups. The developed approach is based on the descriptive variance decomposition (Western and Bloome 2009). I extend this approach to an explanatory framework by modeling a treatment effect on the mean and variance of each group and then determining how these treatment effects affect the variance within groups, between groups, and overall. I demonstrate the utility of the approach by analyzing the changing effect of motherhood on women's earnings and its consequences for women's earnings inequality between 1980 and 2020. The decomposition of this effect reveals that motherhood increases inequality between economic strata but reduces inequality within them. As the within-group effect is larger than the between-group effect, the results show that the changes in the motherhood effect since 1980 have overall reduced earnings inequality among women. This fact is obscured when only mean differences are examined.

In Chapter 2, I develop a Bayesian multilevel model that conceptually reverses the conventional multilevel model setup to model the effect of lower-lever units on an outcome at a higher level. The model allows researchers to derive aggregation functions empirically if the specific functional form is unknown. I accomplish this by including a weighted sum in the linear predictor so that the aggregation weights of lower-level units in their effect on an outcome at a higher level can be modeled as a function of observed explanatory variables. I demonstrate the model's utility with an empirical application to the survival of coalition governments as predicted by parties' financial dependency on their members. The results show

that the more parties' financial resources comprise contributions from their members, the higher the termination hazard of governments including those parties. Analyzing the aggregation function, I find that parties' weight in the effect depends on their relative seat share in parliament. Therefore, when aggregating the effect of parties' financial dependency on government survival, parties should be weighted by their relative seat share rather than evenly averaged.

In Chapter 3, I use exponential random graph modeling and empirically calibrated simulation to examine the determinants of network structure characteristics, such as network cohesion, centralization, clustering, and composition. The idea is to generate synthetic networks from empirically calibrated exponential random graph models in which the modeled tie-formation mechanisms are sequentially activated to determine how they shape structural characteristics at the network level. I employ this approach to examine the degree, pattern, and determinants of socioeconomic segregation and its relationship to racial segregation in friendship networks in high school. The results show that friendship networks are overall less socioeconomically segregated than they are racially segregated. However, the exclusion of low-SES students from high-SES cliques is pronounced and, unlike racial segregation, unilateral rather than mutual: many friendship ties from low-SES students to high-SES peers are unreciprocated. The decomposition of determinants indicates that about half of the socioeconomic segregation in friendship networks can be attributed to differences in socioeconomic composition between schools. The other half is attributable to students' friendship choices within schools and driven by stratified courses (about 13 percent) as well as racial and socioeconomic preferences (about 37 percent). In contrast, relational mechanisms like triadic closure – long assumed to amplify network segregation – have only minor effects on socioeconomic segregation. These results highlight that SES-integrated friendship networks in educational settings are difficult to achieve without also addressing racial segregation.

BIOGRAPHICAL SKETCH

Benjamin (Ben) Rosche grew up in the Southwest of Germany. Prior to his doctoral studies at Cornell University, he obtained a B.A. in Sociology and Economics (minor) from the University of Mannheim, an M.Sc. in Sociology and Social Research as well as an M.Sc. in Methods and Statistics from the University of Utrecht. At Cornell University, he pursued a Ph.D. in Sociology with a minor in Computer Science under supervision of Michael Macy (Cornell Sociology and Information Science), Filiz Garip (Princeton Sociology), Felix Elwert (UW-Madison Sociology and Statistics), Eleonora Patacchini (Cornell Economics), and Lillian Lee (Cornell Computer Science). During his time in Ithaca, Ben led the Sociology Graduate Student Training Seminar (2018-19), the Recreational Squash Club (2018-20), and the Graduate Student Improv Club (2020-24). Ben was also a member of the Cornell Windsurfing Club (2021-24), volunteered at Big Brothers Big Sisters (2023-24), and took Spanish (2021-24) and bachata (2024) classes. In Fall 2024, Ben will begin a postdoctoral position at Princeton University in the Office of Population Research. In his research, Ben leverages and develops quantitative methods to analyze family and network dynamics as drivers of social inequality. Methodologically, he specializes in the modeling of complex data structures, such as network, spatial, multilevel, and text data.

Dedicated to my mother, Claudia Gabriele Rosche (1952-2022).

ACKNOWLEDGEMENTS

Over the course of my PhD, I have had the privilege of meeting, learning from, working with, and being supported by a great number of extraordinary people. I would like to take this opportunity to thank them although words seem insufficient to express the depth of my gratitude. The truth is this dissertation is a testament not only to my efforts but also to the support and belief of many individuals in my life.

First and foremost, I extend my heartfelt thanks to my core advisors, Michael Macy, Filiz Garip, and Felix Elwert, whose thoughtful advice has served as a lighthouse, guiding me through the fog of research.

Michael, it was a pleasure to be your student. Thank you very much for teaching me how to spend my time and energy (and how not to), how to identify and highlight contributions of my research while acknowledging their roots in prior work, and how to write winning grant proposals. Your mentorship has shaped not just this work but also my growth as a scholar.

Filiz, I would like to express my deepest appreciation for always having my back. Like no one else you were ready to assist, advise, and inspire me when I needed it. You have offered invaluable feedback on all my projects (I still have the handwritten letter I got from you in my first year). You also guided me through moments of academic disillusionment by demonstrating that academics can be engaged global citizens working for a better world. I am very excited to begin the postdoc position in your lab!

Felix, I am tremendously grateful for helping me understand, adapt to, and integrate into American sociology—a process that was not always smooth sailing. I also appreciate our many substantive discussions about shared interests. I fondly remember our reading group on peer effect modeling in my first semester and our debate about causal inference in social network analysis on the way to the airport. Your passion for these topics has been incredibly inspiring.

I would also like to extend my sincere thanks to my other committee members, Eleonora Patacchini in economics and Lillian Lee in computer science. Eleonora, many thanks for believing in my research and developing my understanding of peer effect modeling in economics. Lillian, thank you for showing me how natural language processing and machine learning can be applied to study social science questions.

My committee was not the only pivotal factor in my development as a social scientist. I would be remiss not to mention Armin Nassehi at the Ludwig Maximilian University of Munich, who delivered such an inspiring introduction to sociology that I switched my major from computer science to sociology after my first year at university. I would also like to thank Hartmut Esser and Clemens Kroneberg at Mannheim University. The sociology textbooks Hartmut wrote deeply influenced my thinking and inspired my interest in theory-driven, methodologically rigorous, quantitative social research. Clemens, thank you for choosing me as your research assistant, recommending me for a scholarship from the Studienstiftung des Deutschen Volkes, and shaping my interest in social integration. You were my first role model in sociology, and you remain one today. I would also like to thank Vincent Buskens, Ineke Maas, and Jeroen Weesie at Utrecht University. Vincent, Ineke, and Jeroen were the best teachers anyone could ask for in the incredible master's program "Sociology and Social Research". Jeroen motivated me to pursue statistics and to take on a project on multilevel modeling that was so ambitious that it would take my entire PhD to finish it (Chapter 2). He read my master's thesis drafts many times (leaving comments like "meaningless" or "word music"), went through my code (spoiler alert, which he thought was terrible), and gave me individual lectures on multilevel modeling. Jeroen, you taught me to think, write, and program. I cannot thank you enough for this investment in me!

I would like to express my deepest appreciation to Alison Schultz. Alison guided me in ways that no faculty could. She pushed me to view my research outside of the context of an academic discipline, to think about how research can improve the lives of those far away from Ivy League universities without compromising scientific integrity or crossing over into activism. It was Alison who pushed me to walk the talk and do Big Brothers Big Sisters again, in order to create real cross-SES friendships. Alison, thank you for inspiring and showing me what people can and should achieve beyond their narrow self-interests.

I also wish to thank everyone in the Social Dynamics Lab, including Tom Davidson, Ari Decter-Frain, Shiyu Ji, Austen Mack-Crane, Mario Molina, Alex Ruch, Nathália Santos, Michael Tang, and Natalie Tong, and Sterling Williams-Ceci, whose camaraderie and intellectual exchanges have immeasurably enriched my experience at Cornell and beyond. Thanks to the interdisciplinary nature of the lab, I also gained a better understanding of sociology as one of many disciplines working on social science questions, which revived my interest in computer science and led to my PhD minor in computer science.

I am thankful for my other colleagues at Cornell, including Kendra Bischoff (thanks for your mentorship!), Mauricio Bucca, Benjamin Cornwell, Lucas Drouhot (thanks for your mentorship; I will never forget our bromance on Sardinia!), Carrie D'Aprix, SteVon Felton, Lauren Griffin, Matt Hall, Dan Hirschman, Tristan Ivory (thanks for your mentorship!), Wonjeong Jeong, Sang Kyung Lee, Samantha Loew (thanks for your help with the grant!), Vida Maralani (thanks for your mentorship!), Alec McGail, Kelly Musick (thanks for your mentorship!), Duke Mwedzi, Marcus May (I loved our tennis matches!), Susan Meyer, Kaye Nantah, Victor Nee, John Niederbuhl (I loved our runs!), Barum Park (thanks for your mentorship!), Emily Parker, Radu Pârvulescu, Karl Pillemer, Camille Portier, Cody Reed, Peter Rich, Tony Sirianni, David Strang, Kim Weeden (thanks for your mentorship!), Cristobal Young (thanks for the fun evenings around the bonfire!), Haowen Zheng, and Zhipeng Zhou.

I gratefully acknowledge Cornell University, the Department of Sociology, the Brooks School of Public Policy, the Cornell Population Center for the generous funding, the freedom to pursue my research projects, and all the courses, seminars, workshops, and other events they offered. These institutional provisions and a grant from the National Science Foundation (#2049207) allowed me to focus entirely on my research. Graduate students must realize at some point that there are simply too many events to attend them all; for a while, however, I certainly tried.

I would also like to thank the many individuals whose invaluable feedback pushed my projects further. <u>Introduction:</u> Alex Colket, David Kretschmer. <u>Chapter 1:</u> Mauricio Bucca, Felix Elwert, Ian Lundberg, Doug Miller, Kelly Musick, Lucas Sage, Alison Schultz, and Ang Yu. <u>Chapter 2:</u> Fiona Coleman, Filiz Garip, Jeff Gill, Zac Greene, David Kretschmer, George Leckie, Lanny Martin, Alec McGail, Thomas Meyer, Thomas Saalfeld, Alison Schultz, Markus Wagner, Jeroen Weesie, and Cristobal Young. <u>Chapter 3</u>: Richard Alba, Kendra Bischoff, Michelle Crow, Lucas Drouhot, Scott Duxbury, Felix Elwert, Filiz Garip, David Grusky, Tristan Ivory, David Kretschmer, Pavel Krivitsky, Clemens Kroneberg, Diego Leal, Philip

Leifeld, Austen Mack-Crane, Michael Macy, Rourke O'Brien, Barum Park, Eleonora Patacchini, Peter Rich, Matthias Schmidtblaicher, Alison Schultz, Thomas Soehl, and Kim Weeden.

Completing this thesis has led me to reflect on the extraordinary path my life has taken. It seems almost unfathomable that a young boy, raised by a single mother of five, shuffled between social housing and foster homes, and faced with an incarcerated parent, would one day pursue a PhD at Cornell University and later a postdoc at Princeton University. The trials of my early years left an indelible mark, however, by profoundly shaping the direction of my scholarly pursuits. The social and economic disparities between our family and others that I experienced during childhood and adolescence crucially shaped my later interest in family demography and inequality. Unlike many other children in similar situations, I was fortunate to befriend children from middle-class backgrounds whose parents often supported me. Possibly for this reason, I am the only one in my family who had the privilege to attend college and graduate school. This experience is the foundation of my research into cross-SES friendship and socioeconomic attainment.

Therefore, I would like to express my deepest gratitude to my family—my siblings, Julia, Amanda, Alexander, and Angela, and my mother, Gabriele Rosche. We each had our challenges to overcome, my siblings facing even greater obstacles than I did. I am proud of how we all fought for a better life and continue to do so. I am indebted to my older sister Angela, my great aunt Elsa Krämer, and the Tatar family (Cigdem, Özlem, Sebnem, and Imral), who helped raise me and my younger siblings when our mother could not.

Words cannot express my gratitude and appreciation for my surrogate family: Simone Hölzermann, Alison Schultz, Fiona Coleman, Matthias Schmidtblaicher, Edith de Leeuw, and Joop Hox, whose love and support have been the bedrock of my determination and resilience. Simone, you are my best friend! Thank you for being my unwavering ally since high school, for being there for me in the best and worst moments, and for all the adventures, joy, and laughter we have shared over the years. Fiona, my heart is full of gratitude for the countless beautiful memories we have created together. Thank you for exploring Ithaca and the surrounding areas with me through picnics after school and camping on weekend, showing me the value of a healthy lifestyle, bringing running into my life and training me up to a half-marathon, sharing my love for Latin American culture, and your unfailingly good mood. I will cherish these memories forever, as I cherish you. Mathi, I deeply appreciate our long-standing friendship. Thank you for your generous support during my PhD and for always being up for adventures! A special thanks also to Edith and Joop, my surrogate academic parents, who took me under their wings, attended both my master's and my PhD ceremonies, and financially supported my move to the US.

I was so lucky to have met many wonderful people in Ithaca. Audrey and Alex, thank you so much for integrating me into life in Ithaca outside of Cornell. I will forever cherish our conversations about love and life, improv therapy, working on our standup routines, poker tournaments with the gang, pool on Sundays at the Rhinehouse, and, last but not least, wild parties! I also send some love to Austin Shaber, my little brother—an incredibly kind young man—for showing me the world from a child's perspective again. To my friends in Ithaca, including Naman Agrawal, Krisztina Ambrus, Leigh Anne Hahn, Ali Ilahi, Eun A Jo, Seth Mackey, Caro Marraccini, Aaron Munzer, Kimberly Navarro-Vélez, Mike Nave, Rachel Neugarten, Matías Oviedo, Luis Perez (with whom I founded the Cornell graduate student improv club in 2020), Sonia Petrutiu, Sam Porter, Abhishek Ray, Guillaume Reboul, Gemma Del Rossi, Zarmeen Salim, Omar Shehata, Federico Skrzypicki, Janet Smith, Camillo Stubenberg, John Tanis, Ralitsa Todorova, Paul Vermette, Miles Wetherington, and Julia Zakov, to my friends elsewhere, including Danelien van Aalst, Alex Dicks, Loes Bouman, Simon Gann, Jarek Habermann, Lena Hirsch, Marijn Keijzer, Martin Keuler, David Kretschmer, Mathijs Kros, Sam Langton, Daniel Pfaff, Fabian Schuy, and Whitney Yoder, to all the folks I met through squash, windsurfing, improv, werewolf nights, bachata, and Big Brothers Big Sisters, I am forever grateful for the laughter and respite you provided amidst the toil.

I dedicate this thesis to my mother, who passed away in 2022. While she couldn't understand the path I had chosen, she taught me never to give up, not to be intimidated by authority, and to empathize with the truly disadvantaged. She deserves more recognition than I can ever give.

Ithaca, May 1, 2024

TABLE OF CONTENTS

Biographical sketch v

Acknowledgements vii

Table of Contents xii

Introduction 1

Chapter 1: Treatment effects on within-group and between-group inequality. An explanatory decomposition approach 11

Chapter 2: A multilevel model for coalition governments: Uncovering dependencies within and between governments due to parties 47

Chapter 3: Socioeconomic segregation in adolescent friendship networks: A network analysis of social closure in US high schools. 73

Conclusion 123

Supplementary Material: Chapter 1 124

Supplementary Material: Chapter 2 133

Supplementary Material: Chapter 3 158

INTRODUCTION

The explanation of social phenomena through the heterogeneous and interdependent actions of individuals is a core aim of the social sciences. It is what differentiates the social sciences from psychology, in which the explanation of individual behavior is an end in itself. Explaining social phenomena as the aggregate consequences of determinants at the micro level can be a complex undertaking because individuals have different opportunities and restrictions, they follow different logics, and they depend on each other in their actions. Statistical regularities in collective behavior are therefore often unintended by the involved individuals and difficult to explain from their individual properties alone.

Analytical-theoretical research has dedicated much attention to formalizing the link between micro- and macro-level. The explanatory framework was coined "methodological individualism" (Schumpeter 1908; Weber 1922), codified in the well-known "macro-micro-macro" scheme (Coleman 1986), and developed into the field of analytical sociology (Hedström and Bearman 2009). Computational (e.g., agent-based simulation) and mathematical modeling (e.g., game theory) have emerged as key methodological approaches. These methods allow researchers to model complex aggregation processes to demonstrate how macro-level outcomes emerge from their constituting parts (i.e., show generative sufficiency). However, due to their tenuous link to empirical data, it is inherently difficult with analytical-theoretical methods to establish whether modeled explanations are relevant, valid, and generalizable to specific empirical settings.

In empirical research, in contrast, statistical methods that feature trivial aggregation functions dominate and researchers often ignore the aggregation problem (i.e., default to the function implied by their chosen method). For instance, the aggregation function of the workhorse of empirical research, the linear regression model, is a linear equation for the conditional mean. However, many interesting social phenomena are difficult to describe through conditional means, especially those that capture two facts about social life: individual variation and mutual interdependence. Although individual variation can be examined with conditional means, other aggregation functions, such as the variance, are usually more effective for understanding population heterogeneity. Similarly, mutual interdependence is difficult to examine with the linear regression model because it assumes the conditional independence of observations. Overall, empirical-statistical methods for the study of complex aggregation processes are woefully undeveloped.

In this thesis, I contribute to the development of empirical-statistical methods for the study of micro-macro links. In Chapters 1 and 2, I develop methods to model the macro-level consequences of individual variation. In Chapter 3, I develop an approach to model the macro-level consequences of mutual interdependence.

Micro-macro models to examine individual variation

Describing and explaining individual variation is at the core of the social sciences as population science, which perceives population heterogeneity not as a nuisance but as an outcome of interest (Xie 2013). The description of individual variation with statistical methods is relatively straightforward because all necessary quantities can be observed. For example, we can compute the variance across observations to describe heterogeneity in a population. In contrast, explaining individual variation and its macro-level consequences is more challenging with statistical methods because causal effects can be defined but not observed at the individual level. The effect of a treatment x_i on an outcome y_i for an individual i is defined as the difference in potential outcomes associated with the different treatment states: $\delta_i = y_i(x_i = 1) - y_i(x_i = 0)$. The fundamental problem of causal inference is that, for any given i, we only observe either $y_i(x_i = 0)$ or $y_i(x_i = 1)$. In other words, the individual causal effect δ_i cannot be observed. We can, however, observe the average treatment effect (ATE) by invoking two assumptions: no systematic pre-treatment heterogeneity and no systematic treatment effect heterogeneity between those who receive the treatment and those who do not. In this case, $\text{ATE} = E[y_i(x_i = 1)] - E[y_i(x_i = 0)]$. Therefore, we must aggregate (average over) individual-level variation in the target population to identify the causal effect.

This creates an interesting fault line for empirical research aiming to explain social phenomena through their constituting parts. Researchers must average over individual variation to identify causal effects. High levels of aggregation, however, leave them with less variation to model the aggregation process. In contrast, low levels of aggregation leave more variation to model but also require more granular data to identify causal effects. Therefore, how much the black box can be opened depends on the available data and on the capacities of researchers and methods to model the aggregation process. This is the fault line I navigate in the first chapter of this dissertation.

Chapter 1: Treatment effects on within-group and between-group inequality. An explanatory decomposition approach.

In Chapter 1, I contribute to the study of inequality by developing a method to analyze the distributional consequences of heterogeneous treatment effects in a population that can be separated into subgroups. I thus open the black box up to the group level by differentiating a target population into $j = 1, \dots, J$ exhaustive and mutually exclusive groups.

The developed approach extends the descriptive variance decomposition—which is based on the analysis of variance (ANOVA) theorem—to an explanatory framework (Bloome and Schrage 2019; Western and Bloome 2009). The ANOVA theorem states that the variance V of an outcome y in a population of individuals i that can be differentiated into groups j can be decomposed into within- and between-group components: $\mathrm{V}(\mathrm{y_i}) = \underbrace{\sum_{\mathrm{j}} \pi_{\mathrm{j}} \sigma_{\mathrm{j}}^2}_{\text{within-group variance}} + \underbrace{\sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(\mu_{\mathrm{j}} - \sum_{\mathrm{j}} \pi_{\mathrm{j}} \mu_{\mathrm{j}}\right)^2}_{\text{between-group variance}}$, where μ_{j} and σ_{j}^2 are group-specific means and variances and π_{j} is the proportion of individuals in group j. The theorem links the individual observations, $\mathrm{y_i}$, to an aggregate outcome, the variance in y within and between groups, to reveal underlying patterns of inequality in this population.

I modify the ANOVA theorem to accommodate group-specific treatment effects on within-group, between-group, and total variance. Specifically, using grouped heteroscedasticity regression (Harvey 1976), I model a treatment effect on the mean and variance of each group and then determine how these treatment effects affect the variance within groups, between groups, and overall. This approach, therefore, averages over individual-level variation up to the group level to identify treatment effects and then determines mathematically how these group-specific treatment effects impact the population-level variance.

I develop this method both for cross-sectional and longitudinal analyses. With longitudinal analyses, researchers can disentangle compositional changes (i.e., changes in pre-treatment inequality and in the distribution of treatment across groups) from behavioral changes (changes in treatment effects). Both changes relative to a timepoint (e.g., 1980) and changes relative to a counterfactual scenario (e.g., a counterfactual distribution of treatment) can be analyzed. With the paper, I provide the R package "ineqx" that implements both descriptive (Western and Bloome 2009) and explanatory variance decomposition. The paper thus adds to the toolkit of methods to study inequality determinants developed in sociology. While the method is still in an early stage of development, I hope that the paper will inspire subsequent research that will refine and strengthen it.

I apply this approach to examine the how the effect of motherhood on women's earnings inequality has evolved from 1980 to 2020. I differentiate women by total household income into low-, medium-, and high-SES and demonstrate that motherhood has increased inequality between socioeconomic strata but reduced inequality within them. As the within-group effect is larger than the between-group effect, changes in the motherhood effect between 1980 and 2020 have overall reduced earnings inequality among women. This fact is obscured when only mean differences are examined. This result highlights the risk of drawing misleading conclusions when researchers choose aggregation functions that do not properly capture the macro-level consequences of examined micro-level processes.

Chapter 2: "A multilevel model for coalition governments: Uncovering dependencies within and between governments due to parties."

In the first chapter, the micro-macro link is defined by the "analysis of variance" theorem and thus fixed. In the second chapter, I develop a model that allows researchers to derive aggregation functions from data. The developed Bayesian multilevel model conceptually reverses the conventional multilevel model setup by modeling the effect of lower-lever units on an outcome at a higher level. Building on Goldstein (2011) and Snijders (2016), I accomplish this by including a weighted sum in the linear predictor so that the aggregation weights of lower-level units in their effect on an outcome at a higher level can be modeled as a function of observed explanatory variables.

I apply this model to a multilevel structure that emerges in countries with parliamentary democracies in which coalition governments are the norm. To explain coalition outcomes like their formation or termination, political scientists leverage both party- and government-level explanations. Party-level explanations require the aggregation of party features across all parties that make up a government. In the simplest case, $E[y_i] = \beta\left(\frac{1}{n_i}\sum_k x_k\right)$, where y_i is an outcome of a coalition i of size n_i, β is the effect of an aggregated party feature on that outcome, and x_k is the party feature that is averaged across all parties k that make up that coalition. However, there are reasons to believe that the features of some parties matter more than others. For instance, Laver and Shepsle (1998) theorize that party leaders in hierarchically structured parties have more leverage to administer a coalition than party leaders in democratically structured parties as their party base can constrain leaders' room for negotiation (Laver and Shepsle 1998). This idea can be represented by giving parties differential weights: $E[y_i] = \beta(\sum_k w_k x_k)$. To empirically test this idea, however, we must estimate the party weights, which is what I propose to do.

Let y_i^G be a government-level outcome that is determined by a systematic component ($\beta^g x_i^g$) and a random component (u^g) at the government level, a systematic component ($\beta^p x_{ik}^p$) and a random component (u_{ik}^p) at the party level, and let the effect of each party k in the set of parties $p(i)$ that constitute government i be aggregated by a weighted sum:

$$y_i^g = \beta^g x_i^g + u_i^g + \sum_{k \in p(k)} w_{ik}\left(\beta^p x_{ik}^p + u_{ik}^p\right)$$
$$\text{with } w_{ik} = \frac{1}{n_i^{\exp(-\{\beta^w x_{ij}^w\})}} \text{ subject to } \sum_{ik} w_{ik} = 1$$

To estimate the party weights, I propose to model them as a nonlinear regression of unobserved w_{ij} on observed explanatory variables x_{ij}^w instead of assigning fixed weights to each party. The proposed weight function reduces to the arithmetic average (i.e., $w_{ij} = \frac{1}{n_i}$) if the explanatory variables do not impact the weights. If $\beta^w \lesseqgtr 0$, the weights differ between parties, ranging between 0 and 1, and constraining the sum of weights within governments to 1. The weight regression can be seen as a measurement model as β^w measures the relative weight of each party in the aggregated party effect. The aggregated party effect, in contrast, is part of the structural model as β^p estimates the impact of parties and party features on the coalition outcome.

This model makes two contributions. First, the model demonstrates how aggregation functions can be empirically derived if the specific functional forms are unknown a priori. This approach has broader application than just in political science. Aggregation functions and crisscrossing data structures are ubiquitous in empirical social research. In network and spatial analysis, weight matrices define the relative weight of neighbors in their effect on nodes or spatial units. The proposed model allows researchers to estimate these weights rather than specify fixed values. This weight regression can include complex aggregation functions (e.g., min, max, mean, or sum) as the model is estimated with Bayesian MCMC, which does not require analytically tractable likelihood functions. Another example is index building, which often involves the aggregation of information across levels using aggregation functions specified by researchers. The proposed model allows researchers to test the hypothesized aggregation functions against data.

Second, the model accounts for uncertainty at the party level by including a party-level random effect so that standard error estimates reflect uncertainty at that level. A conducted simulation study shows that models ignoring uncertainty at the party level exhibit excessive false positive rates. Even though no effect was specified in the simulation, government-level effects are significant over 40 percent of the time and party-level effects are significant over 60 percent of the time because models assume the independence of observations while, in reality, the same parties participate in governments over and over again.

I demonstrate the model's utility with an empirical application to the survival of coalition governments as predicted by the financial dependencies of parties. The results show that the more parties' financial resources comprise contributions from their members, the higher the termination hazard of governments including those parties. Analyzing the aggregation function, I find that parties' weight in the effect depends on their relative seat share in parliament. Therefore, when aggregating the effect of parties' financial dependency on government survival, parties should be weighted by their relative seat share rather than evenly averaged. With this paper, I also contribute to the research infrastructure by providing the R package "rmm" to estimate this model in JAGS (Plummer 2003) from within R.

Micro-macro models to examine mutual interdependence

Statistical methods used in social research often assume that data obtained from human respondents represent independent replications because models predicated on the independence assumption are mathematically convenient. In many cases, however, social research is concerned with the interdependence among individuals, and about how their interdependent actions shape social phenomena. Research in analytical sociology moves away from the assumption of independence by focusing on how social phenomena emerge from individuals being tied to one another in various ways. In the final chapter of this dissertation, I contribute to this research program, using a network modeling approach to examine how students' interdependent friendship choices shape aggregate segregation patterns in US high schools.

Chapter 3: "Socioeconomic segregation in adolescent friendship networks: A network analysis of social closure in US high schools."

A growing body of research in sociology and adjacent fields focuses on how networks and embedded social capital affect social inequality (Blau 1977; Lenkewitz 2023; Simmel 1908; van Tubergen and Volker 2015). A core finding of this line of research is that social embeddedness consolidates rather than alleviates inequality when networks are socioeconomically segregated.

In this chapter, I examine the determinants of socioeconomic segregation in networks, which are important antecedents of the relationship between networks and inequality. Specifically, I examine the mechanisms underlying socioeconomic segregation in friendship networks in high school. The study addresses a key puzzle raised in Chetty et al. (2022), who identify greater socioeconomic homophily in high school friendship networks than prior studies: what is behind this friending bias? A well-developed literature in sociology on the determinants of racial segregation provides answers. This research highlights the importance of structural barriers, such as neighborhoods (Mouw and Entwisle 2006), courses (Frank, Muller, and Mueller 2013), and extracurricular activities (Schaefer, Simpkins, and Ettekal 2018), as well as intentional and unintentional boundary-making between groups with intersecting attributes (Moody 2001; Wimmer 2013; Zhao 2023). However, while these prior studies have focused on different potential determinants, they omitted measures of alternatives that may confound their results. In this chapter, I integrate the disparate perspectives to disentangle their relative contributions to socioeconomic segregation using data from the National Study of Adolescent Health and exponential random graph modeling (ERGM).

ERGM is a statistical network modeling approach that allows researchers to model friendship formation between interdependent individuals and to examine how these interdependent friendship choices shape the aggregate network structure (Duxbury 2024; Robins, Pattison, and Woolcock 2005; Snijders and Steglich 2015). This is accomplished via empirically calibrated simulation. The approach is to generate synthetic networks from empirically calibrated exponential random graph models, in which the modeled friendship determinants (e.g., homophilous tendencies) are sequentially activated to determine how they shape socioeconomic segregation in the networks. In this way, friendship determinants at the micro level are directly linked to their structural implications at the macro level. While most prior applications of ERGM use the approach to study friendship formation, I show in this paper how it can be leveraged to examine micro-macro links. As such, the paper provides a blueprint for future research into the determinants of other macro-level network features, such as network cohesion, centralization, clustering, and composition.

The results show that socioeconomic segregation in friendship networks in high school is characterized by exclusion of students in the bottom third and closure among students in the upper half of the SES distribution. While friendship networks are overall less socioeconomically segregated than they are racially segregated, the exclusion of low-SES students from high-SES cliques is as pronounced. Moreover, unlike racial segregation, socioeconomic segregation is unilateral rather than mutual: many friendship ties from low-SES students to high-SES peers are unreciprocated. The decomposition of determinants indicates that about half of the socioeconomic segregation in friendship networks can be attributed to differences in socioeconomic composition between schools. The other half is attributable to students' friendship choices within schools and driven by stratified courses (about 13%) as well as racial and socioeconomic preferences (about 37%). In contrast, relational mechanisms like triadic closure – long assumed to amplify network segregation – have only minor effects on socioeconomic segregation. The large impact of racial homophily suggests that SES-integrated friendship networks in educational settings are difficult to achieve without also addressing racial segregation.

REFERENCES

Blau, Peter. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Bloome, Deirdre, and Daniel Schrage. 2019. "Covariance Regression Models for Studying Treatment Effect Heterogeneity Across One or More Outcomes: Understanding How Treatments Shape Inequality." *Sociological Methods & Research* 50(3).

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022. "Social Capital II: Determinants of Economic Connectedness." *Nature* 608(7921):122–34.

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91(6):1309–35. doi: 10.1086/228423.

Duxbury, Scott W. 2024. "Micro Effects on Macro Structure in Social Networks." *Sociological Methodology* 54(1):1–26. doi: 10.1177/00811750231209040.

Frank, Kenneth A., Chandra Muller, and Anna S. Mueller. 2013. "The Embeddedness of Adolescent Friendship Nominations: The Formation of Social Capital in Emergent Network Structures." *American Journal of Sociology* 119(1):216–53. doi: 10.1086/672081.

Goldstein, Harvey. 2011. *Multilevel Statistical Models*. John Wiley & Sons.

Hedström, Peter, and Peter Bearman. 2009. *The Oxford Handbook of Analytical Sociology*. OUP Oxford.

Laver, Michael, and Kenneth A. Shepsle. 1998. "Events, Equilibria, and Government Survival." *American Journal of Political Science* 42(1):28–54. doi: 10.2307/2991746.

Lenkewitz, Sven. 2023. "Limited Opportunities: Adolescents' Access to Social Capital in Secondary Schools in Three European Countries." *Social Networks* 74:245–58. doi: 10.1016/j.socnet.2023.05.001.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107(3):679–716. doi: 10.1086/338954.

Mouw, Ted, and Barbara Entwisle. 2006. "Residential Segregation and Interracial Friendship in Schools." *American Journal of Sociology* 112(2):394–441. doi: 10.1086/506415.

Plummer, Martyn. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

Robins, Garry, Philippa Pattison, and Jodie Woolcock. 2005. "Small and Other Worlds: Global Network Structures from Local Processes." *American Journal of Sociology* 110(4).

Schaefer, David R., Sandra D. Simpkins, and Andrea Vest Ettekal. 2018. "Can Extracurricular Activities Reduce Adolescent Race/Ethnic Friendship Segregation?" Pp. 315–39 in *Social Networks and the Life Course: Integrating the Development of Human Lives and Social Relational Networks*, edited by D. F. Alwin, D. H. Felmlee, and D. A. Kreager. Cham: Springer International Publishing.

Schumpeter, J. A. 1908. *Das Wesen Und Der Hauptinhalt Der Theoretischen Nationalökonomie*. Duncker und Humblot.

Simmel, Georg. 1908. "Die Kreuzung Sozialer Kreise." in *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*. Leipzig: Duncker & Humblot.

Snijders, T. A. 2016. "The Multiple Flavours of Multilevel Issues for Networks." in *Multilevel Network Analysis for the Social Sciences*, edited by E. Lazega and T. A. Snijders. Springer.

Snijders, Tom A. B., and Christian E. G. Steglich. 2015. “Representing Micro–Macro Linkages by Actor-Based Dynamic Network Models.” *Sociological Methods & Research* 44(2):222–71. doi: 10.1177/0049124113494573.

van Tubergen, Frank, and Beate Volker. 2015. “Inequality in Access to Social Capital in the Netherlands.” *Sociology* 49(3):521–38.

Weber, Max. 1922. *Wirtschaft Und Gesellschaft*. Tübingen, Germany: Mohr.

Western, Bruce, and Deirdre Bloome. 2009. “Variance Function Regressions for Studying Inequality.” *Sociological Methodology* 39(1):293–326.

Wimmer, Andreas. 2013. *Ethnic Boundary Making: Institutions, Power, Networks*. Oxford University Press.

Xie, Yu. 2013. “Population Heterogeneity and Causal Inference.” *Proceedings of the National Academy of Sciences* 110(16):6262–68. doi: 10.1073/pnas.1303102110.

Zhao, Linda. 2023. “From Superdiversity to Consolidation: Implications of Structural Intersectionality for Interethnic Friendships.” *American Journal of Sociology* 128(4).

CHAPTER 1

TREATMENT EFFECTS ON WITHIN-GROUP AND BETWEEN-GROUP INEQUALITY AN EXPLANATORY DECOMPOSITION APPROACH

ABSTRACT

Rising inequality has been linked to growing disparities within and between economic strata. Yet, existing approaches to analyzing inequality often disregard within-group inequality and are limited in addressing causal questions about why inequality is changing. This paper introduces an explanatory approach to examining how treatment variables impact within-group, between-group, and total inequality. The method permits both cross-sectional and longitudinal analyses. With longitudinal analyses, researchers can disentangle compositional changes (level of pre-treatment inequality, distribution of treatment across groups) from behavioral changes (changing treatment effects). Moreover, researchers can analyze changes relative to a timepoint (e.g., 1980) or relative to a counterfactual scenario (e.g., a counterfactual distribution of treatment). I demonstrate the utility of the approach by analyzing the changing effect of motherhood on women's earnings and its consequences for women's earnings inequality between 1980 and 2020. The results show that motherhood decreases women's earnings inequality because it reduces inequality within economic strata.

INTRODUCTION

The continuous rise in income inequality in the United States and several other countries over the past thirty years has rekindled interest in analyzing determinants of income inequality (McCall and Percheski 2010; Pew Research Center 2020). Many empirical methods in the social sciences (such as linear regression), however, are ill-suited to the analysis of income inequality because they are tailored toward explaining a distribution's central tendency while inequality statistics measure its dispersion. Popular inequality statistics, such as the Gini coefficient, the variance, or quantile ratios, are not only a function of the mean but also of quantiles or higher moments of the income distribution. Further, approaches to the study of income inequality that are solely based on mean differences between groups, such as the Kitagawa-Blinder-Oaxaca (KBO) decomposition (Blinder 1973; Kitagawa 1955; Oaxaca 1973) or the gap-closing estimand (Lundberg 2022), leave out an important component of the total income inequality—variation within groups. These approaches shift within-group variation to the unexplained part, which limits our understanding of changes in the income distribution as a whole. In fact, studies consistently find that much of the increase in inequality over the past thirty years can been attributed to an increase in "residual inequality", that is, inequality within demographic groups (Juhn, Murphy, and Pierce 1993; Lemieux 2006; Western, Percheski, and Bloome 2008; Wodtke 2016)[1].

The distinction between within- and between-group inequality is not only statistically but also substantively important. Most research on inequality differentiates between groups—be it by race, education, income quantile, or country. Differentiating inequality within and between these groups helps researchers gain a better understanding of the origins and consequences of inequality. Within-group inequality measures disparities between members of the same group, which researchers can use to identify whether some groups are internally more unequal than others. This source of inequality across groups is

[1] The definition and number of groups obviously influences what is identified as within- and between-group change. However, even with fine-grained partitions, studies find that much growth in inequality in the past thirty years is due to an increase of inequality within demographic groups (e.g., Western, Percheski, and Bloome 2008, who differentiate 300 groups).

qualitatively different from disparities in group means (i.e., between-group inequality). Indeed, owing to spatial segregation (Reardon and Bischoff 2011) and social comparison processes (Festinger 1954), people tend to experience within-group inequality much more directly than between-group inequality. Research on relative deprivation further highlights that rising within-group inequality tends to cause frustration with oneself while rising between-group inequality tends to cause frustration with outgroup members (Smith et al. 2012). Therefore, within- and between-group inequality measure distinct facets of social differentiation, differ in their consequences for social cohesion, and are thus both critical for our understanding of inequality. In fact, key fields of inequality research have gained leverage by differentiating within- and between-group inequality, such as work on the gender wage gap (Ferro-Luzzi 2010), class-based disparities (Weeden et al. 2007; Wodtke 2016), or global inequality (Firebaugh 2003; Goesling 2001).

Not all inequality statistics are decomposable into within- and between-group components. Quantile ratios, for instance, are not decomposable in this way. By contrast, the variance, the coefficient of variation, and—under certain conditions—the Gini coefficient can be decomposed into within- and between-group components (Fortin, Lemieux, and Firpo 2011). Western and Bloome (2009) propose an approach to decomposing the variance to analyze trends in within- and between-group inequality. Their method, however, cannot address causal questions as it does not differentiate between covariates determining the groups and treatment variables impacting group outcomes. Despite the descriptive nature of the method, researchers employing it often rationalize their results even though such explanations are speculative. Wodtke (2016), for instance, explains the increase in inequality between social classes with the growing economic concentration, technological displacement of workers, and shifts in bargaining power, but is unable to actually test any of these hypotheses. Accordingly, to make the variance decomposition approach amendable to research of *why* within- and between-group inequality is changing, it must be translated into an explanatory framework.

In this paper, I propose an explanatory variance decomposition approach that enables researchers to measure treatment effects on variance-based inequality statistics and decompose the effects into within-and between-group components. The paper contributes to the development of methods to study inequality in five ways.

First, I advance a treatment effect framework to quantify the effect of a treatment on the variance and variance-based inequality measures—both cross-sectionally and over time. The approach can be employed in experimental research to examine treatment effects and in observational research to examine predictor effects. In experimental research, scientists increasingly employ more complex designs in which treated subjects are differentiated by groups (e.g., by race, village, treatment arm) (Baldassarri and Abascal 2017) and are increasingly interested in analyzing treatment effect heterogeneities along the outcome distribution (Hohberg, Pütz, and Kneib 2020). In observational research, scientists are moving toward experimental thinking and modern methods of causal inference (Morgan and Winship 2014). The proposed approach contributes to these developments by proposing an explanatory framework to measure the impact of treatment variables beyond the mean.

Second, I use this framework to extend the descriptive variance decomposition approach to decomposing treatment effects on the variance into within- and between-group components. This treatment effect decomposition helps researchers to not only identify the causes of rising inequality (why?) but also their mechanisms (how?) by linking individual-level effects to their consequences for group-level and total inequality (Goldthorpe 2015; Jackson 2022; Xie 2007, 2013). This is important because burgeoning causal research indicates that there are different mechanisms underlying changes in within- and between-group inequality. The growth in between-group inequality has been linked to, among other things, increasing returns to education (Autor, Levy, and Murnane 2003) and capital ownership (Piketty 2014). The surge of within-group inequality has been linked to labor market deinstitutionalization causing greater uncertainty within groups (Massey 2007), rising returns to unobserved skills (Juhn et al. 1993), and an increase in measurement error (Lemieux 2006). The proposed explanatory variance decomposition approach contributes to disentangling such within- and between-group mechanisms.

Third, I highlight that careful attention must be paid to the basic axioms of inequality measurement when decomposing the variance into within- and between-group components. Existing variance decomposition approaches often decompose the variance of log-income rather than the variance of income (e.g., Juhn et al. 1993; Lemieux 2006; Weeden et al. 2007; Western and Bloome 2009; Western et al. 2008; Wodtke 2016; Xie, Killewald, and Near 2016). The variance of logarithms, however, is *not* additively decomposable into within- and between group components because it does not respect the Pigou-Dalton principle of transfers (Cowell 1988, 2011; Foster and Ok 1999). In the paper, I discuss this issue and provide recommendations on how to circumvent the issue.

Fourth, I apply the approach to examining the motherhood effect on women's earnings inequality and find that motherhood reduces inequality in women's earnings. This is true on an absolute level (compared to motherhood having no effect on earnings) and relative level (compared to 1980). The decomposition reveals that motherhood primarily reduces inequality within economic strata. The motherhood effect on inequality between economic strata is small relative to the within-group effect. This result highlights the risk of drawing misleading conclusions when inequality researchers base their analyses solely on mean differences between groups.

Fifth, I introduce the R library *ineqx,* which implements both descriptive and explanatory variance decomposition approach. Despite the importance of analyzing within- and between-group inequality, relatively little empirical work has taken this on. This user-friendly R library, which is presented in Appendix 1, is intended to facilitate and popularize research using variance decomposition approaches.

The paper proceeds as follows. In Section 2, I review existing approaches to delineate the contribution. I develop the explanatory variance decomposition in Section 3. Section 4 discusses the issue of using log-income in the context of variance decompositions. In Section 5, I employ both descriptive and explanatory variance decomposition to study trends in women's earnings inequality and examine how the changing impact of motherhood on earnings has contributed to these trends. Section 7 concludes.

METHODS TO STUDY INEQUALITY

Methods to study the effect of covariates on distributional statistics other than the mean has been an active research area in the last decades (Fortin et al. 2011). This literature has produced a number of approaches to examine trends in inequality and their causes. In the following, I review these approaches to provide an overview and to delineate the contribution of the explanatory variance decomposition.

An approach popular in sociology is to decompose inequality statistics into contributions by income source and their inter-correlations (Cancian and Reed 1999; Karoly and Burtless 1995; Schwartz 2010). Family demographers, for instance, have decomposed the variance of couple income into contributions by spouse to model the correlation among the two components (i.e., spousal income correlation). An advantage of this approach is that results are easy to interpret as an increase in the correlation among income components corresponds to an increase in inequality. A disadvantage, however, is that a correlation does not tell us anything about the behavior of individuals (e.g., an increase in the spousal income correlation does not tell us if one spouse is making more or the other one is making less). Another disadvantage is that it is an aggregate-level analysis. Gonalons-Pons, Schwartz, and Musick (2021), for instance, apply this approach to examine the effect of parenthood on inequality by measuring couples' earnings correlation before and after childbirth. The difference in the correlation before and after childbirth, however, cannot be interpreted as a causal parenthood effect, as the correlations as well as their difference are aggregate quantities. The approach proposed in this paper, by contrast, defines treatment effects at the individual level, which, in principle, allows for a more causal interpretation.

Another approach is to decompose inequality statistics by population subgroup (Shorrocks 1984). In this approach, researchers partition a population into mutually exclusive and exhaustive subgroups to then decompose total inequality into inequality within and between groups. The analysis of variance technique (ANOVA) is the basic idea behind this decomposition. Western and Bloome (2009) combine this classic variance decomposition approach with variance regression (Harvey 1976) to model the group means and variances (as functions of covariates) rather than to read them directly from the data. The advantage of this approach is that researchers can examine changes in the effect of variables that determine the groups. In an application of the approach, for instance, Western, Bloome, and Percheski (2008) define 300 groups by intersecting four categorically coded variables (education, race, age, and family type). By comparing the actual development of inequality to a counterfactual development where the regression coefficient of, say, education had not changed during the observed period, they examine changes in the effect of education on within- and between-group inequality. This approach is descriptive because education (is part of what) determines the groups. The approach developed in this paper extends this descriptive approach to an explanatory framework by differentiating between grouping and treatment variable so that changes in the effect of a treatment administered to the groups can be examined.

An approach similar to the approach proposed in this paper is Lemieux (2002), which unifies the KOB-type decomposition of the variance by Juhn, Murphy, and Pierce (JMP) (1993) and the re-weighting procedure by DiNardo, Fortin, and Lemieux (DFL) (1996). Lemieux's method also decomposes the effect of explanatory variables on inequality into within- and between-group inequality. An important difference, however, is that Lemieux (2002) only specifies an earnings equation (i.e., a mean model to quantify the contribution of explanatory variables on the between-group variance) and uses a re-weighting procedure to quantify the contribution of explanatory variables on the within-group variance. The method proposed in this paper, by contrast, models both mean and variance using variance regression (Harvey 1976), which results in a less convoluted approach and offers more transparency and flexibility regarding which variables go into the model and how.

Finally, recent research has developed explanatory approaches based on conditional (Machado and Mata 2005; Melly 2005) and unconditional (Firpo, Fortin, and Lemieux 2009) quantile regression and extensions thereof (Chernozhukov, Fernández-Val, and Melly 2013; Firpo, Fortin, and Lemieux 2018). Quantile regression is a powerful instrument with which it is possible to estimate the effect of an explanatory variable on the entire distribution. A disadvantage of it, however, is that a large number of quantiles has to be estimated to yield a good approximation of the whole distribution, and that estimates must be constrained to avoid model inconsistencies (e.g., crossing quantiles). By contrast, variance regression—if the model is properly specified—estimates the effect of explanatory variables on the entire distribution with much fewer parameters. Extensions of quantile regression, such as recentered influence function (RIF) regression (Firpo et al. 2018), allow to decompose the effect of an explanatory variable on any distributional measure into KOB-type coefficient and endowment effects. RIF regression, while promising due to its flexibility, is a linearization approach, which might poorly approximate the true impact of a variable.

EXPLANATORY VARIANCE DECOMPOSITION

I develop the explanatory variance decomposition in four steps. I start by reviewing the descriptive variance decomposition on which the approach builds. Subsequently, I add the treatment effect framework, which I first do for a single timepoint and then for changes over time. Finally, I discuss the estimation of treatment effects using variance regression.

Descriptive variance decomposition

Take the vector Y_t to be individual[2] incomes at time t, and the vector G_t to be the group to which the individuals belong, where $G_t = j$ is a categorical variable with $j = 1, \dots, J$ categories that represent mutually exclusive and exhaustive groups. The variance V in income at time t can then be expressed as the sum of the variances within and between groups[3]:

$$V_t(Y_t) = E\big(V(Y_t|G_t)\big) + V\big(E(Y_t|G_t)\big) \tag{1}$$
$$= \underbrace{\sum_j \pi_{jt}\sigma_{jt}^2}_{\text{Within-group inequality}} + \underbrace{\sum_j \pi_{jt}\left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2}_{\text{Between-group inequality}}$$

where π_{jt} is the proportion of individuals in group j at time t, μ_{jt} is the mean income in group j at time t, and σ_{jt}^2 is the variance around this mean in group j at time t.

With repeated cross-sectional or panel data, the change in variance from t_0 (baseline) to t (any timepoint post baseline) can then be decomposed into the sum of a within-group effect (δ_W^T), a between-group effect (δ_B^T), and a compositional effect (δ_C^T). That is,

[2] Individuals are indexed by $i = 1,2, \dots N$, which I suppress to simplify notation.
[3] There is no covariance term because the groups are exclusive and exhaustive (law of total variance).

$$V_t - V_{t_0} = \delta_W^T + \delta_B^T + \delta_C^T, \text{ where} \quad (2)$$

$$\delta_W^T = \sum_j \pi_{jt_0}\left(\sigma_{jt}^2 - \sigma_{jt_0}^2\right)$$

$$\delta_B^T = \sum_j \pi_{jt_0}\left(\left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 - \left(\mu_{jt_0} - \sum_j \pi_{jt_0}\mu_{jt_0}\right)^2\right)$$

$$\delta_C^T = \sum_j \left(\pi_{jt} - \pi_{jt_0}\right)\left(\left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 + \sigma_{jt}^2\right)$$

The between-group effect captures the change in total variance induced by changes in the mean of each group. The within-group effect captures the change in total variance induced by changes in the variance around the mean of each group. Finally, the compositional effect captures the change in total variance induced by changes in the relative size of each group. The superscript T on the δs indicates that the change over time is considered. A derivation of equation (2) can be found in appendix 2.

Decomposing the effect of treatment on inequality

I first entertain the treatment effect framework at a single timepoint (and thus suppress subscript t). Let $D \in \{0,1\}$ be a binary treatment, $Y(D)$ be the potential outcome of individual income, and $\tau = Y(1) - Y(0)$ be the intra-individual causal effect of treatment on the outcome. Moreover, assume that the effect of this treatment on the total variance can be fully described by its effect on the group-specific means and variances. The group-specific treatment effect on the *mean* of the treated (ATT) then equals the expected value of the differences between the potential outcomes of the treated individuals in each group:

$$ATT_j = E[Y(1) - Y(0) \mid G = j, D = 1] \quad (3)$$

Similarly, the group-specific treatment effect on the *variance* of the treated (VTT) then equals the difference between the variances in the potential outcomes of the treated individuals in each group:

$$VTT_j = V[Y(1) \mid G = j, D = 1] - V[Y(0) \mid G = j, D = 1] \quad (4)$$

The VTT is thus defined as the difference in each group between the variance of the treated individuals and the counterfactual variance in which these individuals were not treated[4]. The focus lies on the treatment

[4] The VTT is not the variances of the group-specific treatment effects (i.e., $V[Y(1) - Y(0) \mid G = j, D = 1]$).

effect on the treated individuals (i.e., ATT and VTT) rather than all individuals (i.e., ATE and VTE) because only those individuals who receive the treatment will affect the observed variance.

Given these definitions, the effect of treatment on the variance can be decomposed into a within- and between group component:

$$V[Y(1)|D = 1] - V[Y(0)|D = 1] = \delta_B^D + \delta_W^D, \text{ where} \quad (5)$$

$$\delta_B^D = \sum_j \pi_j \left(\left(\mu_j + \beta_j - \sum_j \pi_j (\mu_j + \beta_j) \right)^2 - \left(\mu_j - \sum_j \pi_j \mu_j \right)^2 \right)$$

$$\delta_W^D = \sum_j \pi_j \left(\sigma_j + \lambda_j \right)^2 - \sum_j \pi_j \sigma_j^2$$

The interpretation of the parameters changes in the explanatory variance decomposition. In the descriptive decomposition (equation 2), π_j is the relative size of each group, and μ_j and σ_j are the mean and standard deviation in each group. In the explanatory variance decomposition (equation 5), by contrast, π_j is the proportion receiving treatment in each group, and μ_j and σ_j are the pre-treatment mean and standard deviation in each group[5]. Further, β_j is the treatment effect on the mean in each group (i.e., ATT_j), and λ_j is the treatment effect on the standard deviation in each group (i.e., $\sqrt{VTT_j}$). The superscript D on the δs indicates that the change caused by treatment (at a single timepoint) is considered.

The between-group effect δ_B^D captures the change in total variance induced by the effect of treatment on the mean of each group. The within-group effect δ_W^D captures the change in total variance induced by the effect of treatment on the variance of each group.

Example

The following example illustrates that it is not possible to directly infer whether inequality will go up or down just from the effect of treatment on the group-specific means and variances (i.e., the βs and the λs). Take the stylized case of an organization with two groups—say workers and managers—of which an equal share receives a bonus, and we are interested in how this treatment affects within- and between-group

[5] Note that in the cross-sectional case, the pre-treatment means and standard deviations are equal to the means and standard deviations of the untreated.

inequality. In this example, the effect of treatment on the between-group variance in equation (5) reduces to

$$\delta_B^D = \frac{(\beta_w - \beta_m)(2\mu_w - 2\mu_m - \beta_m + \beta_w)}{4} \tag{6}$$

Equation (6) shows that the impact of treatment on between-group inequality does not only depend on the treatment effects β_w and β_m themselves but also on the groups' position in the income distribution before treatment, that is, the distance between the pre-treatment group means, μ_w and μ_m. Moreover, while absolute inequality measures (like the variance) do not change when treatment affects the group means in the same sign and magnitude, relative inequality measures (like the coefficient of variation) do change even if $\beta_w = \beta_m$ because they also depend on the distance of the group-specific means to the grand mean prior to treatment[6].

The effect of treatment on the within-group variance in equation (5) reduces to

$$\delta_W^D = \frac{\lambda_w^2 + \lambda_m^2 + 2\sigma_w\lambda_w + 2\sigma_m\lambda_m}{2} \tag{7}$$

Equations (7) shows that the impact of treatment on within-group inequality, likewise, does not only depend on the treatment effects λ_w and λ_m but also on the group's pre-treatment standard deviations, σ_w and σ_m. Moreover, in contrast to the effect on between-group variance, the within-group variance will change even if $\lambda_w = -\lambda_m$ because the effect is a multiplicative of the pre-treatment standard deviations. It is, consequently, difficult to gauge the treatment effect on within-group, between-group, and total inequality by "eyeballing" it. Instead, the effect should be calculated as outlined in this section—in particular if there are more than two groups and the proportion receiving treatment differs across groups.

[6] In this example, the effect of treatment on the between-group CV^2 would be $\delta_B^D = \frac{4\left(\beta_w^2\mu_w\mu_m + \beta_m\mu_w\left(-\mu_w^2 + \mu_m(\beta_m + \mu_m)\right) - \beta_w(-\mu_w^2\mu_m + \mu_m^3 + \beta_m(\mu_w^2 + \mu_m^2))\right)}{(\mu_w + \mu_m)^2(\beta_w + \beta_m + \mu_w + \mu_m)^2}$ and the effect of treatment on the within-group CV^2 would be $\delta_W^D = -\frac{2(\sigma_w^2 + \sigma_m^2)}{(\mu_w + \mu_m)^2} + \frac{2((\sigma_w + \lambda_w)^2 + (\sigma_m + \lambda_m)^2)}{(\beta_w + \beta_m + \mu_w + \mu_m)^2}$. These effects do not generally equal 0 when the treatment effects are equal across groups.

Decomposing the change in the effect of treatment on inequality over time

The previous section showed that the treatment effect on the variance depends on the treatment effect on the group-specific means and variances, on the distribution of treatment across groups, and on the level of pre-treatment inequality. Therefore, with repeated cross-sectional or panel data, the change in total variance from t_0 (baseline) to t (any timepoint post baseline) due to a change in the effect of treatment can be decomposed into the sum of a between-group effect ($\delta_B^{D,T}$), within-group effect ($\delta_W^{D,T}$) a compositional effect ($\delta_C^{D,T}$), and a pre-treatment effect ($\delta_P^{D,T}$)[7]:

$$(V[Y_t(1)|D_t = 1] - V[Y_t(0)|D_t = 1]) - \left(V\left[Y_{t_0}(1)|D_{t_0} = 1\right] - V\left[Y_{t_0}(0)|D_{t_0} = 1\right]\right) \tag{8}$$

$= \delta_B^{D,T} + \delta_W^{D,T} + \delta_C^{D,T} + \delta_P^{D,T}$, where

$$\delta_B^{D,T} = B\left(\pi_{t_0}, \mu_{t_0} + \beta_t\right) - B\left(\pi_{t_0}, \mu_{t_0} + \beta_0\right)$$

$$= \sum_j \pi_{j,t_0}\left(\left(\mu_{j,t_0} + \beta_{j,t} - \sum_j \pi_{j,t_0}\left(\mu_{j,t_0} + \beta_{j,t}\right)\right)^2 - \left(\mu_{j,t_0} + \beta_{j,t_0 0} - \sum_j \pi_{j,t_0}\left(\mu_{j,t_0} + \beta_{j,t_0}\right)\right)^2\right)$$

$$\delta_W^{D,T} = W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_t\right) - W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_{t_0}\right)$$

$$= \sum_j \pi_{j,t_0}\left(\left(\sigma_{j,t_0} + \lambda_{j,t}\right)^2 - \left(\sigma_{j,t_0} + \lambda_{j,t_0}\right)^2\right)$$

$$\delta_C^{D,T} = \left(B\left(\pi_t, \mu_{t_0} + \beta_t\right) - B\left(\pi_{t_0}, \mu_{t_0} + \beta_t\right)\right) - \left(B(\pi_t, \mu_t) - B\left(\pi_{t_0}, \mu_t\right)\right)$$

$$+ \left(W\left(\pi_t, \sigma_{t_0} + \lambda_t\right) - W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_t\right)\right) - \left(W(\pi_t, \sigma_t) - W\left(\pi_{t_0}, \sigma_t\right)\right)$$

$$\approx \sum_j\left(\pi_{j,t} - \pi_{j,t_0}\right)\left(\left(\mu_{j,t_0} + \beta_{j,t} - \sum_j \pi_{j,t}\left(\mu_{j,t_0} + \beta_{j,t}\right)\right)^2 - \left(\mu_{j,t} - \sum_j \pi_{j,t}\mu_{j,t}\right)^2 + \left(\sigma_{j,t_0} + \lambda_{j,t}\right)^2 - \sigma_{j,t}^2\right)$$

if $\sum_j \pi_{j,t_0}\mu_{j,t} \approx \sum_j \pi_{j,t}\mu_{j,t}$. The exact equation is given in appendix A2.3.

$$\delta_P^{D,T} = B(\pi_t, \mu_t + \beta_t) - B\left(\pi_t, \mu_{t_0} + \beta_t\right) + W(\pi_t, \sigma_t + \lambda_t) - W\left(\pi_t, \sigma_{t_0} + \lambda_t\right)$$

$$- \left(B\left(\pi_{t_0}, \mu_t\right) - B\left(\pi_{t_0}, \mu_{t_0}\right) + W\left(\pi_{t_0}, \sigma_t\right) - W\left(\pi_{t_0}, \sigma_{t_0}\right)\right)$$

$$= \sum_j \pi_{j,t}\left(\left(\mu_{j,t} + \beta_{j,t} - \sum_j \pi_{j,t}(\mu_{j,t} + \beta_{j,t})\right)^2 - \left(\mu_{j,t_0} + \beta_{j,t} - \sum_j \pi_{j,t}(\mu_{j,t_0} + \beta_{j,t})\right)^2 + \left(\sigma_{j,t} + \lambda_{j,t}\right)^2 - \left(\sigma_{j,t_0} + \lambda_{j,t}\right)^2\right) - \sum_j \pi_{j,t_0}\left(\left(\mu_{j,t} - \sum_j \pi_{j,t_0}\mu_{j,t}\right)^2 - \left(\mu_{j,t_0} - \sum_j \pi_{j,t_0}\mu_{j,t_0}\right)^2 + \sigma_{j,t}^2 - \sigma_{j,t_0}^2\right)$$

[7] Rather than decomposing the change in the effect of treatment on the variance, the change in the post-treatment variance induced by the change in the effect of treatment, i.e., $V[Y_t(1)|D_t = 1] - V\left[Y_{t_0}(1)|D_{t_0} = 1\right]$, can also be decomposed. This decomposition is presented in appendix A2.4. In equation (8), $W(\pi, \sigma^2) = \sum_j \pi_j \sigma_j^2$ and $B(\pi, \mu) = \sum_j \pi_j\left(\mu_j - \sum_j \pi_j \mu_j\right)^2$ are the within- and between-group equation functions of the V that take the parameter vectors $\pi = \pi_1, \dots, \pi_J$, $\sigma^2 = \sigma_1^2, \dots, \sigma_J^2$, and $\mu = \mu_1, \dots, \mu_J$ as input.

The between-group effect captures the change in the effect of treatment on the variance induced by a change in the effect of treatment on the group means. The within-group effect captures the change in the effect of treatment on the variance induced by a change in the effect of treatment on the group variances. These two effects are analogous to the "coefficient effect" of a KBO decomposition across time (Kröger and Hartmann 2021). A KBO decomposition, however, does not calculate the within-group effect as it only considers mean differences. The compositional effect represents the change in the effect of treatment on the variance induced by a change in the distribution of treatment across groups. This effect is analogous to the "endowment effect" of a KBO decomposition across time with a binary covariate. A KBO decomposition, however, ignores the endowment effect on the within-group variance. Finally, the pre-treatment effect captures the change in the effect of treatment on the variance due to a change in pre-treatment inequality (i.e., μ_{j,t_0} and σ_{j,t_0})[8]. The superscript D,T on the δs indicates that the change induced by the change in treatment over time is considered. The derivation of equation (8) can be found in appendix A2.3.

Estimating the effect of treatment on the group-specific means and variances

Having established the aggregate consequences of treatment, next I turn to how the treatment effect on the group-specific means and variances can be estimated in an empirical model. To simplify the exposition, I discuss treatment effect estimation at a single timepoint. The same approach, however, can be used with longitudinal data by estimating treatment effects for each timepoint separately.

In Western and Bloome's (2009) descriptive variance decomposition, the relevant estimand are the group-specific means (μ_j) and variances (σ_j^2) as the interest lies on the effect of belonging to a group. They use variance regression to model the group-specific means and variances (Harvey 1976). Despite the name, the standard deviation σ_j rather than the variance σ_j^2 is typically modeled with variance regression:

$$\mu_j = E(Y|G) = \alpha G \tag{9}$$
$$\log(\sigma_j) = \log(SD(Y|G)) = \delta G$$

[8] The pre-treatment effect does not exist in a descriptive variance decomposition because it does not differentiate between pre- and post-treatment means and variances.

In equation (9), G_t is a matrix of dummy variables representing the groups and the intercepts are omitted so that the coefficients can directly be taken as mean and (log) standard deviation of each group. In the explanatory variance decomposition, by contrast, the relevant estimand is the treatment effect on the group-specific means and variances as the interest lies in the effect of a treatment administered to a group[9]. Accordingly, the treatment effects on the group-specific means and standard deviations are modeled. While any suitable estimator can be used to estimate these effects, the R library accompanying this paper facilitates using the simple difference and the difference-in-difference estimator.

Simple difference estimator

To identify the treatment effect on the group-specific means and variances, the grouping variable G and the treatment variable D must be separated and interacted:

$$\begin{aligned} \mu_j &= E(Y|G) = \alpha G + \beta GD + \gamma Z \\ \log(\sigma_j) &= \log(SD(Y|G)) = \delta G + \lambda^* GD + \rho Z \end{aligned} \tag{10}$$

where α and δ are the group-specific pre-treatment means and standard deviations, β and λ are the group-specific treatment effects on the mean and standard deviation, and γ and ρ are effects of control variables. Since in equation (10) the treatment effect on the standard deviation is modeled multiplicatively, λ_j equals $\exp(\delta_j G + \lambda_j^* GD + \rho Z) - \exp(\delta_j G + \rho Z)$.

To interpret the treatment effects estimated in equation (10) as causal effects, the *conditional independence assumption* must hold. Control variables Z can be included in the model to render this assumption more plausible. The conditional independence assumption states that, conditional on some control variables Z, the potential outcomes under treatment $Y(1)$ and control $Y(0)$ are independent of the treatment status D itself, that is, $Y(1), Y(0) \perp D|G, Z$. In other words, treatment is random given the controls. For the mean, the causal effects are defined as the average difference in potential outcomes in each group.

[9] Note that groups are defined exogenously. Treatment effects will thus be biased if group membership is correlated with unobserved confounders. Treatment effects are, however, unaffected by group-specific selections into treatment or treatment effects since they are modeled.

That is, $\underbrace{E[Y(1) - Y(0)|D = 1, G, Z]}_{\text{ATT}} = \underbrace{E[Y|D = 1, G, Z] - E[Y|D = 0, G, Z]}_{\text{quantity estimated in eq. (10)}}$. For the standard deviation, the causal effects are defined as the average difference in the standard deviation of the potential outcomes in each group: $\underbrace{E\big[SD[Y(1)|D = 1, G, Z] - SD[Y(0)|D = 1, G, Z]\big]}_{\sqrt{\text{VTT}}} = \underbrace{E\big[SD[Y|D = 1, G, Z]\big] - E\big[SD[Y|D = 0, G, Z]\big]}_{\text{quantity estimated in eq. (10)}}$.

Difference-in-difference estimator

If pre- and post-treatment observations are available, a difference-in-difference (DID) estimator can be specified to identify the treatment effects on the mean and variance in each group:

$$\mu_j = E(Y|G) = \alpha G + \varphi GP + \beta GDP + \gamma Z \tag{11}$$
$$\log(\sigma_j) = \log\big(SD(Y|G)\big) = \delta G + \psi GP + \lambda^* GDP + \rho Z$$

where $P \in \{0,1\}$ indicates pre/post-treatment status.

where α and δ are group-specific pre-treatment means and standard deviations, φ and ψ are group-specific post-treatment means and standard deviations, β and λ^* are group-specific treatment effects on the mean and standard deviation, and γ and ρ are effects of control variables. Again, since in equation (11) the treatment effect on the standard deviation is modeled multiplicatively, λ_j equals $\exp\big(\psi_j GP + \lambda_j^* GDP + \rho Z\big) - \exp\big(\psi_j GP + \rho Z\big)$.

The DID estimand is the average difference of potential outcomes post treatment of those who receive treatment: $E[Y(1) - Y(0)|D = 1, P = 1, G]$. The DID estimator captures this effect if the trend from pre- to post-treatment in the absence of treatment, i.e., $E[Y(0)|P = 1, G] - E[Y(0)|P = 0, G]$, is the same for those who receive treatment and those who do not (*common trends assumption*). If this common trends assumptions holds, the treatment effects estimated in equation (11) are causal effects. For the mean, the causal effects are defined as the average difference in potential outcomes post treatment in each group:

$$\underbrace{E[Y(1) - Y(0)|D = 1, P = 1, G]}_{\text{ATT}} =$$
$$\underbrace{\big(E[Y|D = 1, P = 1, G] - E[Y|D = 0, P = 1, G]\big) - \big(E[Y|D = 1, P = 0, G] - E[Y|D = 0, P = 0, G]\big)}_{\text{quantity estimated in eq. (11)}}.$$

For the standard deviation, the causal effects are defined as the average difference in the standard deviation of the potential outcomes in each group: $\underbrace{\mathrm{E}\big[\mathrm{SD}[\mathrm{Y}(1)|\mathrm{D}=1,\mathrm{P}=1,\mathrm{G}]-\mathrm{SD}[\mathrm{Y}(0)|\mathrm{D}=1,\mathrm{P}=1,\mathrm{G}]\big]}_{\sqrt{\mathrm{VTT}}}=$

$$\underbrace{\Big(\mathrm{E}\big[\mathrm{SD}[\mathrm{Y}|\mathrm{D}=1,\mathrm{P}=1,\mathrm{G}]\big]-\mathrm{E}\big[\mathrm{SD}[\mathrm{Y}|\mathrm{D}=0,\mathrm{P}=1,\mathrm{G}]\big]\Big)-\Big(\mathrm{E}\big[\mathrm{SD}[\mathrm{Y}|\mathrm{D}=1,\mathrm{P}=0,\mathrm{G}]\big]-\mathrm{E}\big[\mathrm{SD}[\mathrm{Y}|\mathrm{D}=0,\mathrm{P}=0,\mathrm{G}]\big]\Big)}_{\text{quantity estimated in eq. (11)}}$$

In conclusion, while the simple difference estimator in equation (10) relies on the plausibility of the conditional independence assumption, the difference-in-difference estimator in equation (11) depends on the plausibility of the common trends assumption. If it holds, the DID estimator removes biases in post-treatment comparisons between treated and control that are the result of permanent differences between those groups, as well as biases from comparisons over time in the treated group that are the result of trends due to other causes of the outcome.

VARIANCE-BASED INEQUALITY MEASURES

Three of the most commonly used inequality measures are based on the variance and thus can be decomposed with the approach presented in the previous section: the variance $\mathrm{V}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}}(\mathrm{y_i}-\bar{\mathrm{y}})^2$, the variance of the logarithms $\mathrm{V_L}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i}}\left(\ln \mathrm{y_i}-\frac{1}{\mathrm{n}}\sum \ln \mathrm{y_i}\right)^2$, and the squared coefficient of variation $\mathrm{CV}^2=\mathrm{V}/\bar{\mathrm{y}}^2$, which divides the variance by the squared mean and is a member of the generalized entropy family. In this section, I discuss important properties of these three inequality measures to show that researchers should not incautiously use $\mathrm{V_L}$ in the context of variance decompositions as taking the log can bias the decomposition.

The literature on inequality measurement has developed a list of inequality indices as well as an axiomatic framework to assess those indices with respect to desirable properties of inequality measures. The five inequality axioms are (Allison 1978; Cowell 2011):

1. *Pigou-Dalton principle of transfers*. An inequality measure should decrease after any transfer of income from a higher to lower income earner that does not change the ranking of incomes (and vice versa).

2. *Additive decomposability.* If a population can be partitioned into mutually exclusive and exhaustive subgroups, an inequality measure should be additively decomposable into within- and between-group inequalities.
3. *Scale independence.* An inequality measure should be insensitive to proportional changes in income, including changes in the units of measurement.
4. *Principle of population.* A measure should be invariant to proportional changes in the size of the underlying population.
5. *Anonymity.* Any permutation of incomes should leave the measure unchanged.

In following, I discuss axioms 1 through 3 as 4 and 5 are met by all variance-based inequality measures. The evaluation is summarized in Table 4.

Pigou-Dalton principle of transfers

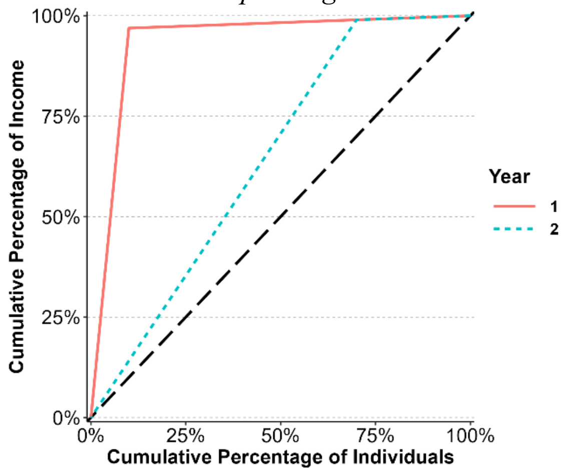
The principle of transfers is the most important of the five axioms because indices violating it move in wrong directions after certain transfers. That is, they may increase after inequality-decreasing transfers (i.e., transfer of income from a higher to lower income earner that does not change the ranking of incomes) and vice versa. The V, CV^2, and Gini coefficient respect the transfer principle. The $\mathrm{V_L}$, however, violates the principle for transfers above 2.7 times the geometric mean (Allison 1978; Cowell 1988, 2011; Foster and Ok 1999; Wolfson 1994). Take the following income distribution as an example:

Table 1: Two hypothetical income distributions

Year	Income distribution	Arithmetic mean	Geometric mean
1	$y_1 = (1,1,1,1,1,1,1,1,1,281)$	$\bar{y}_1 = 29$	$\bar{y}_1 = 1.8$
2	$y_2 = (1,1,1,41,41,41,41,41,41,41)$	$\bar{y}_2 = 29$	$\bar{y}_2 = 13.5$

y_2 can be obtained from y_1 by six progressive transfers (i.e., from rich to poor). Accordingly, the corresponding Lorenz curves in Figure 1 display a decrease of inequality from year 1 to 2. Table 2 shows that this decrease in inequality is captured by V, CV^2, and the Gini coefficient. The $\mathrm{V_L}$, however, *increases*. This behavior has led Wolfson (1994) to call for the $\mathrm{V_L}$ to be banned from inequality analysis.

Figure 1: Lorenz curves corresponding to the distributions in Table 1



The long-dashed line represents equality and deviations from it represent increases in inequality.

Table 2: Inequality statistics corresponding to the distributions in Table 1.

Year	*V*	V_L	CV^2	*Gini*
1	*7840*	*3.18*	*9.32*	*0.87*
2	*373*	*3.22*	*0.44*	*0.29*

All inequality statistics record a decrease of inequality except for the V_L.

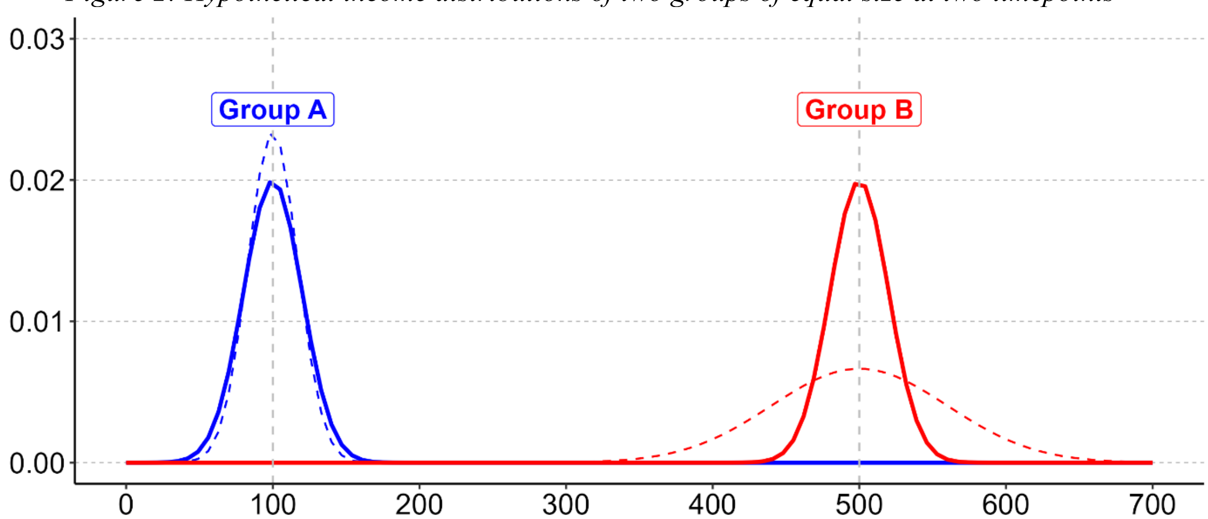
Additive decomposability

While the issue that the V_L violates the transfer principle is not novel, its consequences for the variance decomposition framework have not received careful attention in the literature. The variance decomposition framework is centered around the arithmetic mean (i.e., $\bar{y} = \frac{1}{n}\sum y$) since both within-group variance ($W = \sum\sum(y_{ij} - \bar{y}_j)^2$) and between-group variance ($B = \sum(\bar{y}_j - \bar{y})^2$) include it. The V_L, however, depends on the on the geometric mean (i.e., $\bar{y}^* = \sqrt[n]{y_1 y_2 \dots y_n}$) because $V_L = \frac{1}{n}\sum_i\left(\ln y_i - \frac{1}{n}\sum \ln y\right)^2$ and $\frac{1}{n}\sum \ln y = (\prod y)^{1/n} = \sqrt[n]{y_1 y_2 \dots y_n}$. In expectation, the geometric mean converges to $E\left[(y_1 y_2 \dots y_n)^{1/n}\right] = \exp(E[\log(y)])$ while the arithmetic mean converges to $E[y] = E\left[\exp(\log(y))\right]$. These quantities are not the same. In fact, $E\left[(y_1 y_2 \dots y_n)^{1/n}\right] \leq E[y]$. Because of this, some spreads leave the arithmetic mean unchanged but decrease the geometric mean and vice versa—as is the case in the example of Table 1. The V_L, therefore, is *not* additively (i.e., linearly) decomposable as within-and between-group variances cannot be properly separated (Cowell 1988, 2011; Foster and Ok 1999).

Another concern about the V_L in the context of additive decomposability is that the effect of transfers depends on the income level at which the transfers occur. The V_L is more sensitive to transfers that occur at lower income levels whereas V and CV^2 are equally sensitive to transfers at all income levels. This feature may be desirable, for instance, to represent diminishing marginal utility of income. However, from a measurement perspective, it seems incoherent to linearly decompose an inequality measure with a nonlinear distance concept.

Figure 2 and Table 3 provide an example to illustrate these two issues. The figure displays income distributions of two groups of equal size at two timepoints. The two groups have different means that are constant over time and the same standard deviation (SD) of 20 at timepoint 1. At timepoint 2, the SD of group A decreases to 17 and the SD of group B increases to 60. The table reports the results of a descriptive decomposition (as a percentage of the inequality at timepoint 1). While V and CV^2 record an increase of within-group inequality, V_L records a decrease. This is because the V_L first nonlinearly weighs the changes within the groups due to the log-transformation but then linearly averages them when computing the change in within-group variance. Further, while V and CV^2 correctly record no change in between-group inequality, V_L suggests a decrease owing to its dependence on the geometric mean.

Figure 2: Hypothetical income distributions of two groups of equal size at two timepoints



Solid lines represent the income distribution at timepoint 1, dashed lines represent the income distributions at timepoint 2. The means and standard deviations of the four distributions are as follows: $\mu_{A,1} = \mu_{A,2} = 100$, $\mu_{B,1} = \mu_{B,2} = 500$, $\sigma_{A,1} = \sigma_{B,1} = 20$, $\sigma_{A,2} = 17$, $\sigma_{B,2} = 60$.

Table 3: Results of a descriptive decomposition (as a percentage of the inequality at timepoint 1)

	CV^2		V		V_L	
	Within	Between	Within	Between	Within	Between
Time 1	100	100	100	100	100	100
Time 2	302	100	302	100	**95.4**	**98.8**

A way to restore the V_L as a useful inequality statistic is to interpret $y^* = \ln y$ as the latent utility of income. On this latent scale, the V_L respects the transfer principle and is decomposable because it reduces to the V. Further, the distributional properties of y^* can improve parameter estimation and provide robustness to extreme values[10]. Researchers, however, should not incautiously equate the variance in the utility of income with the variance in actual income. There is no simple correspondence between the two. The variance in utility may increase while the variance in income decreases and the decomposition of utility may indicate a change in within- or between-group inequality while the decomposition of income does not. Further, researchers should understand that, by choosing the latent scale interpretation, they switch from a relative inequality measure (V_L) to an absolute one (V).

Since interpreting $\ln y$ as the utility of income also comes with rather specific assumptions about human nature, I recommend sticking to the original (dollar) scale and using the V (absolute inequality) or the CV^2 (relative inequality) as appropriate because both indices respect the transfer principle and are additively decomposable. I derive the descriptive and explanatory variance decomposition of the V in section 2 and 3 and the descriptive and explanatory variance decomposition of the CV^2 in Appendix 2.

Previous applications of the descriptive variance decomposition all decompose the V_L (e.g., Juhn et al. 1993; Lemieux 2006; Weeden et al. 2007; Western and Bloome 2009; Western et al. 2008; Wodtke 2016;

[10] Note, however, that the V_L is not more robust to extreme values than the CV^2 (Cowell and Victoria-Feser 1996). It should be investigated if other means to improving parameter estimation, such as $y^* = \sqrt{y}$, also lead to a violation of the transfer principle.

Xie, Killewald, and Near 2016)[11]. This is unproblematic if the results are interpreted on a latent scale. I exemplarily replicated the analysis of Wodtke (2016) in Appendix 4 and find that the overall interpretation of the results does not change when I replace the V_L with the CV^2. In the application below, however, I demonstrate that results can deviate substantially. I, therefore, recommend making a deliberate decision about which inequality measure to use and how to interpret it.

Scale independence

An often-cited issue of the V as inequality measure is its scale dependence as an absolute measure of inequality. The variance quadruples when everyone's income doubles even though the relative distance among individuals is unchanged. V_L and CV^2, by contrast, are scale-independent relative inequality measures[12], making it unnecessary to convert currencies when comparing countries or to adjust for inflation when assessing change over time. However, I argue that the decision of whether to use an absolute or a relative inequality measure should be based on a theory of social welfare rather than computational convenience since these operations are simple to perform nowadays. Most inequality research measures inequality relatively because it is assumed that the marginal utility of income declines as income increases. While this argument has an individual-psychological foundation (Kahneman and Tversky 1979), from societal-sociological perspective, it is less convincing to consider inequality unchanged by any proportionate increase in everyone's income. This is because, in absolute amounts, the rich benefit more than the poor from proportionate increases, and many goods and services in society are disproportionally valuable, making it essential to be able to purchase them (e.g., the utility of an ivy league degree compared to a community college degree). Consequently, researchers should carefully consider whether absolute inequality or relative inequality is the quantity of interest in their application when deciding which

[11] The exceptions are Breen and Chung (2015), who decompose the CV^2, and Liao (2019), who decomposes the Gini coefficient and the Theil index.

[12] V_L is scale-independent because $\mathrm{Var}(\ln(2y)) = \mathrm{Var}(\ln(2) + \ln(y)) = \mathrm{Var}(\ln(y))$, and the CV^2 is scale-independent because the variance and the squared mean are on the same scale and their ratio is thus dimensionless.

inequality measure to decompose as the scale dependence of the V is not an issue if all values are brought onto one scale.

Table 4: Properties of variance-based inequality measures

Inequality measure	Scale independence	Transfer principle	Decomposability	Population principle	Anonymity	Sensitivity
V	no	yes	yes	yes	yes	equal
$\mathrm{V_L}$	yes	no	no	yes	yes	lower
CV^2	yes	yes	yes	yes	yes	equal

EMPIRICAL EXAMPLE: THE CHANGING IMPACT OF MOTHERHOOD ON EARNINGS INEQUALITY

In what follows, I apply the variance decomposition approach developed in section 3 to examine the changing impact of motherhood on women's earnings and its consequences for women's earnings inequality between 1980 and 2020. I use this example to demonstrate the utility of the proposed approach for understanding how individual-level change contributes to aggregate inequality trends. I divide the research question into two applications. Application 1 is a descriptive variance decomposition that breaks inequality trends into within- and between-group components to discuss the consequences of using the $\mathrm{V_L}$ in context of a variance decomposition. Application 2 then showcases the explanatory variance decomposition by examining the degree to which these trends are caused by the changing impact of motherhood on women's earnings. In both applications, I group women by household economic position to analyze the degree to which inequality trends are differentiating women within or between economic strata.

Data and measurements

Current Population Survey Panels. My data source is the 1980-2020 Current Population Surveys (CPS), which have been extensively used to study changes in inequality in the US. I restrict my sample to adult, partnered women of childbearing age (18-49) and weigh observations by their inverse sample inclusion probability.

Women's earnings. I use the Outgoing Rotations Group supplement of the CPS collected in months-in-sample 4 and 8 to measure women's weekly pre-tax wage and salary income adjusted for inflation to 2010 dollars using the consumer price index (CPI-U) (Crawford, Church, and Akin 2014). I exclude earnings from the self-employed but include zero-earners and those not working for pay.

Motherhood. Owing to the sampling design of the CPS, about 70 percent of respondents can be linked across subsequent months of survey participation (Drew, Flood, and Warren 2014). I use these newly available identifiers to create short-term panels, tracking earnings across two timepoints (months-in-sample 4 and 8) that are 12 months apart. Following Musick and Jeong (2021), I measure the change in earnings of women who gave birth in between the two timepoints. This approach thus identifies the immediate impact of motherhood on earnings. To also capture longer-term effects, I additionally measure the change in earnings of young mothers with an eldest child up to the age of 5. By modeling motherhood as a categorical variable, I can evaluate both the effect of transitioning to motherhood and the effect of having a young child *compared to not having children*. I then calculate the average within-person change in earnings following motherhood up to child age 5 by averaging across these effects (1-year-old vs. no child, 2-year-old vs. no child, etc.). Therefore, the estimated motherhood effect captures both immediate and longer-term effects of motherhood on earnings. For different-sex couples, I take her information on own children in the household, and, for same-sex couples households, I take the household head's information on own children in the household. I restrict my sample to partnered women because the single women in the sample exhibit a high panel attrition that may be endogenous to the motherhood effect.

Household economic position. I measure women's economic position prior to childbirth by the total annual household income. This variable has been collected as part of the basic monthly survey since 1982 in categorical form and includes earnings and nonlabor income of all household members. I take the middle values of each category, adjust them for inflation to 2010 dollars, and then compute the 20 and 80 percentiles to categorize women into low, medium, and high economic position.

Time. I pool the data in 5-year intervals to reduce noise in the estimates.

For the explanatory variance decomposition, I use the following control variables (measured prior to childbirth): age, education, race, marriage status, and family size.

The final sample includes 498,388 adult women of the childbearing age (18-49) between 1982 and 2020 for which two observations are available (84 percent link rate) and with no missing values on all relevant variables. 69,922 of those women (~1793 per year) either become mothers at the second timepoint or already are mothers at the first timepoint with an eldest child up to age 5 at the second timepoint.

Application 1: Descriptive decomposition of women's earnings inequality

Figure 3 depicts the mean and standard deviation of women's earnings by economic strata and over time estimated using variance regression (equation 9). We can see that both mean and standard deviation have been growing steadily across economic strata in the past four decades. One might be inclined to take the steeper slopes for women in high-income households as evidence for a steady rise of inequality. Figure 4 shows that this is true in absolute terms (as measured by V). Relative inequality, however, declined substantially between 1980 and 2000 and has since remained constant (as measured by Gini or CV^2) or increased again (as measured by $\mathrm{V_L}$).

In the following, I examine these trends using a descriptive variance decomposition (equation 2). I decompose the CV^2 rather than the V to compare the results to a decomposition of the $\mathrm{V_L}$. Figure 5 presents the results of this decomposition graphically, splitting the overall change in the CV^2 into changes in within-group inequality, between-group inequality, and changes in the size of economic strata[13]. For the CV^2, the figure shows that inequality between economic strata increased in the early 1980s but has since remained constant and, thus, cannot explain the decline in total inequality between 1980 and 2000. Changes in the size of economic strata do not account for much of the decline in total inequality either. Instead, the decomposition reveals that the decline owes to a decrease of inequality within economic strata during this period[14]. This result demonstrates that researchers run the risk of ignoring significant changes in inequality by solely focusing on mean differences across groups (i.e., between-group inequality).

[13] Table A3.1 in Appendix 3 provides the results in table form.

[14] Figure A3.1 in the appendix confirms that the group-specific CV_j^2 (i.e., σ_j^2/μ_j^2) declined during this time for all three groups.

The decomposition of the V_L, in contrast, suggests that both within- and between-group inequality have been increasing since 2000. It difficult to discern the extent to which these contradictory results are a consequence of the V_L violating of the transfer principle or the V_L using a nonlinear distance concept. Regardless the reason, the stark differences between CV^2 and V_L highlight that researchers should be cautious about logging income in variance decompositions.

Application 2: Explanatory decomposition of the motherhood effect

There is an extensive body of literature on the motherhood penalty (Budig and England 2001; Cukrowska-Torzewska and Matysiak 2020). By contrast, we know much less about how the motherhood effect shapes women's earnings in ways that contribute to aggregate inequality (but see Gonalons-Pons, Schwartz, and Musick 2021; Harkness 2013; Kleven, Landais, and Søgaard 2019).

Therefore, as a next step, I examine how the changing effect of motherhood on women's earnings has impacted women's earnings inequality using a explanatory variance decomposition. In keeping with prior work in this literature, I focus on relative inequality and therefore decompose the CV^2 rather than the V. The decomposition breaks the motherhood effect into four components—a within-group effect, a between-group effect, a compositional effect, and a pre-treatment effect. The within- and between-group components describe how much changes in the motherhood effect (i.e., changes in employment, hours, and wages) have affected the inequality of women within and between economic strata. The compositional component describes how much changes in the share of mothers have affected inequality of women within and between economic strata. Finally, the pre-treatment component describes how much changes in the inequality prior to motherhood have affected inequality of women within and between economic strata. This decomposition, therefore, disentangles the distinct behavioral and compositional mechanisms underlying the total impact of motherhood on inequality.

Figure 3: Weekly earnings of women of childbearing age

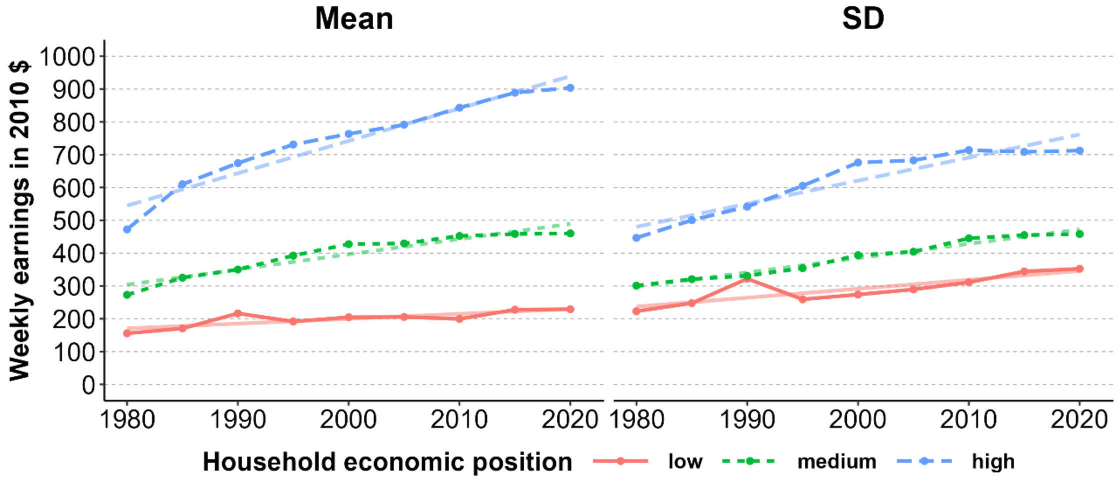


Figure 4: Absolute and relative earnings inequality of women of childbearing age

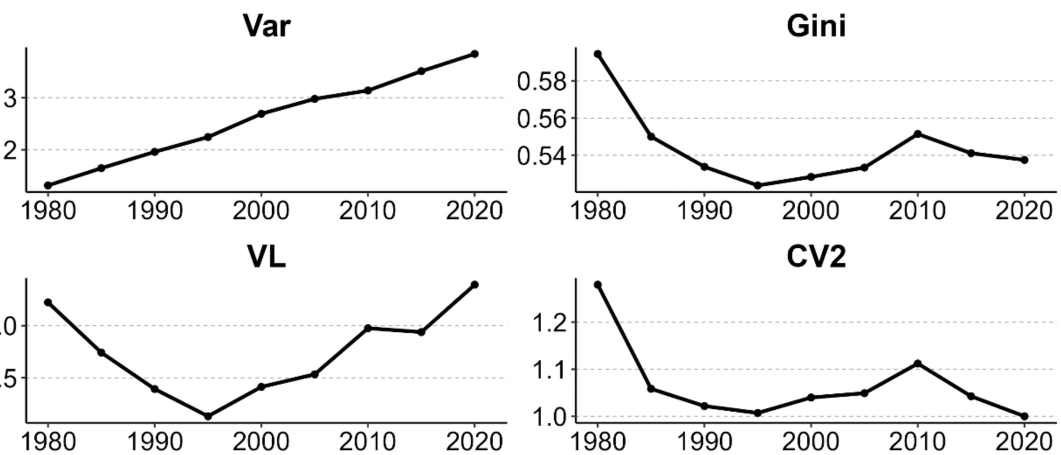


Figure 5: Descriptive decomposition of the change in the CV^2 and the V_L since 1982 (in points)

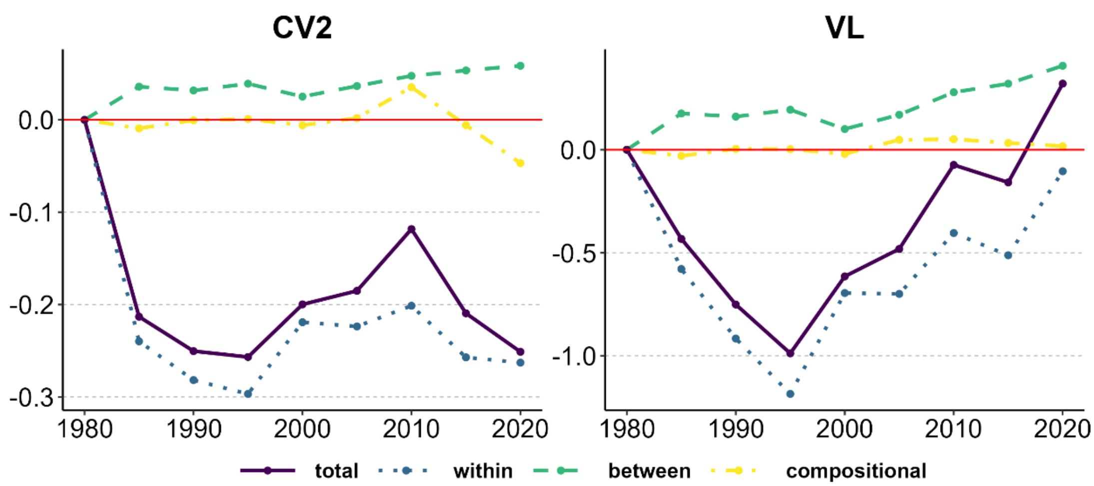


Figure 6: Motherhood effect on the mean (β) and standard deviation (λ) of earnings

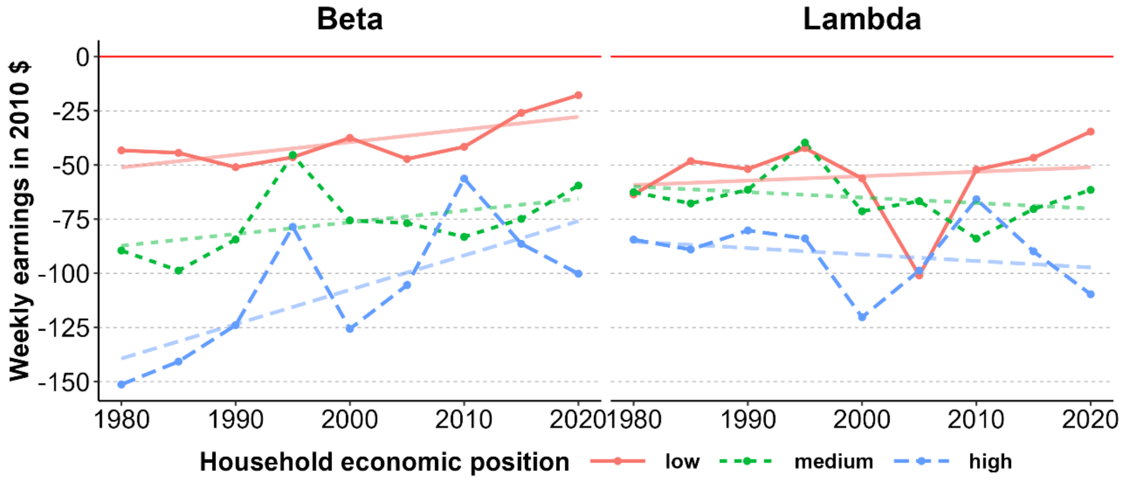


Figure 7: Decomposition of the motherhood effect on the CV^2

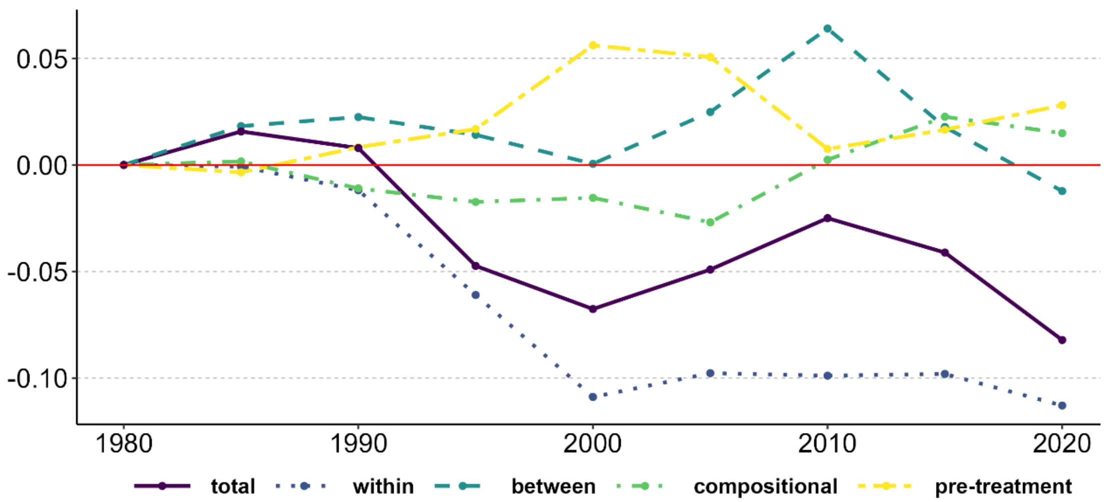
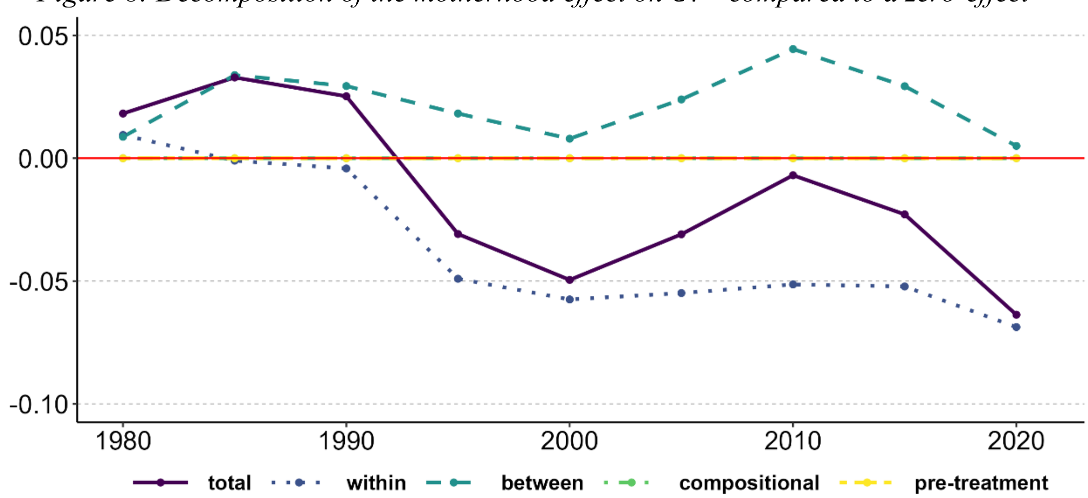


Figure 8: Decomposition of the motherhood effect on CV^2 compared to a zero-effect



I use a difference-in-difference estimator and demographic controls to estimate the within-person effect on the mean and variance of women's earnings of having a child up to age 5 in the household, compared to not having children. The difference-in-difference estimator removes all time-constant differences between nonmothers and mothers as well as any time trends affecting both nonmothers and mothers. The demographic controls (age, education, race, marriage status, and family size) are included to ensure that observations are conditionally independent should the common trends assumption not hold.

This research design has several limitations. First, the plausibility of the common trends assumption cannot be examined as only one pre-treatment observation is available. Second, the effect of motherhood on women's earnings may be underestimated if expectant mothers reduce their work commitments prior to giving birth as earnings recorded before childbirth would be lower than usual. This underestimation will affect the impact estimates of motherhood on inequality within and between groups if women reduce their work commitments differently across economic strata.

Figure 6 displays the estimates of the motherhood effect (in dollars per week) on the mean and standard deviation of women's earnings. The figure shows that the motherhood effect on mean earnings is negative, and that the earnings losses following motherhood are lowest for women in low-income households and highest for women in high-income households[15]. In line with previous research (Glauber 2018), we can see that the earnings losses have been declining since the 1980s—most significantly for women in high-income households. The motherhood effect on the standard deviation of earnings is likewise negative. That is, women within economic strata differ in earnings more before than after motherhood. This convergence of earnings following motherhood since the 1980s has been decreasing for women in low-income households, increasing for women in high-income households, and relatively stable for women in medium-income households.

[15] I estimate the absolute effect in dollars rather than the relative effect in percent of pre-birth earnings because the dollar amount is what determines the impact on aggregate inequality. Further, note that the motherhood effect is strongest following childbirth and attenuates as children age. I average the effect across child age to describe the motherhood effect for a larger population than just new mothers.

In the following, I employ the explanatory variance decomposition to disentangle the consequences of these changes for aggregate inequality. Figure 7 presents the results of the decomposition, showing four effects—within, between, compositional, pre-treatment—and their total[16].

The within- and between-group effects capture the distributional consequences of the behavioral changes shown in Figure 6. The within-group effect is negative, indicating that motherhood reduces within-group inequality. This inequality-reducing effect, further, increased substantially between 1980 and 2000 and has since remained constant. The between-group effect, by contrast, increased inequality up until 2010, but has since been declining again. Indeed, the 2020 estimate suggests that motherhood no longer increases earnings inequality between women in low-, medium-, and high-income households.

The compositional effect captures the distributional consequences of changes in the share of mothers across economic strata. The effect—overall small—decreased inequality up until 2010 and has since been increasing inequality. Figure A3.2 in Appendix 3, which displays the share of mothers by economic strata over time, shows that motherhood up until 2010 was constant in high-income households but declined slightly in low- and medium-income households. Thus, fewer women in low- and medium-income households incurred income losses due to motherhood. Since around the 2008 financial crisis, however, motherhood in high-income households has started to decline as well, which reversed the effect.

Finally, the pre-treatment effect captures the distributional consequences of changes in the impact of motherhood that are driven by changes in inequality prior to motherhood. This effect has increased inequality since the 1980s. The reason is that pre-childbirth inequality has declined (as shown in Figure 5), which lowered the inequality-reducing effect of motherhood.

Adding up the four components, I find that—overall—the changing effect of motherhood on earnings since the 1980s has reduced women's earnings inequality. In fact, the CV^2 is about 1.6 percent lower in 2020 than in 1980 due the changing motherhood effect[17]. The primary reason for it is that the inequality-

[16] Table A3.1 in Appendix 3 provides results in table form.
[17] Details on the calculations can be found in Appendix 3.

decreasing effect of motherhood on within-group inequality is about 11.5 percent higher in 2020 than in 1980 while the inequality-increasing effect on between-group inequality has not changed much.

Instead of comparing the motherhood effect in each year to its effect in the 1980s, we can also compare the motherhood effect in each year to a zero-effect. That is, a counterfactual scenario in which motherhood has no effect on the mean and variance of women's earnings. While this counterfactual scenario is unrealistic, it is a useful benchmark to understand the absolute effect of motherhood on inequality. Figure 8 displays the results of this decomposition[18]. The figure shows that, between 1980 and 1990, motherhood increased total inequality by increasing inequality between economic strata without affecting inequality within them. Between 1990 and 2000, however, motherhood begun to homogenize earnings within economic strata, which caused motherhood to reverse its effect on total inequality. In fact, the CV^2 is about 5.5 percent lower in 2020 than if motherhood had no effect on earnings.

CONCLUSION AND DISCUSSION

In this paper, I introduce a novel approach that enables researchers to measure treatment effects on variance-based inequality statistics and decompose the effects into within- and between-group components. The approach is based on the descriptive variance decomposition (e.g., Western and Bloome 2009), which it extends to an explanatory framework. It can be employed in observational research studying predictor effects and in experimental research studying treatment effects. The approach is more transparent and flexible than related approaches, such as Lemieux (2002), because variance regression (Harvey 1976) rather than a convoluted re-weighting procedure is employed to quantify the treatment effect on the within-group (residual) variance. The approach permits researchers to analyze inequality at a single timepoint using cross-sectional data as well as inequality trends over time using repeated cross-sectional or panel data. As treatment effects are defined at the individual level, the approach links individual-level effects to their

[18] The decomposition returns only a within- and between-group effect, as the level of pre-treatment inequality and the number of mothers in each group are kept at their actual levels.

consequences for group-level and total inequality (micro-to-macro link). As such, the approach bears the potential to not only advance our understanding of *why* but also of exactly *how* inequality is changing[19].

I apply the approach to examine the decline in women's earnings inequality between 1980 and 2020. I demonstrate that changes in the effect of motherhood on earnings is one reason why inequality declined during this period. The CV^2 is about 1.6 percent lower in 2020 due to changes in the motherhood effect since 1980 and about 5.5 percent lower in 2020 than if motherhood had no effect on earnings. The decomposition of this effect reveals that motherhood increases inequality between economic strata but reduces inequality within them. Moreover, the within-group effect increased between 1990 and 2000 and is substantially larger than the between-group effect. This result highlights the risk of drawing misleading conclusions when inequality researchers base their analyses solely on mean differences between groups.

I further highlight that careful attention must be paid to the basic axioms of inequality measurement when decomposing the variance into within- and between-group components. Previous applications of the variance decomposition approach decompose the variance of log-income rather than the variance of income. The variance of logarithms, however, is *not* additively decomposable into within- and between group components because it does not respect the Pigou-Dalton principle of transfers (Cowell 1988, 2011; Foster and Ok 1999). I demonstrate that this issue can be addressed—either by switching to a latent utility interpretation or by using the squared coefficient of variation instead of the variance of logarithms.

With this paper, I provide the R library "ineqx" to facilitate research using the variance decomposition approach. The library implements both descriptive and explanatory variance decomposition. The descriptive variance decomposition can be applied to a wide range of questions central to understanding dimensions of inequality, such as "Is there more inequality within races, genders, or economic strata or more between them?" and "Is the inequality within and between these groups changing over time?". The

[19] Future research is needed to firmly establish the explanatory variance decomposition approach. First, it should be examined how the standard errors from variance regression can be translated into uncertainty estimates for the decomposition components. Second, the approach should be generalized from dichotomous to categorical treatment variables. Lastly, the approach could be extended to other decomposable inequality statistics, such as the Gini coefficient or Theil index.

explanatory variance decomposition approach can be employed to ask questions, such as "What is the causal effect of college on total inequality and on inequality within and between these groups?" and "To what extent do changes in the effect of college over time simply reflect changes in pre-college inequality and changes in group composition?". The library allows to estimate and decompose such effects on inequality with a single command but also gives the possibility to draw on outside estimators. Further, the library makes it possible to decompose inequality at a single timepoint (absolute decomposition) or relative to a reference level (relative decomposition). The reference level can be a point in time (e.g., 1980) or a counterfactual level of inequality or effect of treatment (e.g., motherhood effect on the mean and variance is zero). With the counterfactual approach, questions, such as "What would the effect of higher education be if there was no pre-college inequality?" or "How would women's earnings inequality change if there was no effect of motherhood on women's earnings?", can be analyzed.

To conclude, the proposed explanatory variance decomposition provides researchers with a novel approach to measuring the distributional consequences of group heterogeneity in treatment effect analyses. The method considers heterogeneity in the distribution and effect of treatment across groups, heterogeneity of treatment within groups, and heterogeneity in pre-treatment inequality. By defining treatment effects at the individual level, measuring them at the group level, and quantifying their consequences at the aggregate level, the approach negotiates the impossibility of drawing causal inferences at the individual and aggregate level (Xie 2013).

REFERENCES

Allison, Paul D. 1978. "Measures of Inequality." *American Sociological Review* 865–80.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118(4):1279–1333. doi: 10.1162/003355303322552801.

Baldassarri, Delia, and Maria Abascal. 2017. "Field Experiments Across the Social Sciences." *Https://Doi.Org/10.1146/Annurev-Soc-073014-112445* 43:41–73. doi: 10.1146/ANNUREV-SOC-073014-112445.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8(4):436. doi: 10.2307/144855.

Breen, Richard, and Inkwan Chung. 2015. "Income Inequality and Education." *Sociological Science* 2:454–77.

Budig, M. J., and P. England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66(2):204–25. doi: 10.2307/2657415.

Cancian, Maria, and Deborah Reed. 1999. "The Impact of Wives' Earnings on Income Inequality: Issues and Estimates." *Demography* 36(2):173–84. doi: 10.2307/2648106.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81(6):2205–68. doi: 10.3982/ECTA10582.

Cowell, Frank. 1988. "Inequality Decomposition: Three Bad Measures." *Bulletin of Economic Research* 40(4):309–12.

Cowell, Frank. 2011. *Measuring Inequality*. Oxford University Press.

Cowell, Frank A., and Maria-Pia Victoria-Feser. 1996. "Robustness Properties of Inequality Measures." *Econometrica* 64(1):77. doi: 10.2307/2171925.

Crawford, M., J. Church, and B. Akin. 2014. *CPI Detailed Report. Data for December 2014.*

Cukrowska-Torzewska, Ewa, and Anna Matysiak. 2020. "The Motherhood Wage Penalty: A Meta-Analysis." *Social Science Research* 88–89:102416. doi: 10.1016/J.SSRESEARCH.2020.102416.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64(5):1001. doi: 10.2307/2171954.

Drew, R. J. A., S. Flood, and J. R. Warren. 2014. "Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months." *Journal of Economic and Social Measurement* 39(3):121–44.

Ferro-Luzzi, G. 2010. "A Note on the Determinants of Wage Inequality Between and Within Genders." in *The Measurement of Individual Well-Being and Group Inequalities*, edited by J. Deutsch and J. Silber. Routledge Economics, Taylor and Francis Group.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes:" *Human Relations* 7(2):117–40. doi: 10.1177/001872675400700202.

Firebaugh, Glenn. 2003. *The New Geography of Global Income Inequality*. Harvard University Press.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* 77(3):953–73. doi: 10.3982/ECTA6822.

Firpo, Sergio P., Nicole M. Fortin, and Thomas Lemieux. 2018. "Decomposing Wage Distributions Using Recentered Influence Function Regressions." *Econometrics 2018, Vol. 6, Page 28* 6(2):28. doi: 10.3390/ECONOMETRICS6020028.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. *Decomposition Methods in Economics*. Vol. 4. Elsevier Inc.

Foster, James E., and Efe A. Ok. 1999. "Lorenz Dominance and the Variance of Logarithms." *Econometrica* 67(4):901–7.

Glauber, Rebecca. 2018. "Trends in the Motherhood Wage Penalty and Fatherhood Wage Premium for Low, Middle, and High Earners." *Demography* 55(5):1663–80. doi: 10.1007/S13524-018-0712-5.

Goesling, B. 2001. "Changing Income Inequalities within and between Nations: New Evidence." *American Sociological Review* 66(5):745–61. doi: 10.2307/3088956.

Goldthorpe, John H. 2015. *Sociology as a Population Science*. Cambridge University Press.

Gonalons-Pons, Pilar, Christine R. Schwartz, and Kelly Musick. 2021. "Changes in Couples' Earnings Following Parenthood and Trends in Family Earnings Inequality." *Demography* 58(3):1093–1117. doi: 10.1215/00703370-9160055.

Harkness, Susan. 2013. "Women's Employment and Household Income Inequality." Pp. 207–33 in *Income inequality: Economic disparities and the middle class in affluent countries*, edited by J. C. Gornick and M. Jäntti. Stanford University Press.

Harvey, A. C. 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity." *Econometrica* 44(3):461. doi: 10.2307/1913974.

Hohberg, Maike, Peter Pütz, and Thomas Kneib. 2020. "Treatment Effects beyond the Mean Using Distributional Regression: Methods and Guidance." *PLOS ONE* 15(2):e0226514. doi: 10.1371/JOURNAL.PONE.0226514.

Jackson, Michelle. 2022. "Order from Chaos: Sociology as a Population Science." Pp. 21–37 in *Handbook of Sociological Science: Contributions to Rigorous Sociology*, edited by K. Gërxhani, N. D. de Graaf, and W. Raub. Edward Elgar Publishing.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101(3).

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2):263–92. doi: 10.2307/1914185.

Karoly, Lynn A., and Gary Burtless. 1995. "Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well-Being, 1959-1989." *Demography* 32(3):379–405. doi: 10.2307/2061687.

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates*." *Journal of the American Statistical Association* 50(272):1168–94. doi: 10.1080/01621459.1955.10501299.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11(4):181–209. doi: 10.1257/APP.20180010.

Kröger, Hannes, and Jörg Hartmann. 2021. "Extending the Kitagawa–Oaxaca–Blinder Decomposition Approach to Panel Data:" *The Stata Journal* 21(2):360–410. doi: 10.1177/1536867X211025800.

Lemieux, Thomas. 2002. "Decomposing Changes in Wage Distributions: A Unified Approach." *Canadian Journal of Economics/Revue Canadienne d'économique* 35(4):646–88. doi: 10.1111/1540-5982.00149.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review* 96(3):461–98.

Liao, Tim Futing. 2019. "Individual Components of Three Inequality Measures for Analyzing Shapes of Inequality:" *Https://Doi.Org/10.1177/0049124119875961* 51(3):1325–56. doi: 10.1177/0049124119875961.

Lundberg, Ian. 2022. "The Gap-Closing Estimand: A Causal Approach to Study Interventions That Close Disparities Across Social Categories." *Sociological Methods & Research*. doi: 10.1177/00491241211055769.

Machado, José A. F., and José Mata. 2005. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics* 20(4):445–65. doi: 10.1002/JAE.788.

Massey, Douglas S. 2007. *Categorically Unequal. The American Stratification System*. Russell Sage Foundation.

McCall, Leslie, and Christine Percheski. 2010. "Income Inequality: New Trends and Research Directions." *Https://Doi.Org/10.1146/Annurev.Soc.012809.102541* 36:329–47. doi: 10.1146/ANNUREV.SOC.012809.102541.

Melly, Blaise. 2005. "Decomposition of Differences in Distribution Using Quantile Regression." *Labour Economics* 12(4):577–90. doi: 10.1016/J.LABECO.2005.05.006.

Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

Musick, Kelly, and Wonjeong Jeong. 2021. *His and Her Long Work Hours and Short-Term Employment Changes Following First Birth*. PAA 2021 submission.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3):693. doi: 10.2307/2525981.

Pew Research Center. 2020. "Trends in Income and Wealth Inequality." Retrieved August 9, 2022 (https://www.pewresearch.org/social-trends/2020/01/09/trends-in-income-and-wealth-inequality/).

Piketty, T. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Reardon, Sean F., and Kendra Bischoff. 2011. "Income Inequality and Income Segregation." *American Journal of Sociology* 116(4):1092–1153. doi: 10.1086/657114/ASSET/IMAGES/LARGE/FG12.JPEG.

Rigby, R. A., and D. M. Stasinopoulos. 2005. "Generalized Additive Models for Location, Scale and Shape." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(3):507–54. doi: 10.1111/J.1467-9876.2005.00510.X.

Schwartz, Christine R. 2010. "Earnings Inequality and the Changing Association between Spouses' Earnings." *American Journal of Sociology* 115(5):1524–57. doi: 10.1086/651373.

Shorrocks, Anthony F. 1984. "Inequality Decomposition by Population Subgroups." *Econometrica* 52(6):1369. doi: 10.2307/1913511.

Smith, Heather J., Thomas F. Pettigrew, Gina M. Pippin, and Silvana Bialosiewicz. 2012. "Relative Deprivation: A Theoretical and Meta-Analytic Review." *Personality and Social Psychology Review* 16(3):203–32. doi: 10.1177/1088868311430825.

Weeden, Kim A., Young Mi Kim, Matthew di Carlo, and David B. Grusky. 2007. "Social Class and Earnings Inequality:" *American Behavioral Scientist* 50(5):702–36. doi: 10.1177/0002764206295015.

Western, Bruce, and Deirdre Bloome. 2009. "Variance Function Regressions for Studying Inequality." *Sociological Methodology* 39(1):293–326. doi: 10.1111/J.1467-9531.2009.01222.X.

Western, Bruce, Christine Percheski, and Deirdre Bloome. 2008. “Inequality among American Families with Children, 1975 to 2005.” *American Sociological Review* 73(6):903–20. doi: 10.1177/000312240807300602.

Wodtke, Geoffrey T. 2016. “Social Class and Income Inequality in the United States: Ownership, Authority, and Personal Income Distribution from 1980 to 2010.” *American Journal of Sociology* 121(5):1375–1415. doi: 10.1086/684273.

Wolfson, Michael. 1994. “When Inequalities Diverge.” *The American Economic Review* 84(2):353–58.

Xie, Yu. 2007. “Otis Dudley Duncan’s Legacy: The Demographic Approach to Quantitative Reasoning in Social Science.” *Research in Social Stratification and Mobility* 25(2):141–56. doi: 10.1016/J.RSSM.2007.05.006.

Xie, Yu. 2013. “Population Heterogeneity and Causal Inference.” *Proceedings of the National Academy of Sciences* 110(16):6262–68. doi: 10.1073/PNAS.1303102110.

Xie, Yu, Alexandra Killewald, and Christopher Near. 2016. “Between- and Within-Occupation Inequality: The Case of High-Status Professions.” *The Annals of the American Academy of Political and Social Science* 663(1):53–79. doi: 10.1177/0002716215596958.

CHAPTER 2

A MULTILEVEL MODEL FOR COALITION GOVERNMENTS: UNCOVERING DEPENDENCIES WITHIN AND BETWEEN GOVERNMENTS DUE TO PARTIES

ABSTRACT

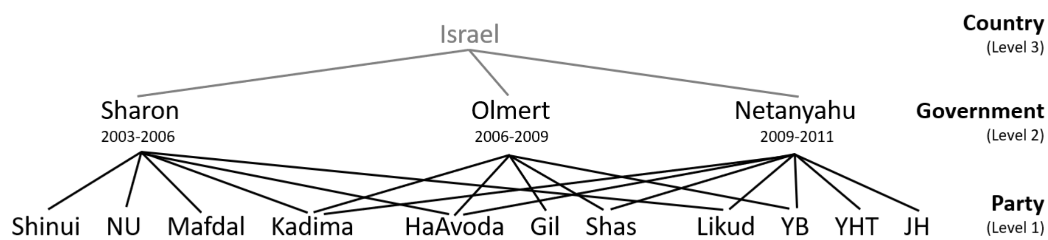
Coalition research increasingly focuses on party-level explanations of coalition outcomes. This work, however, ignores the complex multilevel structure between parties and governments: Many parties become nested in multiple governments over time, and governments are often nested in coalitions of multiple parties. In this paper, I show that this crisscrossing structure (i) induces dependencies among observations that inflate both Type-I and Type-II error rates if ignored, and (ii) reflects an inherent aggregation problem in party-level explanations. I then advance a novel multilevel model to account for the dependencies among observations and explicitly model the aggregation of party effects to the government level. I demonstrate the model's ability to better integrate theoretical expectations into empirical models with an application to coalition government survival as predicted by parties' financial dependencies. The results show that the more parties' financial resources comprise of member contributions, the higher the termination hazard of governments including those parties.

INTRODUCTION

The majority of advanced industrial democracies are parliamentary systems in which coalition governments are the norm. Accordingly, the study of the formation, governance, and termination of coalition governments is a central focus of comparative scholarship. A vast body of literature has established that such outcomes depend on (i) critical events, such as economic recessions, (ii) systemic properties at the country level, such as electoral institutions, and (iii) structural attributes at the government and parliament level, such as ideological heterogeneity or minimal winningness (Strøm, Müller, and Bergman 2008).

In recent years, coalition research has directed attention to party-level explanations of coalition outcomes (Giannetti and Benoit 2009). Efforts to redress the lack of large-scale comparative data on party organization[20] have led to a rise of empirical contributions (Bäck 2008; Bäck, Debus, and Dumont 2011; Ceron 2016; Druckman 1996; Greene 2017; Martínez-Cantó and Bergmann 2019; Martínez-Gallardo 2010; Saalfeld 2009). These studies, however, ignore the complex nonstandard multilevel structure between parties and governments. As Figure 1 shows, in addition to their nesting in countries, parties often become nested in multiple governments over time, and governments are often nested in multiple parties (i.e., are coalition governments). Acknowledging this crisscrossing relationship between parties and governments is key to studying party effects on government outcomes for two reasons.

Figure 1: The multilevel structure of coalition governments in Israel



[20] New data collection efforts include the Political Party Database Project (Poguntke, Scarrow, and Webb 2017), ParlGov (Döring and Manow 2016), the Integrated Party Organization dataset (Giger and Schumacher 2015), and the Party Facts Project (Bederke, Döring, and Regel 2020).

First, from a statistical perspective, the multilevel structure implies dependencies among observations. While dependencies arising from the clustering within countries have been considered by previous research, dependencies that arise from the clustering within parties are ignored in all existing empirical analyses. These dependencies manifest in substantial correlations among outcomes of governments containing the same parties. A conducted simulation study shows that party-level dependencies affect regression coefficients and standard errors not only at the party level but also at the government, and country level. Even though no effect was specified in the simulation, regression coefficients were significant 20 to 60 percent of the time depending on outcome and level of covariate (false positive rate). It is, therefore, crucial to model the multilevel structure to not primarily capture spurious effects.

Second, from a theoretical perspective, the multilevel structure reflects interdependencies of political processes in the parliamentary arena. The outcomes of coalition governments depend on all constituent parties and on the interdependencies among them. Consequently, the interplay of coalition parties in their joint effect on the government must be considered. Analyzing parties' joint effect constitutes cross-level inference as the effects of parties are aggregated to the government level. Multilevel analysis provides a suitable framework to model and test theories on how this aggregation process depends on the features and the configuration of coalition parties. While emphasized in theoretical work (Cross and Katz 2013; Lupia and Strøm 1995; Müller and Strøm 1999; Strøm 1990), the interplay of intra-party and inter-party politics is not properly modeled in empirical research on coalition government outcomes.

In this paper, I propose a new statistical technique to represent the complex multilevel structure of coalition governments, which is based on the multiple membership multilevel model (MMMM) (Goldstein 2011). The proposed model makes explicit that many parties are members of several governments over time and that many governments are multiparty coalitions. This is accomplished by including effects of parties in all governments in which they have participated and modeling the total party effect on a government as the aggregated effect of all parties within that coalition. In this way, dependencies within and across governments that originate at the party level are considered.

The proposed technique further allows to consider interdependencies among coalition parties in their effect on government outcomes. This is accomplished by endogenizing the weights of the MMMM so that the importance of parties in their effect on governments may depend on their relationship to each other (e.g., balance of power). The approach relaxes the implicit assumption of conventional single-level models (SLMs) that party weights are fixed within and across governments. In doing so, the MMMM opens up opportunities to model and test theories on the interplay of parties in their collective impact on government outcomes, which are impossible to test with SLMs.

I illustrate the importance of recognizing the multilevel structure of coalition government outcomes by simulation and with an empirical application. In the simulation study, I use real variables from widely employed datasets and manipulate their effect on simulated outcomes (linear and survival) to assess the consequences of party-level clustering and interdependencies for recovering the true effects. The results show that the proposed generalized MMMM recovers the true regression coefficients and standard errors while conventional SLMs do not.

I demonstrate the model's ability to better integrate theoretical expectations into empirical models with an application to coalition government survival as predicted by parties' financial dependency on their members. The results indicate that (i) the more a party's financial resources comprise of contributions from its members, the higher the termination hazard of a government including this party, and that (ii) this effect is contingent upon the power balance in the coalition.

While developed to study coalition outcomes, the proposed method has wide applicability to cases characterized by similar interdependence structures, such as data on multi-party wars, treaties, and international organizations. With this paper, I provide the R package 'rmm' available on CRAN to estimate the generalized MMMM for a variety of outcomes (linear, logit, conditional logit, Cox, Weibull) in a user-friendly way.

THE MULTILEVEL STRUCTURE OF COALITION GOVERNMENT DATA

Coalition government data exhibit a complex multilevel structure that induces dependencies among observations. First, governments are nested in countries. Since each government is nested in a single country, hierarchical multilevel modeling (termed shared frailty modeling in survival analysis) can be used to account for this hierarchical multilevel structure (Goldstein 2011). The second structure concerns the relationship between governments and parties. This relationship is not hierarchical: parties are members of no, one, or several governments over time and, vice versa, many governments are multiparty coalitions. As shown in Figure 1, Israel's labor party, HaAvoda, was a member of all three governments between 2003 and 2011, each of which was a coalition government. This crisscrossing data structure can be taken into account with the multiple membership multilevel model (MMMM) (Goldstein 2011).

The MMMM was developed in educational research to accommodate that the hierarchical nesting of students in schools is broken up for mobile students who change schools over time (Browne, Goldstein, and Rasbash 2001; Goldstein, Burgess, and McConnell 2007; Leckie 2009). However, while in this application only a small fraction of students is mobile, in coalition government data, the multiple membership structure is the rule rather than the exception. Figure 2 shows the distribution of government participations in the Woldendorp, Keman and Budge dataset (2000). In these data, parties participate in 7.2 coalition governments on average since WWII. However, the distribution is right-skewed with some parties being part of over 40 governments, such as the Italian Christian democrats.

In linear regression, ignoring dependencies among observations will lead to downward-biased standard errors while regression coefficients will be unaffected. In nonlinear models, such as logistic and survival models, both regression coefficients and standard errors will be downward-biased. That is, similarities among observations affect model parameters even in the absence of omitted variable bias (i.e., dependencies are uncorrelated with the covariates in the model) (Box-Steffensmeier and Jones 2004, 141–54).

I conducted a simulation study to assess the consequences of ignoring the complex multilevel structure of coalition government outcomes. To assess the consequences in a realistic setting, I use covariates from datasets widely employed in the literature and only simulate their effect on synthetic (linear and survival) outcomes. The results are presented in Appendix A3 and show that the bias is pronounced, both Type-I and Type-II error rates are affected, and that robust standard errors do not redress the problem. Even though no effect was specified, the regression coefficients of a model ignoring dependencies among observations due to parties were significant 20–60 percent of the time depending on outcome and level of covariate. These results affect not only studies examining party effects but all studies examining coalition government outcomes because party-level dependencies affect parameter estimates at the party, government, and country level. Consequently, it is crucial to model this multilevel structure in order not to primarily capture spurious effects. The bias is likely also pronounced in other fields of political science in which multiple membership structures are frequently encountered, such as multiparty wars, treaties, and international organizations.

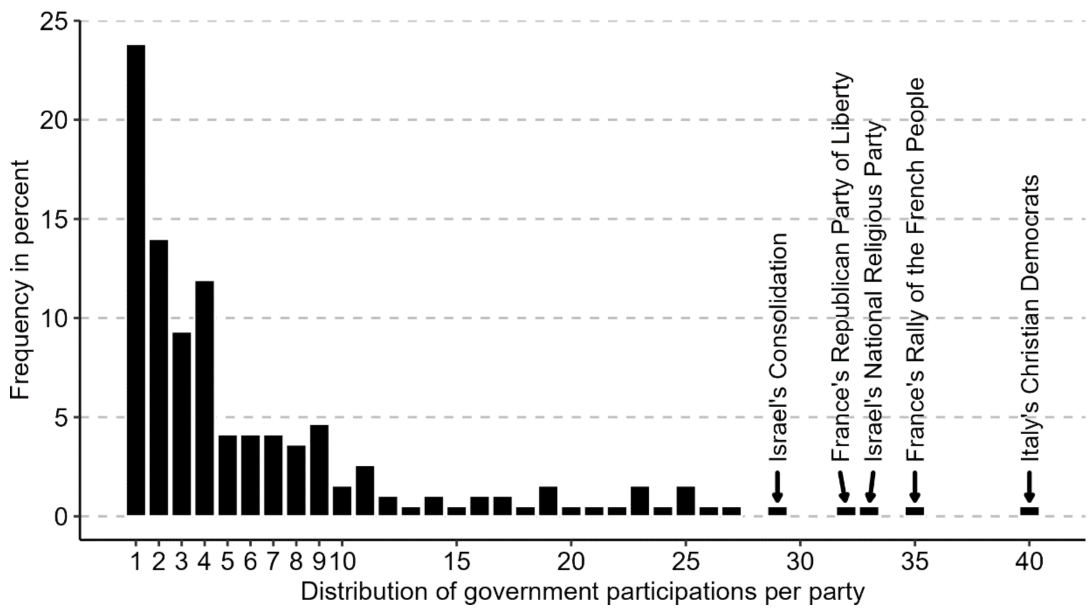
Previous research has employed two strategies to overcome rather than incorporate the multilevel structure – disaggregating government outcomes to the party level or aggregating party features to the government level. Both strategies are problematic. The disaggregation strategy assigns coalition outcomes to all parties in the coalition and thereby treats the outcomes as independent realizations of each coalition party. This approach thus ignores interdependences among coalition partners that may have led to the observed outcome. It also ignores dependencies of government outcomes containing the same parties. Moreover, the approach induces Type-I error because data are artificially multiplied. A three-way coalition, for instance, generates three observations in the dataset even though only one outcome is observed, inflating the sample size on which standard error estimates are based. Finally, the estimated effect sizes are often misinterpreted. As I show in the simulation study in Appendix A3, regression coefficients estimated at the party level should not be interpreted as government-level effects.

Aggregating party features to the government level recognizes that government outcomes depend on all constituent parties. However, while aggregating party features accounts for observed dependencies, similarity among observations that result from unobserved causes are still ignored with this approach.

Moreover, by imposing a fixed aggregation function, such as the arithmetic mean, the aggregation strategy also precludes analyzing the interdependencies of parties in their effect on government outcomes. Finally, owing to defaults in popular statistical programs such as Stata, it is common practice to take the mean of party features *over available information* and thus to ignore missingness data at the party level. For instance, in a three-way coalition with missing information on one party, Stata's mean function will calculate the mean over the features of the available two parties. Research on imputation demonstrates that in the presence of nonrandom missingness and small groups, this strategy induces bias and almost any imputation strategy outperforms listwise deletion (Newman and Sin 2009). Such coding choices are often obscured to the reader (and sometimes the researcher) because the party level is not explicitly modeled.

The following model proposal produces correct standard errors, allows for explicit theorization, and makes coding decisions more transparent, which expose missingness or ad-hoc aggregation procedures.

Figure 2: The distribution of government participations in the WKB dataset



MODELING THE MULTILEVEL STRUCTURE

Let $y^g = [y_1^g, \dots, y_I^g]$ be a coalition government outcome, where subscript $i = 1, \dots, I$ indexes governments and superscript g indicates that the outcome is located at the government level. I propose to model this outcome in terms of a government-level effect θ_i^g, a country-level effect $\theta_{c(i)}^c$, and an aggregated party-level effect θ_i^p to recognize the three levels at which explanations of coalition government outcomes originate:

$$y_i^g = \theta_i^g + \theta_{c(i)}^c + \theta_i^p \tag{1}$$

To simplify the exposition, I propose the generalized MMMM as a linear random intercept model. However, the model structure can be implemented for a large class of models, including generalized linear models and survival models, and can be extended to accommodate random slopes.

The government-level effect

The government-level effect θ_i^g in equation (1) is modeled in terms of a systematic component ($\boldsymbol{\beta}^{\mathbf{g}} \cdot \mathbf{x}_i^{\mathbf{g}}$) to represent the effect of observed covariates at the government level and a random component (u^G) that captures the joint impact of unobserved government-level effects:

$$\begin{aligned} \theta_i^g &= \boldsymbol{\beta}^{\mathbf{g}} \cdot \mathbf{x}_i^{\mathbf{g}} + u_i^g \\ u_i^g &\sim N\left(0, \sigma_{u^g}^2\right), \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}^{\mathbf{g}} \cdot \mathbf{x}_i^{\mathbf{g}}$ is the dot product of $\boldsymbol{\beta}^{\mathbf{g}} = [\beta_1^g, \dots, \beta_G^g]$ and $\mathbf{x}_i^{\mathbf{g}} = [1, x_{i1}^g, \dots, x_{iG}^g]$, and u_i^g at this level is an error term, which is assumed to be normally distributed with a mean of zero, a constant variance of $\sigma_{u^g}^2$, and zero covariance.

The country-level effect

The country-level effect $\theta_{c(i)}^c$ in equation (1) models the hierarchical relationship between governments and countries in which each government i is associated with one country j. That is, subscript $j = 1, \dots, J$ indexes countries and the indexing function $c(i)$ returns the country j to which government i belongs. The country-level effect is also modeled in terms of a systematic and a random component:

$$\theta^c_j = \boldsymbol{\beta^c} \cdot \mathbf{x^c_j} + u^c_j$$
$$u^c_j \sim N(0, \sigma^2_{u^c}), \tag{3}$$

where $\boldsymbol{\beta^c} \cdot \mathbf{x^c_j}$ is defined analogously to the above and u^c_j captures the joint impact of unobserved country-level effects. Including the same random component in all governments of the same country allows coalition governments in the same country to be more similar after conditioning on observed variables. The random components are assumed to be normally distributed with a mean of zero, a constant variance of $\sigma^2_{u^c}$, and zero covariance. The variance measures heterogeneity at the country-level due to unobserved country-level effects.

The party-level effect

The aggregated party-level effect θ^p_i in equation (1) models the relationship between parties and governments. It is a weighted sum of the effect of each party k in the set of parties that constitute government i. That is, subscript $k = 1, \ldots, K$ indexes parties and the indexing function $p(i)$ returns all parties that participate in government i. The individual party effects are determined by a systematic and a random component and then aggregated by calculating their weighted sum with weights $0 \leq w_{ik} \leq 1$:

$$\theta^p_i = \sum_{k \in p(i)} w_{ik}(\boldsymbol{\beta^p} \cdot \mathbf{x^p_{ik}} + u^p_{ik}) = \underbrace{\boldsymbol{\beta^p} \sum_{k \in p(i)} w_{ik}\mathbf{x^p_{ik}}}_{\text{systematic component}} + \underbrace{\sum_{k \in p(i)} w_{ik}u^p_{ik}}_{\text{random component}} \tag{4}$$
$$\text{where } u^P_k \sim N(0, \sigma^2_{u^p})$$

and where $\boldsymbol{\beta^p} \cdot \mathbf{x^p_{ik}}$ is defined analogously to the above and u^p_{ik} captures the joint impact of unobserved party-level effects[21]. Including the same random component in all governments a party participated allows coalition governments that contain the same party to be more similar after conditioning on observed variables. The random components are assumed to be normally distributed with a mean of zero, a constant

[21] If we assume that unobserved party-level effects are constant across parties' government participations, parties have one random component: u^p_k. If we assume that unobserved party-level effects change across parties' government participations, parties have as many random components as participations in government: u^p_{ik} (see Appendix A5).

variance of $\sigma^2_{u^p}$, and zero covariance. The variance measures heterogeneity at the party-level due to unobserved party-level effects.

In contrast to the conventional MMMM, which aggregates only the random component, I propose to include both systematic and random component in the aggregation function in equation (4)[22]. Including the systematic component (i.e., observed party-level variables) in the aggregation function has two advantages. First, researchers incorporate the party level more explicitly in their analysis because the proposed model requires data at the party level. In doing so, both modeling choices and data quality issues become more transparent than when modeling aggregated data. Second, making the aggregation of the systematic component a modeling choice allows researchers to examine the interplay of parties in their collective impact on coalition government outcomes. This is accomplished by endogenizing the weights—an issue to which I turn next.

ENDOGENIZING THE WEIGHT FUNCTION TO MODEL INTERDEPENDENCIES

Prior research studying party effects often calculates the arithmetic mean of features of coalition parties to model the party effect on coalition governments (e.g., Bäck 2008; Druckman 1996; Greene 2017). In equation (4), this is equivalent to $w_{ij} = 1/n_i$ with n_i being the number of parties in coalition i. This modeling choice tacitly assumes that (i) the weights of coalition parties are identical and (ii) the sum of weights is constant across governments. However, not only is it an empirical question whether these assumptions hold, but also theoretical work on the interplay of intra- and inter-party politics offers reasons as to when and why these assumptions will be violated. For instance, violating (i), Gamson (1961) argues that parties' share of government portfolios should be proportional to their legislative seat share; violating (ii), Tsebelis (2002) argues that party effects should be more important in ideologically heterogenous coalitions as intra- and inter-party conflicts amplify each other.

[22] In equation (4), the same weights are used to aggregate systematic and random components. However, the model is readily extended to allow researchers to specify different weights for different covariates as well as the random component. The model can also be extended to include opposition parties (see Appendix A3).

To relax the two assumptions, I propose to model party weights as a nonlinear regression of unobserved w on observed explanatory variables $\mathbf{x}^{\mathbf{w}}$ instead of assigning fixed weights to each party. I suggest the following general form for the weight regression function:

$$\mathrm{w}_{\mathrm{ik}} = \frac{1}{\mathrm{n_i}^{\exp(-\{\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{\mathrm{ik}}^{\mathbf{w}}\})}} \qquad (6)$$
$$\text{subject to } \textstyle\sum_{\mathrm{i}} \sum_{\mathrm{k}} \mathrm{w}_{\mathrm{ik}} = \mathrm{N}$$

where $\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{\mathrm{ik}}^{\mathbf{w}}$ is defined analogously to the above, $\mathrm{n_i}$ is the number of parties in coalition i, and N equals the total number of observations at the party level.

The weight regression coefficients $\boldsymbol{\beta}^{\mathbf{w}}$ measure the effect of explanatory variables $\mathbf{x}^{\mathbf{w}}$ on a party's weight in the aggregated party-level effect θ^{p}. In contrast to $\boldsymbol{\beta}^{\mathbf{p}}$, which describe structural effects of aggregated party-level covariates on government outcomes, $\boldsymbol{\beta}^{\mathbf{w}}$ describe how party-level covariates $\sum \mathrm{w}\mathbf{x}^{\mathbf{p}}$ and the random component $\sum \mathrm{w}\mathrm{u}^{\mathrm{p}}$ should be measured at the government level.

If weight variables have no impact on the aggregation process ($\boldsymbol{\beta}^{\mathbf{w}} = \mathbf{0}$), the weights reduce to $\mathrm{w} = 1/\mathrm{n}$ (mean aggregation) and assumptions (i) and (ii) hold. If $\boldsymbol{\beta}^{\mathbf{w}} \gtrless \mathbf{0}$, the weights reveal a more complex interplay of parties in their effect on governments. That is, weights will deviate from $\mathrm{w} = 1/\mathrm{n}$ and are no longer constant within and/or between governments. Instead, they depend on depend on attributes at the party, government, or country level.

Weight variables at the party level vary within governments and thus allow examining whether some parties are more important than others in their effect on the coalition outcome. Weight variables at the government or country level vary between governments and thus allow examining whether party effects are more important in some governments (or countries) than in others. Such cross-level interactions make it possible to examine the interplay of intra- and inter-party processes.

The functional form $\mathrm{w}_{\mathrm{ik}} = \frac{1}{\mathrm{n_i}^{\exp(-\{\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{\mathrm{ik}}^{\mathbf{w}}\})}}$ constrains the weights between 0 and 1 and the sum of weights within governments to, at most, $\mathrm{n_i}$. That is, $0 \leq \mathrm{w}_{\mathrm{ik}} \leq 1$ and $\forall \mathrm{i}\colon \sum_{\mathrm{k}} \mathrm{w}_{\mathrm{ik}} \leq \mathrm{n_i}$. In governments in which all $\mathrm{w}_{\mathrm{ik}} = 0$ and, accordingly, $\sum_{\mathrm{k} \in \mathrm{p(i)}} \mathrm{w}_{\mathrm{ik}} = 0$, party effects do not matter in determining the outcome. In governments in which all $w_{ik} = 1/n_i$ and, accordingly, $\sum_{k \in p(i)} w_{ik} = 1$, the aggregated party effect is the arithmetic mean of the effect of each government party. In such governments, the outcome changes by $\boldsymbol{\beta}^{\mathbf{w}}$ if all coalition parties jointly increase $\mathbf{x}_{ik}^{\mathbf{w}}$ by one unit. Such an effect is *contextual* in that it requires all coalition parties to change to for the effect to be realized. In governments in which all $w_{ij} = 1$ and, accordingly, $\sum_k w_{ik} = n_i$, the aggregated party effect is the unweighted sum of the effect of each government party. Such an effect is *autonomous* in that increasing $\mathbf{x}_{ik}^{\mathbf{w}}$ of one party alone by one unit impacts the outcome by $\boldsymbol{\beta}^{\mathbf{w}}$ and, therefore, each additional coalition party will increase the total impact of party effects on government outcomes.

To provide identification and prevent a rescaling of the party effects themselves, I constrain the sum of all weights to equal the total number of governments N, i.e. $\sum_i \sum_k w_{ik} = N$. Mean aggregation also constrains the sum of all weights to N by constraining the weights to sum to 1 in each government. In contrast to the constraint imposed by mean aggregation, this constraint allows some governments to have sums larger than 1 and some governments to have sums smaller than 1. However, it does not change the overall sum of weights, which still equals N.

Whether a covariate should enter the model as weight or structural variable must be guided by theory. This is no different from the conventional method of selecting weight variables when preparing the dataset. However, the generalized MMMM makes this modeling choice more transparent and enables researchers to confront their choices with data rather than to impose them unchecked. The simulation study below further shows that the generalized MMMM can safeguard against some types of model misspecification. Of course, weights can remain unmodeled if there is no reason to suspect that they vary between parties, governments, or countries.

Two examples from government survival

1. *Weight variable varies between parties.* Let x^p describe parties' internal cohesion, x^w indicate if they hold the prime ministership, and the outcome be government survival. If $\beta^w = 0$, government survival

depends on the cohesion of the prime minister's party as much as the cohesion of the coalition partners. Accordingly, the effect of party cohesion is best described by the average cohesion among the coalition parties. In contrast, if $\beta^{w} \gtrless 0$, government survival depends on the cohesion of the prime minister's party more (less) than the cohesion of the coalition partners.

2. *Weight variable varies between governments.* Let x^{P} describe parties' internal cohesion, x^{w} describe coalitions' ideological heterogeneity, and the outcome be government survival. If $\beta^{w} \gtrless 0$, the effect of party cohesion on government survival is more (less) important in ideologically dispersed coalitions. Such cross-level interactions allow researchers to examine how intra- and inter-party processes affect each other. Assume the model estimates that $\sum_{k} w_{ik} \approx n_{i}$ for ideologically dispersed coalitions. In that case, the effect of party cohesion is best described by the sum of the effect of each coalition party.

In conclusion, the generalized MMMM makes three contributions. First, it accounts for dependencies among observations due to parties by including party-level random effects. Second, it incorporates the aggregation of party effects into the model so that the aggregation function can be theoretically grounded. I suggested a functional form, but other forms can also be implemented. Finally, it enables researchers to test theories on the interplay of parties in their collective impact on government outcomes by making it possible to endogenize the aggregation function.

ESTIMATION

Estimation of model parameters is done using Bayesian Markov chain Monte Carlo (MCMC) as the likelihood function of the generalized MMMM has no closed-form solution and because Bayesian estimation of multilevel models offers several advantages (see Appendix A1). Accordingly, proper priors must be specified across all parameters to define the posterior distribution, which I discuss in more detail in the empirical application below.

The model can be fitted using the `rmm` R package, which is provided with this paper and available on CRAN. This package offers a user-friendly interface to estimate the generalized MMMM in JAGS

(Plummer 2015; v.4.3) from within R for a variety of models (linear, logit, conditional logit, Cox, Weibull). It allows users to model hierarchical and multiple membership multilevel structures and develop their own weight function(s).

MODEL PERFORMANCE

The simulation study in Appendix A3 also examines parameter bias and Type-I and Type-II error rates of the generalized MMMM. Frequentist properties of Bayesian models are regularly studied in objective Bayesian statistics (Bayarri and Berger 2004; Berger 2006). The Type-I error rate of Bayesian models, for instance, is not known a priori (as compared to the frequentist promise of 5% over the iterations of the simulation when setting the nominal α to 0.05) (Gelman and Tuerlinckx 2000). Bayesian models, however, often have good frequentist properties and it is, therefore, useful to study them. The results show that MMMM estimates both regression coefficients and standard deviations at all three levels without bias.

RELATIONSHIP TO ALBARELLO (2024)

Recently, Albarello (2024) prosed a special case of the generalized MMMM:

$$y_i^g = \boldsymbol{\beta}^{\mathbf{g}} \cdot \mathbf{x}_i^{\mathbf{g}} + \mathbf{1} \cdot \sum_{k \in p(i)} \frac{\exp(\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{ik}^{\mathbf{w}})}{\sum \exp(\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{ik}^{\mathbf{w}})} \mathbf{x}_{ik}^{\mathbf{p}} + u_i^g$$

This model has two major disadvantages. First, it does not differentiate between weight and structural regression coefficients since $\boldsymbol{\beta}^{\mathbf{p}} = \mathbf{1}$. Accordingly, the model is a measurement model and can be only applied to cases where the outcome y^g and the party-level variable $\mathbf{x}^{\mathbf{p}}$ are the same. In their application, for instance, the model measures the extent to which a coalition's policy positions y^g is the average of the coalition parties' policy positions x^p weighted by their legislative seat share x^w. In contrast, the generalized MMMM differentiates between measurement ($\boldsymbol{\beta}^{\mathbf{w}}$) and structural effects ($\boldsymbol{\beta}^{\mathbf{p}}$) but allows researchers to specify offset parameters if so desired (e.g., $\boldsymbol{\beta}^{\mathbf{p}} = \mathbf{1}$).

Second, and most importantly, the model ignores the dependencies among observations due to parties (apart from the modeled covariates) as it omits party-level random components (u^p). Accordingly, regression coefficients and standard deviations suffer from a similar bias as those of conventional SLMs.

APPLICATION: PARTIES' FINANCIAL DEPENDENCY ON THEIR MEMBERS AND THE INTERDEPENDENCE STRUCTURE OF GOVERNMENT SURVIVAL

To illustrate the utility of the generalized MMMM and the importance of modeling the crisscrossing relationship between parties and government, I examine the effect of parties' financial dependency on their members on the survival of coalition governments.

Theoretical considerations

I hypothesize that the more a party's financial resources comprise of contributions from its members, the higher the termination hazard of a government including this party. The reason is that the less financial capital a party receives from sources other than its members, the more it depends on them. Most members cannot be rewarded with office spoils but are reimbursed with promises about future public policy. Such promises confine party leaders who need to build and maintain inter-party agreement and, consequently, increase the risk of political gridlock (Müller and Strøm 1999; Strøm 1990).

Regarding the aggregation of party effects, I examine whether Gamson's law (1961) extends to government survival by testing whether the impact of parties on government survival is proportional to their legislative seat share. Allowing party weights to depend on their seat share relaxes the first of the two tacit assumptions made when mean-aggregating party effects—that all coalition parties are equally important. Furthermore, I hypothesize that party effects are more important in ideologically heterogenous coalitions because intra- and inter-party conflicts amplify each other (Tsebelis 2002). Allowing the sum of weights to depend on the coalition's ideological heterogeneity relaxes the second tacit assumption of mean aggregation— that party effects are equally important across governments.

Data and Measurement

The analysis is based on the Woldendorp, Keman and Budge dataset (2000) to source start and end date of each government, reason for termination, government parties, and their distribution of seats in parliament, on the Comparative Manifestos Project (Volkens et al. 2016) to source political positions of the government parties, and on Round 1a of the Political Party Database (Scarrow, Poguntke, and Webb 2017) to operationalize parties' financial dependency on their rank-and-file.

Government duration is measured as the time between the investiture of a government and new elections. Due to the focus on political gridlock, I differentiate terminations by their underlying motivation (escape gridlock / seize opportunity) rather than by the employed constitutional mechanism (nonelectoral replacements / early elections). To do so, I only consider terminations following some form of conflict and more than one year before the official end of term as events and censor other terminations.

Parties' financial dependency is measured by the share of financial contributions from party members in the total funding of parties, which the Political Party Database obtained from parties' financial reports.

Parties' relative seat share in government is measured as $S_{ij} = \left(\frac{seats_{ij}}{\sum_j seats_{ij}} - \frac{1}{N_i}\right)\left(\frac{N_i}{N_i-1}\right)$, where $seats_{ij}$ is the seat share of party j in government i, and N_i is the number of government parties. $S_{ij} = 0$ if j has as many seats as the other parties, $S_{ij} = 1$ if j holds all seats, and $S = -1$ if j holds no seats within the coalition.

Following Warwick (1994), I compute the standard deviation of coalition parties' right-left score to measure governments' ideological heterogeneity. Benoit and Laver (2007) point to the incomparability of the left-right score across countries and time, which is why I divide the coalition's standard deviation by the left-right standard deviation of all parties in the parliament that year. The measure is therefore relative to the ideological distribution at time and place.

I control for whether the government is in the majority in parliament, a minimal winning coalition, and country fixed effects to eliminate all unobserved time-constant country confounds. The final dataset comprises 401 governments from 18 countries (Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Hungary, Ireland, Israel, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden, United Kingdom) between 1944 and 2014, and a total of 194 parties. Continuous variables are standardized, dividing them by two times their standard deviation, so that the regression coefficients are comparable to those of binary predictors (Gelman 2008). I refer to Appendix A4 for more details on theoretical considerations, data, and measurement.

Model specification

The generalized MMMM as Weibull accelerated failure time model is parameterized as follows:

$$\log\left(t_i^g\right) = \boldsymbol{\beta}^{\mathbf{g}} \cdot \mathbf{x}_i^{\mathbf{g}} + u_i^g + \boldsymbol{\beta}^{\mathbf{c}} \cdot \mathbf{x}_{c(i)}^{\mathbf{c}} + u_{c(i)}^c + \sum_{k \in p(i)} \frac{1}{{n_i}^{\exp(-\{\boldsymbol{\beta}^{\mathbf{w}} \cdot \mathbf{x}_{ik}^{\mathbf{w}}\})}} \left(\boldsymbol{\beta}^{\mathbf{p}} \cdot \mathbf{x}_{ik}^{\mathbf{p}} + u_{ik}^p\right)$$

where the dependent variable t_i^g is time until government termination, the systematic components are defined as above, $u_i^g \sim \text{Gumbel}\left(0, \frac{1}{\rho}\right)$ and thus $\sigma_{u^g}^2 = \frac{\pi^2}{6\rho^2}$, ρ is the shape parameter of the Weibull distribution, $u_j^c \sim N\left(0, \sigma_{u^c}^2\right)$, $u_k^P \sim N\left(0, \sigma_{u^p}^2\right)$, and s.t. $\sum_i \sum_k \frac{1}{{n_i}^{\exp(-\{\beta^w \cdot x_{ik}^w\})}} = I$.

Parameters to be estimated are the vectors of regression coefficients $\boldsymbol{\beta} = [\boldsymbol{\beta}^{\mathbf{g}}, \boldsymbol{\beta}^{\mathbf{c}}, \boldsymbol{\beta}^{\mathbf{p}}, \boldsymbol{\beta}^{\mathbf{w}}]$**,** the variances of the random terms $\boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{2}} = [\sigma_{u^p}^2, \sigma_{u^c}^2]$, and the shape parameter ρ. I use weakly informative priors for computational stability: normal distributions for the regression coefficients: $\boldsymbol{\beta} \sim N(0, 0.001)$, scaled half-t distributions for the standard deviations of the random components: $\boldsymbol{\sigma}_{\mathbf{U}} \sim \text{Half-T}(S = 25, df = 1)$ (Gelman 2006), and an exponential distribution for the shape parameter: $\rho \sim \exp(0.001)$.

Collecting $\boldsymbol{\Theta} = [\boldsymbol{\beta}, \boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{2}}, \rho]$ and $\mathbf{x} = [\mathbf{x}^{\mathbf{g}}, \mathbf{x}^{\mathbf{c}}, \mathbf{x}^{\mathbf{p}}]$, I obtain the posterior distribution $p(\boldsymbol{\Theta}|t, \mathbf{x})$ using Bayesian MCMC in JAGS. I specify 5 chains, a chain-length of 100,000 with a burn-in of 10,000. The MCMC diagnostics suggest that the chains successfully converged to the posterior distributions and robustness checks show that the results are not sensitive to competing prior settings. A detailed discussion of model estimation, convergence, and goodness-of-fit can be found in Appendix A4.

Results

The results are presented in three parts. First, I study the origin of variance in government survival by using the generalized MMMM to decompose the total variance in government survival into party, government, and country components. Next, I present the results of using the generalized MMMM to examine the effect of parties' financial dependency on their members. Finally, I demonstrate how estimated weight coefficients shed light on the interdependence structure of coalition governments.

The origin of variance in government survival

Table 3 presents the results of an intercept-only model that separates the total variance in government survival into party, government, and country components. Unsurprisingly, most variability in government survival is located at the government level. However, 34 percent of the variation originates from differences between governments at the party level, and 24 percent originates from differences between governments at the country level. Consequently, while it is standard in the literature to include country-level predictors, this result vindicates recent efforts to integrate party-level explanations into explanations of government survival.

Table 3: The variance in government survival at each level

		Variance	Percent of total variance
Level	Party	0.811 (0.328)	29.8
	Government	1.319 (0.177)	48.5
	Country	0.589 (0.223)	21.7

MMMM without covariates to estimate the variance at each level.
Mode (standard deviation) of the posterior distribution.

The effect of financial dependency

Table 4 presents the regression results. Model 1 is an SLM estimated on party-level data (i.e., each row constitutes a party's participation in government). All variables are significant because this approach drastically underestimates the standard error at all three levels. (In Appendix A3, I demonstrate by simulation that this strategy underestimates standard errors and that regression coefficients cannot be interpreted at the government level.)

Model 2 is an SLM estimated on government-level data, in which the share of contributions from party members in the total funding of parties has been aggregated using the arithmetic mean. The effect of parties' financial dependency is nonsignificant despite a larger effect size because the standard error too is larger.

Model 3 is an MMMM that accounts for the similarity of observations caused by parties repeatedly appearing in the dataset, but it also uses the arithmetic mean as aggregation function. The effect of parties' financial dependency is considerably larger but remains nonsignificant because the MMMM estimates the actual uncertainty in the effect – roughly double of what the SLM estimates.

Table 4: Regression results

Level	Variable type	Variable	(1) SLM at party level	(2) SLM at gov level	(3) MMMM with w=1/N	(4) MMMM with w=f(Z)
Party	Structural	Financial dependency	**0.134†** (0.0976)	0.293 (0.2429)	0.376 (0.4655)	**0.692*** (0.4003)
	Weight	Relative legislative seat share	-	-	-	**8.617*** (5.0884)
Government	Structural	Majority	**-0.279*** (0.1201)	**-0.385*** (0.2154)	**-0.418*** (0.2358)	**-0.388*** (0.2397)
	Structural	Minimum winning	**-0.897***** (0.1170)	**-0.843***** (0.2083)	**-0.888***** (0.2313)	**-0.903***** (0.2251)
	Weight	Ideological heterogeneity	-	-	-	1.375 (4.8609)
Country	Structural	Fixed effects	yes	yes	yes	yes
N		Parties	1285		194	194
		Governments		401	401	401
		Countries			18	18
DIC			9171	2806	2851	2812
Predictive R^2			0.039	0.039	0.060	0.075

All four models are estimated with Bayesian MCMC. Details on estimation can be found in Appendix A4. The MCMC diagnostics suggest that the chains successfully converged to the posterior distributions. The dependent variable is the hazard rate of government termination, and the reported estimates are the mean and standard deviation of the posterior distribution. I refer to the standard deviation of the posterior as the standard error and to the one-sided tail probability of the posterior as p-value. Note, however, that the standard deviation of the posterior is a more direct measure of uncertainty than the standard error, and the Bayesian analog to the frequentist p-value is a test of direction rather than existence (Marsman and Wagenmakers 2016). The results can be interpreted without this reference to the frequentist paradigm. I chose this framing so that the advantages of modeling the multilevel structure are not mistaken for advantages of a Bayesian approach. Continuous variables are standardized, dividing them by two times their standard deviation, so that the regression coefficients are comparable to those of binary predictors (Gelman 2008). ***p<0.001, **p<0.01, *p<0.05, †p<0.1.

Model 4 is another MMMM that includes parties' relative seat share and coalitions' ideological heterogeneity as weight predictors. Instead of using a static function, the model allows the aggregation process to depend on the coalition's power balance and potential for conflict. The effect of parties' financial dependency is two times as large as the effect in the SLM. Moreover, the standard error is smaller than in the MMMM using mean aggregation. Therefore, the effect of parties' financial dependency has been suppressed by forcing the weights to be 1/N instead of letting them depend on the coalition's interdependence structure.

This result reinforces the conclusion from the simulation study that modeling the aggregation process is key to uncovering party effects when weights are not uniform within and across governments. The impact of parties' financial dependency is considerable. The probability of reaching the 4-year mark decreases by about 13 percentage points per 10 percent of funding by members. This more than the difference in survival probability between minority and majority governments (10 percentage points).

The simulation study also shows that regression coefficients and standard errors of variables at other levels are affected. This is visible, for instance, for the effect of minimal winningness, which is larger in the MMMM than in the SLM.

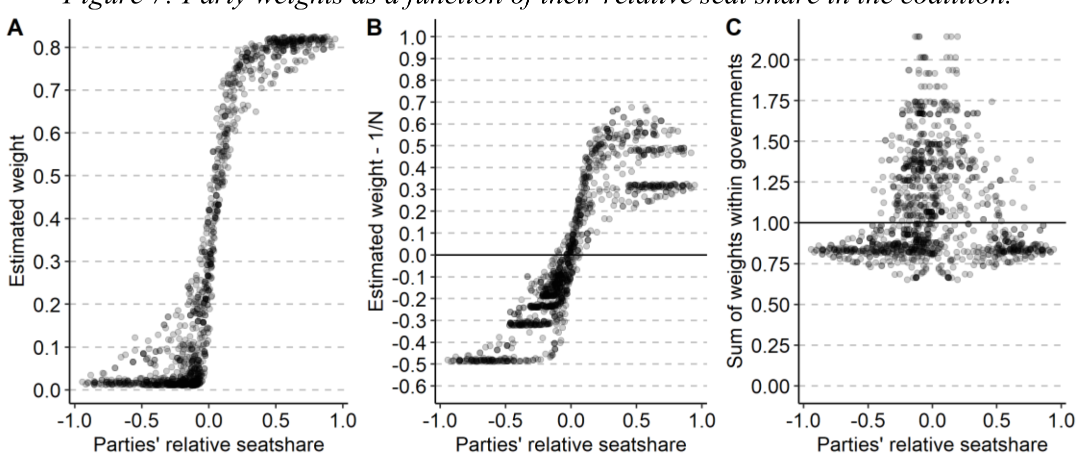
The MMMM enjoys good model-fit and higher predictive power than the SLM. The reported predictive R^2 measures the proportion of explained variance by the linearly corrected prediction function, which quantifies the potential predictive power of survival models (Li and Wang 2019). This power is two times higher for the MMMM ($R^2_{M4} = 0.08$) than for the SLM ($R^2_{M2} = 0.04$). The deviance information criterion (DIC) indicates that modeling the aggregation process improves model-fit considerably. Like the AIC, the DIC is based on the trade-off between model-fit and complexity and allows comparing non-nested models. Smaller DICs indicate better model-fit and differences greater than 10 are considered significant (Spiegelhalter et al. 2002). The DIC of the MMMM including weight predictors is 39 points lower than the MMMM with mean aggregation. The DIC, however, is 6 points higher than the SLM. This is because the DIC favors models with a smaller number of parameters. Unbiased hypothesis tests, however, are worth the increase in complexity in my view.

The interdependence structure of government survival

The results provide strong evidence that the balance of power in the coalition influences the degree to which a coalition party's financial dependency impacts government survival. The financial dependency of a party with more seats relative to other coalitions partners is more important in the effect on government survival than the financial dependency of a party with fewer seats. Panel A of Figure 7 illustrates this relationship. Mean aggregation thus overestimates the impact of parties with fewer seats and underestimates the impact of parties with more seats, which is shown in Panel B of Figure 7. To give an example, in a 3-party coalition with one party holding 60 percent ($S = 0.4$) of the coalition's seat share and the other two each holding 20 percent ($S =$ -0.2), the MMMM estimates the weights to be $w = (0.7, 0.1, 0.1)$. That is, the MMMM reveals that primarily the first party's financial dependency impacts the coalition's termination hazard. Mean aggregation, by contrast, assigns $w = (0.\overline{3}, 0.\overline{3}, 0.\overline{3})$ without consulting the data, which is the source of effect suppression in model (2) and (3). Unequal seat distributions like in this example ($S \geq |0.4|$) are very common; 59 percent of the coalitions fall into this category.

The results do not provide evidence that parties' financial dependency is more or less important in ideologically heterogenous coalitions. Panel C of Figure 7, however, shows that the importance of parties' financial dependency nonetheless varies across governments. The sum of weights peak at $S = 0$, which suggests that parties' financial dependency matters most in coalitions with equal power distribution.

Figure 7: Party weights as a function of their relative seat share in the coalition.



CONCLUSIONS

Research on coalition governments increasingly directs attention to party-level explanations of coalition outcomes (Giannetti and Benoit 2009). This research, however, ignores the complex multilevel structure between parties and governments. In addition to their nesting in countries, parties often become nested in multiple governments over time and governments are often nested in coalitions of multiple parties.

In this paper, I show that the crisscrossing relationship between parties and governments (i) implies dependencies among observations that need statistical control, and that it (ii) exposes interdependencies between parties in their effect on governments that need theorization. I then advance a new statistical method to address both issues. The proposed method represents the multilevel structure of coalition governments by modeling the total party effect on a government as the aggregated effect of all parties in the coalition and by including effects of parties in all governments in which they have participated.

The approach is based on the multiple membership multilevel model (MMMM), which I generalize in two ways. First, while a standard MMMM only aggregates the random part of the model, the proposed model aggregates both fixed and random parts to provide a consistent aggregation of party effects to the government level. Second, I propose to endogenize the weights of an MMMM so that the weights of parties in their effect on a government are no longer fixed but can depend on interdependencies among coalition partners. With this paper, I provide the R package ‘rmm’ to estimate the generalized MMMM for a variety of outcomes (linear, logit, conditional logit, Cox, Weibull) in a user-friendly way.

I illustrate the importance of modeling the multilevel structure by simulation and with an empirical application. The simulation demonstrates that the generalized MMMM recovers the correct regression coefficients and standard errors in the presence of party-level clustering and interdependencies and enjoys good model-fit and higher predictive accuracy than the conventional single-level model (SLM). The SLM, ignoring party-level clustering and interdependencies, by contrast, is biased.

Ignoring party-level clustering leads to underestimated standard errors in linear regression and underestimated regression coefficients and standard errors in survival analysis. This behavior has considerable consequences for hypothesis testing and a precise identification of causal effects. Both false positive rate and power are affected: On datasets widely employed in the literature, government-level regression coefficients were significant over 40 percent and party- and country-level regression coefficients were significant in over 60 percent of the time even though no effect was specified in the simulation. The power to detect effects went down by 20 percentage points for higher degrees of party-level clustering because the variance of the estimator surges.

In survival analysis, the government-level regression coefficients were significant over 20 percent and party- and country-level regression coefficients were significant over 40 percent of the time even though no effect was specified. The power to detect effects went down by 25 percentage points for higher degrees of party-level clustering, owing to the downward bias in the regression coefficients and an increased variance of the estimator.

Ignoring interdependencies in the aggregation process biases both regression coefficients and standard errors at all three levels both in linear regression and survival analysis. The bias is particularly consequential in survival analysis where it further decreases the model's power. Consequently, when interdependencies are present, modeling them is key to uncover effects that, otherwise, are suppressed.

To summarize, the bias of ignoring party-level clustering and interdependencies is substantial and affects covariates at all three levels. Worse yet, the bias continues to grow as parties continue to accumulate government participations over time and their vote shares continue to decline, leading to larger coalitions (Drummond 2006). The simulation shows that available robust standard errors do not redress the problem. It is, therefore, key to model the dependencies within and across governments regardless the level of the covariates considered.

In the empirical application, I employ the MMMM to examine the effect of a party's financial dependency on their members on government survival. The results indicate that the more a party's financial resources comprise of contributions from its members, the higher the termination hazard of a government including this party. This finding provides evidence for the idea that intra-party politics confines party leaders who need to build and maintain inter-party agreement. The results also show that the financial dependency of parties with more seats relative to other coalitions partners is more important in the effect on government survival. Thus, the effect of a party's financial dependency is contingent upon the power balance in the coalition. Owing to the downward bias in the regression coefficients, the effect of a party's financial dependency is not detected using a SLM. Finally, the MMMM allows partitioning the variation in government survival into components at each level. While, unsurprisingly, most variance is located at the government level (42%), the results reveal more variance at the party level (35%) than at the country level (23%). This finding, therefore, vindicates efforts to integrate intra-party politics into explanations of government survival.

In conclusion, this paper contributes to the literature on coalition governments by dissecting the analytical levels of explanations of coalition outcomes and the link between those levels. The paper also proposes a novel way to model the interdependence structure of political processes, not wholly unlike the spatial autoregressive model (SARM) (Juhl 2020). The difference is that the SARM models how parties affect each other's outcomes while the MMMM models interdependences in the joint effect of parties on outcomes at a higher level. This feature makes it possible to test theoretical expectations on aggregation processes. Examples include the aggregation of player strategies in game-theoretic models (Lupia and Strøm 1995) and spatial theory (Laver and Shepsle 1996), or indices that aggregate party features to a higher level, such as power indices (Banzhaf 1964). Theoretical expectations on such processes can be tested with the MMMM by translating them into weight functions.

The paper demonstrates what a multilevel approach can add to research on coalition governments. The proposed approach, however, has the potential to advance our understanding of other political research areas characterized by similar interdependence structures, such as multiparty wars, treaties, and international organizations.

REFERENCES

Bäck, Hanna. 2008. "Intra-Party Politics and Coalition Formation." *Party Politics* 14 (1): 71–89.

Bäck, Hanna, Marc Debus, and Patrick Dumont. 2011. "Who gets what in coalition governments?: Predictors of portfolio allocation in parliamentary democracies." *European Journal of Political Research* 50 (4): 441–78.

Banzhaf, John F. 1964. "Weighted voting doesn't work: A mathematical analysis." *Rutgers Law Review* 19 (2): 317–43.

Bayarri, M. J., and J. O. Berger. 2004. "The Interplay of Bayesian and Frequentist Analysis." *Statistical Science* 19 (1).

Bederke, Paul, Holger Döring, and Sven Regel. 2020. "Party Facts project." 2020. https://partyfacts.herokuapp.com/.

Benoit, Kenneth, and Michael Laver. 2007. "Estimating party policy positions: Comparing expert surveys and hand-coded content analysis." *Electoral Studies* 26 (1): 90–107.

Berger, James. 2006. "The case for objective Bayesian analysis." *Bayesian Analysis* 1 (3).

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge: Cambridge University Press.

Ceron, Andrea. 2016. "Inter-factional conflicts and government formation." *Party Politics* 22 (6): 797–808.

Cross, William P., and Richard S. Katz. 2013. *The challenges of intra-party democracy*. *Comparative Politics*. Oxford: Oxford University Press.

Döring, Holger, and Philip Manow. 2016. *ParlGov 2015 Release*.

Druckman, J. N. 1996. "Party Factionalism and Cabinet Durability." *Party Politics* 2 (3): 397–407.

Drummond, Andrew J. 2006. "Electoral Volatility and Party Decline in Western Democracies: 1970–1995." *Political Studies* 54 (3): 628–47.

Gamson, William A. 1961. "A Theory of Coalition Formation." *American Sociological Review* 26 (3): 373–82.

Gelman, Andrew, and Francis Tuerlinckx. 2000. "Type S error rates for classical and Bayesian single and multiple comparison procedures." *Computational Statistics* 15 (3): 373–90.

Giannetti, Daniela, and Kenneth Benoit. 2009. "Intra-Party Politics and Coalition Governments." In *Intra-Party Politics and Coalition Governments*, eds. Daniela Giannetti and Kenneth Benoit. London, New York: Routledge, 3–24.

Giger, Nathalie, and Gijs Schumacher. 2015. *Integrated Party Organization Dataset (IPOD)*.

Goldstein, Harvey. 2011. *Multilevel statistical models*. 4th ed. Oxford: Wiley.

Grabisch, Michel, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. 2009. *Aggregation functions*. Cambridge: Cambridge University Press.

Greene, Zachary. 2017. "Working through the issues: How issue diversity and ideological disagreement influence coalition duration." *European Political Science Review* 9 (4): 561–85.

Hodges, James S. 2014. *Richly parameterized linear models: Additive, time series, and spatial models using random effects*. Boca Raton: CRC Press.

Hox, Joop J., Mirjam Moerbeek, and Rens van de Schoot. 2017. *Multilevel Analysis: Techniques and Applications*. 3rd ed. *Quantitative methodology series*. New York: Routledge.

Juhl, Sebastian. 2020. "The Sensitivity of Spatial Regression Models to Network Misspecification." *Political Analysis* 28 (1): 1–19.

Laver, Michael, and Kenneth A. Shepsle. 1996. *Making and breaking governments: Cabinets and legislatures in parliamentary democracies*. Cambridge: Cambridge University Press.

Li, Gang, and Xiaoyan Wang. 2019. "Prediction Accuracy Measures for a Nonlinear Model and for Right-Censored Time-to-Event Data." *Journal of the American Statistical Association* 114 (528): 1815–25.

Lupia, Arthur, and Kaare Strøm. 1995. "Coalition Termination and the Strategic Timing of Parliamentary Elections." *American Political Science Review* 89 (3): 648–65.

Marsman, Maarten, and Eric-Jan Wagenmakers. 2016. "Three Insights from a Bayesian Interpretation of the One-Sided P Value." *Educational and Psychological Measurement* 77 (3): 529–39.

Martínez-Cantó, Javier, and Henning Bergmann. 2019. "Government termination in multilevel settings. How party congruence affects the survival of sub-national governments in Germany and Spain." *Journal of Elections, Public Opinion and Parties* 97 (2): 1–21.

Martínez-Gallardo, Cecilia. 2010. "Out of the Cabinet: What Drives Defections From the Government in Presidential Systems?" *Comparative Political Studies* 45 (1): 62–90.

Müller, Wolfgang C., and Kaare Strøm, eds. 1999. *Policy, office, or votes?: How Political Parties in Western Europe Make Hard Decisions*. Cambridge: Cambridge University Press.

Newman, Daniel A., and Hock-Peng Sin. 2009. "How Do Missing Data Bias Estimates of Within-Group Agreement?" *Organizational Research Methods* 12 (1): 113–47.

Poguntke, Thomas, Susan Scarrow, and Paul Webb. 2017. *Political Party Database Version 1a.3*.

Saalfeld, Thomas. 2009. "Intra-Party Conflict and Cabinet Survival in 17 West European Democracies, 1945-1999." In *Intra-Party Politics and Coalition Governments*, eds. Daniela Giannetti and Kenneth Benoit. London, New York: Routledge, 169–86.

Scarrow, Susan, Thomas Poguntke, and Paul Webb, eds. 2017. *Organizing Representation. Political Parties, Participation, and Power*. Oxford: Oxford University Press.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. 2002. "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B* 64 (4): 583–639.

Strøm, Kaare. 1990. "A Behavioral Theory of Competitive Political Parties." *American journal of political science:* 565–98.

Strøm, Kaare, W. C. Müller, and T. Bergman, eds. 2008. *Cabinets and Coalition Bargaining: The Democratic Life Cycle in Western Europe*. *Comparative Politics*. Oxford: Oxford University Press.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton University Press.

Volkens, Andrea, Pola Lehmann, Theres Matthieß, Nicolas Merz, and Sven Regel. 2016. "The Manifesto Data Collection. Version 2016b." Berlin: WZB.

Warwick, Paul. 1994. *Government Survival in Parliamentary Democracies*. Cambridge University Press.

Woldendorp, Jaap, Hans Keman, and Ian Budge. 2000. *Party Government in 48 Democracies (1945 -1998): Composition-Duration-Personnel*. Dordrecht: Springer.

CHAPTER 3

SOCIOECONOMIC SEGREGATION IN ADOLESCENT FRIENDSHIP NETWORKS: A NETWORK ANALYSIS OF SOCIAL CLOSURE IN US HIGH SCHOOLS.

ABSTRACT

Adolescent friendship networks are characterized by low interaction across both socioeconomic and racial lines. Using data from the National Study of Adolescent Health and a new exponential random graph modeling approach, this study examines the degree, pattern, and determinants of socioeconomic segregation and its relationship to racial segregation in friendship networks in high school. The results show that friendship networks are overall less socioeconomically segregated than they are racially segregated. However, the exclusion of low-SES students from high-SES cliques is pronounced and, unlike racial segregation, unilateral rather than mutual: many friendship ties from low-SES students to high-SES peers are unreciprocated. The decomposition of determinants indicates that about half of the socioeconomic segregation in friendship networks can be attributed to differences in socioeconomic composition between schools. The other half is attributable to students' friendship choices within schools and driven by stratified courses (about 13 percent) as well as racial and socioeconomic preferences (about 37 percent). In contrast, relational mechanisms like triadic closure – long assumed to amplify network segregation – have only minor effects on socioeconomic segregation. These results highlight that SES-integrated friendship networks in educational settings are difficult to achieve without also addressing racial segregation. Implications for policymakers and educators are discussed.

INTRODUCTION

An extensive literature on social capital highlights how friendships not only promote psychological well-being but also provide access to the human, cultural, and economic capital of peers.[1] Recent research using sociometric network data indicates that peer resources impact life trajectories of adolescents with low socioeconomic status (SES). For low-SES youth, friendships that cross socioeconomic boundaries predict higher attainment in middle school (Lessard and Juvonen 2019) and high school (Chung 2020), and greater socioeconomic success later in life (Chetty et al. 2022a). For high-SES youth, in contrast, peer resources appear redundant with the resources provided by family and school, and ties to low-SES peers do not seem to affect socioeconomic outcomes (Chetty et al. 2022a; Chung 2020; Lessard and Juvonen 2019).[2] Cross-SES interaction, therefore, has the potential to increase upward mobility of disadvantaged youth without compromising the life-chances of their high-SES peers.

This potential is constrained in the United States by socioeconomic segregation, separating adolescents not only physically into different neighborhoods and schools, but also socially into different friendship circles (Chetty et al. 2022b; Kao, Joyner, and Balistreri 2019; Malacarne 2017; McFarland et al. 2014; McMillan 2022; Mouw and Entwisle 2006; Smith, McPherson, and Smith-Lovin 2014; Zeng and Xie 2008). The effects of physical distance on limiting interaction across socioeconomic lines are intuitively obvious and well-documented (for an overview, see Mijs and Roe 2021). In contrast, the determinants of social distance are much less straightforward theoretically and the empirical data are far more difficult to acquire. As a result, the mechanisms underlying socioeconomic segregation in friendship networks are less well-understood. Accordingly, this paper seeks a deeper understanding of socioeconomic segregation in friendship networks and the processes that create it.

I focus on friendship networks in high school because high schools offer a more diverse socioeconomic composition than primary schools, middle schools, and universities (Chetty et al. 2022b; Kalogrides and Loeb 2013). Yet, despite greater contact opportunities between peers with different SES backgrounds, high school students are no more likely to form friendships across socioeconomic lines (Chetty et al. 2022b;

Malacarne 2017). The underlying puzzle motivating my research is: Why are high school students not forming cross-SES friendships despite these opportunities? The significance of cross-SES interaction in high school goes beyond opportunities, however. Early adolescence is the developmental period in which important group and identity formation processes unfold to shape life-long social identities (Cotterell 2007). The presence or absence of cross-SES interaction during this period will not only affect the socioeconomic attainment of disadvantaged youth but also shape future social and political divisions in society (Gidron and Hall 2020; Putnam 2016). Accordingly, interventions to promote social integration may strategically target high schools, but their effectiveness will depend on a better understanding of socioeconomic segregation in high school friendship networks.

Two challenges stand in the way of progress. First, diverging results in the extant literature expose limitations in our understanding of the degree and patterns of socioeconomic segregation. For example, Chetty et al. (2022b) found greater socioeconomic segregation in high school friendship networks than prior studies (Kao et al. 2019; Malacarne 2017; McFarland et al. 2014; McMillan 2022; Mouw and Entwisle 2006; Smith et al. 2014; Zeng and Xie 2008). This discrepancy may or may not reflect changing friendship patterns because data and measurement differences across studies make it difficult to compare results. More descriptive research in a single comprehensive framework is needed to better understand the degree and specific patterns of socioeconomic segregation and how it compares to other dimensions of segregation.

Second and most importantly, we lack understanding of the determinants of socioeconomic segregation, especially within schools. Most racial and socioeconomic desegregation efforts to date, such as busing, zoning, and affirmative action, are designed to reduce segregation between schools. Chetty and colleagues (2022b), however, find that between-school segregation explains only half of the total socioeconomic segregation in friendship networks. The other half is determined by processes within schools, which Chetty and colleagues subsume under “friending bias” (i.e., apparent SES homophily). However, aside from SES homophily, several other processes can induce socioeconomic segregation. Prior research suggests that stratified settings (e.g., courses), homophily[3] on correlates of SES (e.g., racial homophily), popularity differences, and relational mechanisms (e.g., triadic closure) can also contribute to socioeconomic segregation within schools. Prior studies have focused on different potential determinants but omitted measures of alternatives that may confound their results. A comprehensive decomposition of determinants is therefore needed to disentangle their relative contributions. In particular, the extent to which racial homophily shapes socioeconomic segregation remains unclear.

Given these limitations in prior research, the present study contributes to our understanding of socioeconomic segregation in high school friendship networks in two ways. First, the study offers a rich description of the degree and patterns of socioeconomic segregation and its relationship to racial segregation. Second, the study provides a comprehensive decomposition of the determinants of socioeconomic segregation. The decomposition not only splits segregation into within- and between-school components but also moves beyond friending bias as a singular concept to disentangle the contribution of multiple determinants within schools:

- Neighborhood, course, and extracurricular activities that may be socioeconomically stratified and thus impose structural barriers to cross-SES interaction.
- Student preferences, which can indicate both intentional (preferences regarding peers' SES) and unintentional socioeconomic segregation (preferences regarding peers' race or academic performance).
- Relational mechanisms like triadic closure that reflect students' tendency to base friendship choices on already existing friendship ties and thus may amplify existing tendencies.

The study uses data on friendship networks in the National Study of Adolescent Health (Add Health). Add Health is a nationally representative sample of 7th to 12th graders across a wide range of school contexts that includes sociometric and sociodemographic data. The sociometric network data provide information both on exposure (attending the same school) and interaction (friendship), which is a key distinction between this study and work on socioeconomic segregation using compositional data (e.g., Owens, Reardon, and Jencks 2016) or egocentric network data (e.g., J. A. Smith et al. 2014). The sociodemographic data provide measures of key determinants of friendship segregation. These data have been extensively used to study friendship segregation along racial lines. I integrate prior efforts that studied as determinants students' residential locations (Mouw and Entwisle 2006), course selections (Frank, Muller, and Mueller 2013), extracurricular participation (Schaefer, Simpkins, and Vest Ettekal 2018), and academic performances (Flashman 2012) to disentangle their relative contributions to socioeconomic segregation.

The Add Health data are not without limitations, but I overcome these by employing multiple imputation to target the missing data problem and by using survey weights to make the results generalizable to the larger population of adolescents. Generalizability sets this study apart from network studies based on samples confined to single locations or regions (e.g., McMillan 2022) and network studies using Add Health without survey weights even though select groups are oversampled (e.g., Kao et al. 2019). Chetty et al. (2022b) also provide generalizable results but have to predict SES background entirely based on aggregate (zip-code level) data. The Add Health data, in contrast, are collected directly from students, parents, and school administrators. Moreover, Chetty and colleagues measure friendships on Facebook. It is anything but guaranteed and hard to verify if Facebook friendships reflect offline interaction (Dunbar 2016). In contrast, Add Health measured not only if students are friends but also if they spend time together.

I examine friendship networks using exponential random graph modeling (ERGM). ERGM is a principled statistical approach to modeling social networks in which the whole network of interdependent dyads is considered a single observation. ERGM allows researchers to test competing and intersecting tie-formation mechanisms and examine how these mechanisms shape aggregate segregation patterns (Duxbury 2023; Robins, Pattison, and Woolcock 2005; Snijders and Steglich 2015). This approach to the study of segregation determinants is preferable to dyadic network models as employed, for instance, by Chetty et al. (2022b). Dyadic network models ignore tie-formation mechanisms that depend on other ties (e.g., triadic closure) and thus risk overestimating homophilous tendencies that are artifacts of unmeasured structural confounds. ERGM, in contrast, allows me to specify a comprehensive friendship formation model that captures the most important tie-formation mechanisms shaping socioeconomic segregation.

The results of this study advance knowledge of adolescent friendship segregation in three critical respects. First, socioeconomic segregation in friendship networks in high school is characterized by exclusion of students in the bottom third and closure among students in the upper half of the SES distribution. While friendship networks are overall less socioeconomically segregated than they are racially segregated, the exclusion of low-SES students from high-SES cliques is as pronounced. Moreover, whereas racial segregation is marked by mutual exclusion, socioeconomic segregation is asymmetric in that many friendship ties from low-SES students to peers with higher SES backgrounds are unreciprocated.

Second, in line with Chetty and colleagues (2022b), about half of the socioeconomic segregation in friendship networks can be attributed to disparities in the socioeconomic composition between schools while the other half is attributable to friending bias within schools. Moving beyond friending bias as a singular concept, the decomposition reveals that, within schools, socioeconomic segregation is shaped by stratified courses (about 13 percent) and students' (revealed) racial and socioeconomic preferences (about 37 percent). Among students' revealed preferences (i.e., friending rates net of other modeled mechanisms), racial homophily has the greatest impact on socioeconomic segregation albeit promoting both same- and cross-SES ties. In contrast to students in European secondary schools (Chabot 2024; Lenkewitz 2023; Zwier and Geven 2023), high schoolers also reveal socioeconomic preferences. They prefer peers in the top third of the SES distribution regardless of their own SES background, suggesting an overlap of social and socioeconomic status in high school.

Finally, relational mechanisms like triadic closure increase clustering but have only minor effects on school-wide socioeconomic segregation. This result challenges a long-held assumption that triadic closure amplifies network segregation (e.g., Goodreau, Kitts, and Morris 2009; Kossinets and Watts 2009; Tóth et al. 2021; Wimmer and Lewis 2010).

Taken together, the results show that socioeconomic segregation in high school is characterized by exclusion of low-SES students from high-SES cliques and shaped by compositional differences between schools, stratified courses within schools, and students' racial and socioeconomic preferences. The large impact of racial homophily suggests that SES-integrated friendship networks in educational settings are difficult to achieve without also addressing racial segregation.

DETERMINANTS OF SOCIOECONOMIC SEGREGATION IN HIGH SCHOOL FRIENDSHIP NETWORKS

Researchers began collecting sociometric network data as early as the 1920s, and soon thereafter, they also began studying patterns of socioeconomic segregation in friendship networks (Coleman 1961; Hollingshead 1949, 1975; Neugarten 1946; Tuma and Hallinan 1979). This work is largely univariate and confined to single localities (e.g., schools). The inclusion of a network module in the 1985 General Social Survey spurred work based on ego-centric network data that is generalizable to larger populations (Marsden 1987; Smith et al. 2014; Verbrugge 1977; Wright 1997). The systematic analysis of determinants of segregation patterns using sociometric network data began with the development of statistical network models in 1980s, such as the ERGM (Holland and Leinhardt 1981). This literature (reviewed in Appendix C1) emphasizes five key determinants of socioeconomic segregation in high school friendship networks: compositional differences between schools, and, within schools, stratified settings, homophilous tendencies, popularity differences, and relational mechanisms (Rivera, Soderstrom, and Uzzi 2010). I discuss these determinants (summarized in Figure 1) in the following.

Socioeconomic segregation between schools

Parents' choices on where to live and which school to enroll their children produce differences in SES composition between school districts and schools that limit students' opportunities to form friendship across socioeconomic lines (Bischoff and Reardon 2014; Caetano and Macartney 2021; Owens et al. 2016). Chetty and colleagues (2022b) find that about half of the socioeconomic segregation in friendship networks is determined by such compositional differences between schools while students' friendship choices within schools account for the other half.

Socioeconomic segregation within schools: Stratified settings

Within schools, sorting and self-selection processes create settings (or social foci) in which students are more likely to become friends simply because they meet regularly and participate together in course, extracurricular, and neighborhood activities (Feld 1981). School settings can induce socioeconomic segregation in friendship networks if their SES composition is more homogenous than the school overall (i.e., if settings are stratified). In particular, courses are stratified due to academic tracking (Chmielewski 2014; Frank et al. 2013; Kalogrides and Loeb 2013). In contrast, if school settings are less homogenous than the school overall, proximity effects on friendship formation will reduce socioeconomic segregation in friendship networks. It has been suggested that extracurricular activities, such as sports teams, may constitute such settings within schools (Lessard and Juvonen 2019; Malacarne 2017). Lastly, while neighborhoods drive socioeconomic segregation between schools, it is unclear whether residential proximity (e.g., taking the same bus to school) also creates stratified settings within schools. In traditional neighborhood schools, neighborhood and school socioeconomic composition largely overlap. The expansion of alternatives since 1990, however, may have created this new layer of stratification within schools (Bischoff and Tach 2020; Rich and Owens 2023).

Besides structural barriers (i.e., differences in SES composition between schools and stratified settings within schools), students' preferences may also explain socioeconomic segregation. Prior research differentiates between a desire to befriend similar others (homophilous tendencies) and an aspiration to befriend others with greater characteristics (aspirational tendencies) that produces popularity differences across attributes (Snijders and Lomi 2019).

Socioeconomic segregation within schools: Homophilous tendencies

Social-psychological research shows that people tend to choose friends similar to themselves on a wide variety of characteristics (McPherson, Smith-Lovin, and Cook 2001). Crucially, not homophily with respect to SES background will induce socioeconomic segregation but also homophily with respect to attributes correlated with SES, such as academic performance and race. Their impact on socioeconomic segregation depends on the strength of homophily with respect to the attribute and on the correlation between the attribute and SES (Blau 1977; Simmel 1908).

SES homophily. In contrast to racial homophily, SES homophily is thought to be less a result of intentional ingroup favoritism (or outgroup exclusion) and more an unintended consequence of mutual attractions based on shared cultural tastes, interests, beliefs, and behaviors (Edelmann and Vaisey 2014; Lewis and Kaufman 2018; Lizardo 2006; Mark 2003). A long tradition of sociological research provides thick descriptions of how these mutual attractions correlate with SES background to create more or less salient boundaries (Bourdieu 1984; Ferguson and Lareau 2021; Khan 2011; Lamont and Fournier 1992; Lamont and Lareau 1988; Small 2009). Quantitative evidence on SES homophily, by contrast, is inconclusive. Not only do we lack causal evidence for SES homophily, but also studies using observational data provide limited evidence because SES homophily is often not sufficiently separated from correlated drivers. Moreover, even though homophily is sometimes portrayed as an invariable law, empirically, homophilous tendencies show dependencies on age, period, cohort, as well as institutional and cultural context (McFarland et al. 2014). For example, Hollingshead's (1949) study of Elmtown High was among the first to uncover socioeconomic clustering among students. When Coleman revisited this school ten years later, he found little evidence for it (Cohen 1979; Coleman 1961). Similarly, while recent studies find evidence for SES homophily in the US high schools (Malacarne 2017) and Chinese middle schools (An 2022), there is little evidence of it in European secondary schools (Chabot 2024; Lenkewitz 2023; Zwier and Geven 2023).

GPA homophily. Students are more likely to befriend peers with similar academic performance (Flashman 2012; Quillian and Campbell 2003; Tuma and Hallinan 1979). Since GPA is also correlated with SES background (Hanushek et al. 2022; Sirin 2005), homophily based on academic performance can be expected to induce socioeconomic segregation in friendship networks.

Race homophily. Whereas racial segregation in friendship networks is *not* a by-product of SES homophily (Moody 2001; Zeng and Xie 2008), the extent to which racial homophily shapes socioeconomic segregation is unclear. This question hinges upon whether race and SES background correlate within schools. While there is a persistent race-SES correlation at aggregate levels, it would be a fallacy to infer from it a correlation within schools. Schools are socioeconomically selective, and this selectiveness reshapes the link between race and SES background within schools. In the Add Health schools, the race-

SES correlation between schools is $\rho_B = 0.24$. Within schools, in contrast, it is only $\rho_W = 0.11$ (see Appendix C2 for details). In fact, within schools, SES disparities are greater within racial groups than between them. SES disparities within racial groups produce a countervailing integrative effect of racial homophily by inducing cross-SES ties within racial groups.

Socioeconomic segregation within schools: Popularity differences

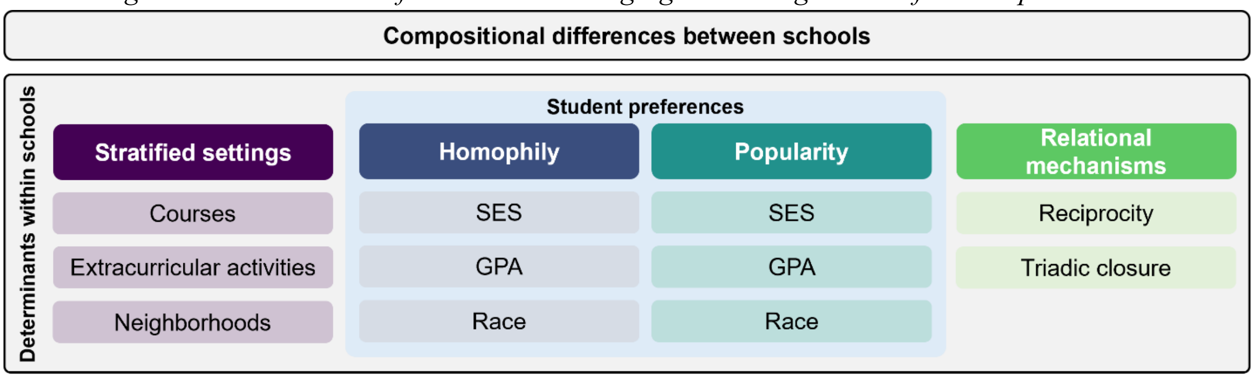
Rather than a preference to befriend similar peers, students may also aspire to befriend peers with a higher status than themselves (An 2022; An and McConnell 2015; Dijkstra, Cillessen, and Borch 2013; Homans 1961; Laumann 1965; Mark 2003; Skvoretz 2013). Such aspirational tendencies can affect socioeconomic segregation if social and socioeconomic status overlap in high school. Popularity differences by SES background attest to this consolidation of social and socioeconomic status (Hjalmarsson and Mood 2015; Malacarne 2017; Raabe, la Roi, and Plenty 2024; Zeng and Xie 2008). The consequences for socioeconomic segregation are unclear, however. The aspiration to befriend high-SES peers nudges all but the students at the top of the SES distribution to form cross-SES friendships. The (so far unexamined) question is whether high-SES students reciprocate such ties. In the same vein can popularity differences on correlates of SES affect socioeconomic segregation in friendship networks.

Socioeconomic segregation within schools: Relational mechanisms

Socioeconomic segregation in friendship networks can also be affected by relational mechanisms: tie-formation mechanisms that depend on the structure of other ties in the network. Relational mechanisms reflect students' tendency to base friendship choices on already existing friendship ties. Much like Schelling's (1971) classic model of residential segregation, relational mechanisms have the potential to dynamically amplify existing tendencies (e.g., modest preferences) to substantial segregation patterns. They must therefore be considered so as not to overestimate other modeled determinants and because such "social multiplier" effects are interesting in and of themselves.

The most important relational mechanisms are the tendency to reciprocate friendships and the tendency to befriend others with whom we have a friend in common (triadic closure). Reciprocation and balance norms are thought to amplify socioeconomic segregation (e.g., Goodreau, Kitts, and Morris 2009; Kossinets and Watts 2009; Tóth et al. 2021; Wimmer and Lewis 2010). The intuition is that, if high-SES students are more likely to befriend other high-SES peers, the tendency to reciprocate such friendships and close open triads will produce even more of such ties. However, not only do we lack convincing evidence whether this is this case empirically,[4] but also analytical work on the effect of balancing mechanisms on network segregation comes to contradictory conclusions (Abebe et al. 2022; Asikainen et al. 2020; Grund 2014).

Figure 1: Determinants of socioeconomic segregation in high school friendship networks.



DATA AND METHODS

Data

The data for this study come from the National Study of Adolescent Health (Add Health) (Harris 2018). Add Health is a nationally representative sample of more than 90,000 adolescents in grades 7-12. The first wave was collected during the 1994-95 school year from a carefully stratified sample of 132 middle and high schools (to include public and private and to vary by size, location, racial and socioeconomic composition). Since then, four follow-up waves have been conducted. I limit my sample to the first wave in which all students were asked to nominate up to five closest male and five closest female friends from a roster. In addition to the sociometric and sociodemographic data in Add Health, I draw on measures in an expansion of Add Health, the Adolescent Health and Academic Achievement study (AHAA) (Muller et al. 2007). AHAA collected high school transcripts of more than 12,000 participants in wave 3 to provide students' GPA and course selections in the first wave for all respondents that did not drop out of the panel. The analyses in this paper are based on two samples of these data.

Sample 1. The description of socioeconomic segregation within and between schools is based on the core sample of more than 83,000 adolescents in 128 school in which both in-school and in-home interviews were administered (about 200 in-home interviews per school). Since Add Health oversampled select groups, I employ survey weights (design and nonresponse adjustment) to provide population-averaged measures of socioeconomic segregation within and between schools. Computational details are provided in Appendix A4.

Sample 2. The decomposition of determinants of socioeconomic segregation within schools requires detailed information and is thus based on a sample of 5,942 adolescents in the 15 schools with the most complete data. This sample largely coincides with the "saturated" sample, a subset of schools in which all students were interviewed in school and at home. The saturated sample includes two large schools (with a total combined enrollment exceeding 3,100) and 14 small schools (with enrollments of fewer than 300). One of the large schools is predominantly white and is located in a mid-sized town. The other is ethnically heterogeneous and is located in a major metropolitan area. The 14 small schools, some public and some private, are located in both rural and urban areas. Therefore, the saturated sample is diverse but not a random sample. Effect estimates are therefore not population-averaged associations but average associations in the sample. However, comparisons show that students in both samples exhibit similar sociodemographic profiles and friendship networks are similarly segregated along sociodemographic lines.

I employ multiple imputation using predictive mean matching to eliminate any missing values in both samples because network analyses like this one require complete data and because listwise deletion of students with incomplete information can lead to incorrect inference. Listwise deletion is particularly problematic for the purposes of this paper as removing students from the networks distorts the analyzed network structure (i.e., changes the number of dyads, triads, and other network structures), which could bias the analysis of relational mechanisms. Moreover, the primary cause for missing data in the two samples is that the in-home interviews were only administered to a subset of students to create a self-weighting nationally representative subsample. The missing data are thus not completely at random but at random conditional on race, SES background, school type, and region (the basis on which students were selected). In this context, multiple imputation is particular effective at reducing bias and increasing power and outperforms other methods (Krause et al. 2020; Smith, Morgan, and Moody 2022). A detailed discussion of the imputation model can be found in Appendix A2.

Measurement

Socioeconomic status. For the main analyses, I use a pre-constructed socioeconomic status (SES) score on a z-scale, which is based on a principal component analysis using as input parental education, parental occupation, parental income, and whether a student receives free lunch. This information is collected from a parent during the in-home interviews. For students without in-home interviews, the SES score is predicted. Note that the imputation model includes variables that are highly predictive of this score (e.g., parental education and occupation as reported by the student during the in-school interview). To compare the different dimensions of SES background, I separately analyze parental education (Census educational groups), parental occupation (Census occupational groups), and parental income (in 1994 dollars).

Race. I use a pre-constructed variable that categorizes students into White, Black, Native American, Asian, and Hispanic based on student, parent, and interviewer reports. I also differentiate 15 ethnoracial groups (European, African, Native, Chinese, Filipino, Japanese, Indian, Korean, Vietnamese, other Asian, Mexican, Cuban, Puerto Rican, Central & South American, and other Hispanic) to examine the degree to which ethnoracial homophily underlies racial homophily.

GPA. I calculate the average GPA across English, Math, Social Science, and Science. For the most part, grades are reported by the students themselves. For 11 percent of students, I could take this information directly from their transcripts.

Residential proximity. I measure students' residential proximity by the inverse of the log-Euclidean distance (in meters) between students' residential locations. For students without in-home interviews, residential proximity is predicted. Note that the imputation model includes variables that are highly predictive of residential proximity (e.g., residential location of friends).

Overlap courses. Students' course selections are based on their transcripts collected in AHAA. AHAA measures course overlap of pairs of students as the number of courses taken together, weighted by course size and contact time (Muller et al. 2007). This score operationalizes the degree to which students share a social space in school by virtue of their patterns of course participation. For students without transcript information, course overlap is predicted. Note that the imputation model includes variables that are highly predictive of course overlap (e.g., academic performance).

Overlap extracurricular activities. During the in-school interview, students indicated their participation in a list of 20 possible school clubs (e.g., band) and 11 sports teams (e.g., basketball). I measure the degree of overlap of pairs of students as the number of extracurricular activities in which they participate together.

Friendship networks. Based on students' nominations of their five closest male and five closest female friends, I create friendship networks of directed ties in each school. Students were also asked to describe their interactions with friends: if they hang out, discuss problems, talk on the phone, spend time together on the weekend, or go to each other's homes. I use this information to define student i a friend of j if i nominated j as a friend and also reported that they interact with j in at least one of these ways. This definition of friendship therefore goes beyond mere nominations. Further, I differentiate whether student i or j is reporting the friendship to identify friendship asymmetries. While asymmetric ties can reflect measurement error (e.g., nomination limits in the survey instrument) (Lee and Butts 2018), they also indicate asymmetries in friendship definition and strength (An and McConnell 2015; Ball and Newman 2013; Gould 2002; Kitts and Leal 2021). In the Add Health networks, asymmetric ties are unlikely to reflect nomination limits since 60 percent of all ties are unreciprocated but only 3 percent of students nominated 10 friends and only 24 percent hit the constraint on one of the genders. Instead, they are likely to reflect asymmetries in friendship definition and strength, especially because I only measure friendships where students report spending time together. Such asymmetries are interesting in the context of socioeconomic segregation as they shed light on whether all socioeconomic groups display social closure, or whether only some groups exclude and only some groups are being excluded.

Socioeconomic segregation. I measure socioeconomic segregation at a dyadic level to examine socioeconomic mixing patterns in greater detail than is possible with aggregate network segregation indices (Bojanowski and Corten 2014). To do so, I first divide students globally (i.e., across schools) into low-, medium-, and high-SES at the tercile cut-off points of the continuous SES score and subsequently group friendships with respect to these categories (i.e., low→low, low→med, low→high, etc.).[5] I call this the socioeconomic mixing matrix. I then measure socioeconomic segregation as the surplus of same-SES ties (e.g., low→low) and the absence of cross-SES ties (e.g., low→high) in a network relative to a network in which students select friends at random. Akin to a χ^2 test, I calculate $S_t = \left(\pi_t^{obs} - \pi_t^{exp}\right)/\pi_t^{exp} \cdot 100$ for each combination t in the socioeconomic mixing matrix (where π_t^{exp} is the expected and π_t^{obs} is the observed proportion). The segregation statistic S thus measures in percent how many more or fewer ties of each type we observe than expected if students selected friends at random[6]. S ranges from 0 to infinity.[7]

Figure 2: The socioeconomic mixing matrix.

$$M_{SES} = \begin{array}{cc} & alter \\ ego & \begin{bmatrix} low{\to}low & low{\to}med & low{\to}high \\ med{\to}low & med{\to}med & med{\to}high \\ high{\to}low & high{\to}med & high{\to}high \end{bmatrix} \end{array}$$

Descriptive statistics and a more detailed summary of all employed variables can be found in Appendix A1.

Methodological approach

I draw on exponential random graph models (ERGMs) for the analyses in this paper. In contrast to dyadic network models, which disregard fundamental properties of networks by assuming the conditional independence across dyads, ERGMs explicitly model dependencies among dyads that result from network embedding (Lusher, Koskinen, and Robins 2012). Given a network Y with y_{ij} ties connecting actors i and j, ERGM estimates the probability of observing Y as a function of exogenous features and endogenous graph statistics. ERGM has the following probability mass function: $\Pr\left(Y = y|\theta, g(x,y)\right) = \frac{\exp(\theta^T g(x,y))}{\kappa(\theta)}$, where θ is the parameter vector, $g(x, y)$ is a matrix of exogeneous characteristics x and endogenous graph statistics computed on y, and $\kappa(\theta) = \sum \exp\left(\theta^T g(x, y)\right)$ is a normalizing constant ensuring that the sum of equation (1) over all possible networks equals 1. Due to the intractability of $\kappa(\theta)$, the denominator is typically approximated using Markov chain Monte Carlo (MCMC) sampling.

To model the ensemble of networks across schools, prior studies employ a two-step approach in which researchers first estimate separate models per school and then average model coefficients in a meta-analytic approach (An 2015; Lubbers 2003). This method has come under criticism because the resulting average model coefficients may be inaccurate for two reasons. First, as networks vary by size, researchers using the two-step approach are forced to discard small networks when fitting more realistic models, overfit models to small networks, or specify parsimonious models that can be fit on all networks but may omit important tie-formation mechanisms. To the extent that coefficients reflect only a subset of networks, are poorly estimated, or biased, their averages will not accurately approximate average tendencies in the networks (Slaughter and Koehly 2016; Tolochko and Boomgaarden 2024). Second, ERGM coefficients –like in any other nonlinear regression model – are confounded with residual variation and thus is only identified up to a scale factor (Duxbury 2021; Duxbury and Wertsching 2023). Pooling parameters across models is only valid if we can assume that the residual variation is constant across schools, which is unlikely to hold in practice.

Multilevel ERGM make it now possible to resolve these issues by jointly estimating model parameters across networks (Krivitsky et al. 2023; Lazega and Snijders 2016; Slaughter and Koehly 2016; Stewart et al. 2019). I take the simplest multilevel approach and model the ensemble of networks $Y_1, Y_2, \ldots, Y_n$ by jointly estimating model parameters across networks while letting exogenous features and endogenous graph statistics vary arbitrarily between them. This pooling of information is crucial for the purposes of this paper. It allows me to estimate a friendship formation model complex enough to capture the most important tie-formation mechanisms shaping socioeconomic segregation. While simple network models are invaluable to test the existence of tie-formation mechanisms, more complex models are necessary to examine their interplay and impact on network structure characteristics, such as segregation patterns. In particular, interactions among tie-formation mechanisms must be considered to avoid results that are artifacts of restrictive functional forms (Block 2015, 2018; Grund and Densley 2015). Since network density varies across schools, I preserve schools' network densities by constraining the outdegree distribution of each network to the observed one during estimation and simulation.

Recent advances in simulation approaches to evaluate micro-macro linkages show that ERGM can be used to study how local tie-formation mechanisms locally shape network structures globally (Duxbury 2023; Robins et al. 2005; Snijders and Steglich 2015). These approaches shift away from significance testing of model coefficients towards predictive modeling. Following Duxbury (2023), I use ERGMs to simulate networks and then measure in the synthetic networks how the modeled tie-formation mechanisms shape socioeconomic segregation. By constraining the simulation space to networks that exhibit the same outdegree distribution as the observed networks, I ensure comparability across simulation conditions and render the results independent of Add Health's outdegree limitation. While ERGM has not yet been extended to integrate survey weights during estimation, the chosen simulation approach allows me to use survey weights post-estimation to reflect that dyads differ in their inclusion probability in the sample.

I fit all ERGMs using MCMC maximum likelihood estimation with a burn-in of 200,000, a MCMC sample size of 500,000, a thinning interval of 3,000, and the Hummel termination criterion. Convergence statistics suggest that the models have converged, and goodness-of-fit statistics indicate that they fit the observed network structure quite well. See Appendices C3-6, for more details on model estimation and fit.

Analysis 1: Dividing socioeconomic segregation in friendship networks into within- and between-school components

To divide socioeconomic segregation in friendship networks into within- and between-school components, I draw on sample 1 and use ERGM to generate two types of random networks. In the first model (M1), students select friends at random and are not confined to their school but can nominate anyone in the sample. The synthetic networks generated by M1 are thus void of disparities in the SES composition between schools and friending biases within schools. In Bernoulli random graphs, the expected proportion of each tie combination in the socioeconomic mixing matrix equals their unconditional probability, which is $E(\pi) = (1/3)^2$ since students are separated into three equal-sized groups. Rather than Bernoulli random graphs, however, I generate random networks with the same outdegree distribution as the observed network. These random networks provide more realistic baselines and ensure that differences between expected and observed mixing patterns are not artifacts of Add Health's outdegree limitation. Due to this constraint, the expected mixing pattern deviates somewhat from 11.1 percent.

In the other model (M2), students select friends at random but only within their school. The synthetic networks generated by M2 are thus void of friending biases within schools but retain disparities in the SES composition between schools. Based on these two null models, the proportion of each type of tie t in the socioeconomic mixing matrix can be divided into within- and between-school components:

$$\hat{S}_t^{\text{total}} = \left(\pi_t^{\text{obs}} - \hat{\pi}_t^{\text{M1}}\right)/\hat{\pi}_t^{\text{M1}} \cdot 100$$
$$\hat{S}_t^{\text{between}} = \left(\pi_t^{\text{M2}} - \hat{\pi}_t^{\text{M1}}\right)/\hat{\pi}_t^{\text{M1}} \cdot 100$$
$$\hat{S}_t^{\text{within}} = \hat{S}_t^{\text{total}} - \hat{S}_t^{\text{between}}$$

I repeat the simulation process 1,000,000 times, base it on all 10 imputations (100,000 per imputation), and draw simulation coefficients from a multivariate normal distribution based on the ERGM parameter and variance estimates. In this way, I create null distributions that reflect the inherent randomness of friendship formation, as well as sampling, estimation, and imputation uncertainty. I ensure that the calculated proportions are representative of high school friendship networks in 1995 by weighing synthetic and observed friendship ties by the product of two friends' sample inclusion probabilities.

Analysis 2: Parsing the determinants of socioeconomic segregation within schools

To parse the determinants of socioeconomic segregation within schools, I fit another model (M3) using sample 2 in which students select friends based on the hypothesized mechanisms. Subsequently, I generate synthetic networks from M3 in which the estimated conditional effect of each mechanism k on friendship formation, $\hat{\theta}_k$, is turned on, one by one. The impact of k on socioeconomic segregation then is the change in the proportion of each type of tie in the socioeconomic mixing matrix, $\hat{\pi}_t$, measured in the synthetic networks when the mechanism is turned on: $\hat{\tau}_{t,\theta_k} = \hat{\pi}^{M3}_{t,\theta_k=\hat{\theta}_k} - \hat{\pi}^{M3}_{t,\theta_k=0}$. Duxbury (2023) calls these in silico experiments "micro effects on the macro structure". I repeat the simulation process 1,000,000 times, base it on all 10 imputations (100,000 per imputation), and draw simulation coefficients from a multivariate normal distribution based on the ERGM parameter and variance estimates. In this way, I create distributions of $\hat{\pi}^{M3}_{t,\theta_k}$ that reflect the inherent randomness of friendship formation, as well as sampling, estimation, and imputation uncertainty.

The main friendship formation model (M3) includes stratified settings (neighborhoods, courses, and extracurricular activities), homophilous tendencies (GPA, race, and SES), popularity differences (GPA, race, and SES), and relational mechanisms (preferential attachment, reciprocity, and triadic closure). I also include interactions between race and SES homophily, race and SES popularity, and homophilous tendencies and relational mechanisms. Appendix C3 provides a more detailed exposition and justification of the model specifications and details how the chosen operationalization builds and improves on prior work that separates homophily and popularity mechanisms using mixing matrices (e.g., An 2022; An and McConnell 2015).

I start by turning on the effects of stratified settings to examine how neighborhoods, courses, and extracurricular activities contribute to socioeconomic segregation in the absence of homophilous tendencies, popularity differences, and relational mechanisms. Subsequently, I switch on homophilous tendencies to analyze how homophily on the basis of GPA, race, and SES background within these settings contribute to socioeconomic segregation. Then, I turn on popularity differences by GPA, race, and SES background to examine the extent to which segregation is the result of an unequal rejection of outgroup members rather than an equal one (i.e., homophily). Finally, I switch on relational mechanisms to analyze if they amplify the existing socioeconomic segregation.

While I chose this order carefully, Appendix C5 shows that the impact of each mechanism on socioeconomic segregation does not depend much on the ordering. In other words, $\hat{\tau}_{t,\theta_k} + \hat{\tau}_{t,\theta_l} \approx \tau_{t,\theta_k+\theta_l}$ for any order and $\sum \tau_{t,\theta_k} \approx \hat{\pi}^{M3,full}_{t} - \hat{\pi}^{M3,empty}_{t}$. The reason is that I switch on conditional effects estimated in a single model rather than unconditional effects estimated in a sequence of models. An exception is relational mechanisms; their impact on socioeconomic segregation depends on the existing segregation in the network, which is why I add them last.[8] As with any other ceteris paribus approach, changing a single model parameter while holding the other parameters constant is unrealistic. Simulation-based thought experiments provide intuitions about relative effect sizes, but they cannot predict what will happen in reality.

RESULTS

Degree and patterns of socioeconomic segregation in high school friendship networks

Friendship networks in high schools are segregated along multiple dimensions. The friendship network in Figure 3, for instance, shows signs of clustering by gender, race, and SES background. Therefore, I first compare socioeconomic segregation to segregation on the basis of other demographic categories. The results are reported in Table 1.

The grey-shaded row provides the overall segregation S for each attribute, comparing the observed proportion of in-group ties to the expected proportion if the attribute was uniformly distributed between schools and students selected friends at random within schools. To give an example, 95 percent of random networks exhibit a proportion of same-SES friendships between 33.7 and 34.6 percent. Since the observed 39.2 percent lies outside of this interval, the socioeconomic segregation in high schools is systematic: friendship ties between students with the same SES background are $S = 15$ percent more likely than if students selected friends at random.

The results in Table 1 show that the socioeconomic segregation in friendship networks is less pronounced than segregation based on race (+54%), gender (+43%), and GPA (+31%), but more pronounced than the segregation based on the language spoken at home (+3%). Comparing the different indicators of students' SES background, I find that friendship networks are more segregated in terms of parental education (+17%) and parental occupation (+15%) than in terms of parental income (+12%). The evidence of socioeconomic segregation notwithstanding, these results show that there is quite a bit of socioeconomic mixing in high school: about 60 percent of all friendship ties cross socioeconomic lines.

In Table 1, continuous attributes were divided into low, medium, and high categories at the tercile cutoff points and collapsed into a dichotomous indicator of same- vs. cross-group ties, which can obscure important heterogeneities in how groups relate to each other. To understand whether interaction patterns differ between different points in the SES distribution, in Figure 4, I distinguish all tie combinations in the socioeconomic mixing matrix. The figure shows that low-SES students do not select friends based on their SES background. They nominate low-SES students 5 percent more often, medium-SES students 2 percent more often, and high-SES students 8 percent less often than expected in random graphs with the same outdegree distribution. They also nominate fewer friends since the observed proportions all fall below 11.1 percent, which is what we would expect if outdegree did not differ by SES background. The situation differs for medium-SES students. They nominate low-SES students 14 percent less often, medium-SES students 2 percent more often, and high-SES students 12 percent more often than expected. Most clearly, high-SES students select friends based on their SES background. They nominate high-SES students 34 percent more often, medium-SES students 2 percent less often, and low-SES students 32 percent less often than expected. Both medium- and high-SES students nominate low-SES peers as friends less often and high-SES peers as friends more often than expected. Socioeconomic segregation in high school friendship networks is therefore characterized by unilateral exclusion of students in the bottom third of the SES distribution and closure among students in upper half of the SES distribution. Robustness checks in which I differentiate the SES score into more than three categories, choose different percentiles as cutoff points, examine the continuous SES score, and analyze reciprocated ties clearly confirm this pattern (see Appendix B).

Differentiating interaction patterns between different points in the SES distribution also reveals that socioeconomic segregation is more pronounced than is apparent in Table 1. In particular, the boundary between low- and high-SES students is bright. Much prior work collapses socioeconomic mixing into a dichotomous indicator of same- vs. cross-SES ties (e.g., Malacarne 2017) or a linear measure of SES distance between peers (e.g., McMillan 2022). These operationalizations mask the unilateral and nonlinear segregation pattern to some extent (see Appendix B), which explains why prior studies find less evidence for socioeconomic segregation in the Add Health networks. Chetty et al. (2022b) also identify a greater degree of socioeconomic segregation. To understand whether the data (Facebook friendships) may be responsible for the divergent conclusions, I replicate their analysis on the Add Health networks.[9] I find that friendship networks in the Add Health schools are more segregated than the friendship networks on Facebook. Consequently, measurement and interpretation differences are the reason why this study and Chetty et al. (2022b) identify a greater degree of socioeconomic segregation than prior work.

Determinants: Compositional differences between schools

Having described socioeconomic segregation in friendship networks, next I parse its determinants. I start by decomposing socioeconomic segregation into compositional differences between schools and students' friendship choices within schools.

Table 1 indicates that schools differ in composition much more by race (34%) than by SES (7%). Within schools, too, socioeconomic segregation is low (7%) compared to segregation based on race (20%), academic performance (22%), and gender (39%). While compositional differences between schools are similar across the different indicators of SES background, within schools, friendship networks are segregated more by parental education (8%) and occupation (7%) than by income (3%). Overall, disparities in the SES composition between schools account for about $7/14.5 \cdot 100 \approx 48$ percent of the socioeconomic segregation in friendship networks while students' friendship choices explain the remaining 52 percent. In contrast, 62 percent of the racial segregation can be attributed to compositional differences between schools while students' friendship choices within schools only explain 38 percent.

Figure 3: Segregation patterns in a high school friendship network. *The figure depicts three times the same friendship network in a school in the Northeast. The network shows signs of clustering by gender, race, and SES background. In the displayed connected component, same-gender ties are 20.9%, same-race ties are 24.3%, and same-SES ties are 14.2% more likely than if students selected friends at random.*

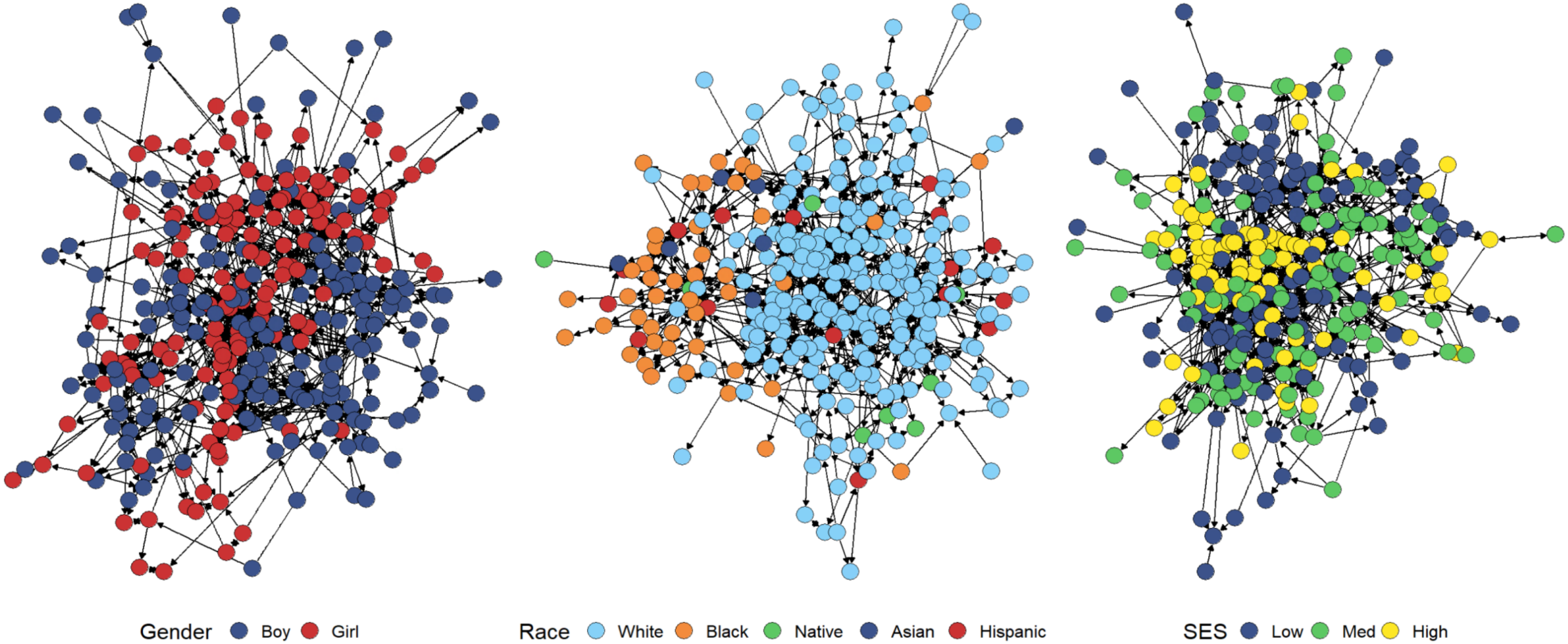


Table 1: Segregation in high school friendship networks by social category. *The table reports the expected proportion of in-group ties if the attribute was uniformly distributed between schools and students selected friends at random within schools ($\hat{\pi}^{M1}$), the expected proportion if students selected friends at random within schools ($\hat{\pi}^{M2}$), and the observed proportion (π^{obs}). Total segregation then equals $\hat{S}_t^{total} = (\pi^{obs} - \hat{\pi}^{M1})/\hat{\pi}^{M1} \cdot 100$, segregation between schools equals $\hat{S}_t^{between} = (\hat{\pi}^{M2} - \hat{\pi}^{M1})/\hat{\pi}^{M1} \cdot 100$, and segregation within schools equals $\hat{S}_t^{within} = \hat{S}_t^{total} - \hat{S}_t^{between}$.*

Interpretation example: We expect 51.9% of all friendships to be same-gender ties if schools did not differ in their gender composition and students selected friends at random. We observe 74.2% of all friendships to be same-gender ties, which is 43% more than expected. Of this segregation, 39.7 percentage points are within schools and 3.3 percentage points are between schools.

Calculation notes: The expected proportions are based on random networks with the same outdegree distribution as the observed networks. The 95% predictive interval reflects the inherent randomness of friendship formation as well as sampling, estimation, and imputation uncertainty. Continuous attributes are categorized into low, medium, and high at the tercile cut-off points.

						Indicators of SES		
	Gender	**Race**	**Language**	**GPA**	**SES**	**Parental education**	**Parental occupation**	**Parental income**
M1: Expected	51.9	50.6	85.2	35.0	34.2	33.6	41.4	35.1
95% PI	51.4, 52.3	50.3, 50.9	84.7, 85.8	34.6, 35.5	33.7, 34.6	33.1, 34.0	41.0, 41.9	34.3, 35.6
M2: Expected within	53.6	64.7	86.4	37.8	36.5	36.7	44.7	37.9
95% PI	52.9, 56.2	64.2, 66.1	86.0, 87.0	37.3, 38.7	36.0, 37.1	36.2, 37.2	44.2, 45.2	37.1, 38.6
Observed	74.2	77.9	87.6	46.1	39.2	39.3	47.7	39.2
Segregation S (in %)	43.0	53.8	2.8	31.6	14.5	17.1	15.0	11.8
- within schools	39.7	20.4	1.4	23.6	7.5	7.8	7.3	3.6
- between schools	3.3	33.4	1.4	8.0	7.0	9.3	7.7	8.2

Figure 4: Socioeconomic segregation by tie combination in the socioeconomic mixing matrix. *The figure differentiates segregation by tie combination in the socioeconomic mixing matrix. The bars are based on the proportion of same- and cross-SES ties t expected if SES was uniformly distributed between schools and students selected friends at random within schools* ($\hat{\pi}_t^{M1}$)*, students selected friends at random within schools* ($\hat{\pi}_t^{M2}$)*, and the observed proportion* (π_t^{obs})*. Total segregation then equals* $\hat{S}_t^{total} = \left(\pi_t^{obs} - \hat{\pi}_t^{M1}\right)/\hat{\pi}_t^{M1} \cdot 100$*, segregation between schools equals* $\hat{S}_t^{between} = \left(\pi_t^{M2} - \hat{\pi}_t^{M1}\right)/\hat{\pi}_t^{M1} \cdot 100$*, and segregation within schools equals* $\hat{S}_t^{within} = \hat{S}_t^{total} - \hat{S}_t^{between}$*.* $\hat{\pi}_t^{M1}$*,* $\hat{\pi}_t^{M2}$*, and* $\hat{\pi}_t^{obs}$ *can be found in Appendix B.*

Interpretation example: We observe 34% more friendship ties among high-SES students than expected. Of this segregation, 18 percentage points are within schools and 16 percentage points are between schools.

Calculation notes: The expected proportions are based on random networks with the same outdegree distribution as the observed networks. The 95% predictive interval reflects the inherent randomness of friendship formation as well as sampling, estimation, and imputation uncertainty. To differentiate tie combinations, students are categorized into low-, med-, and high-SES at the tercile cut-off points of the continuous SES score. Differences in the between-school effect by tie direction are due to outdegree differences by SES background.

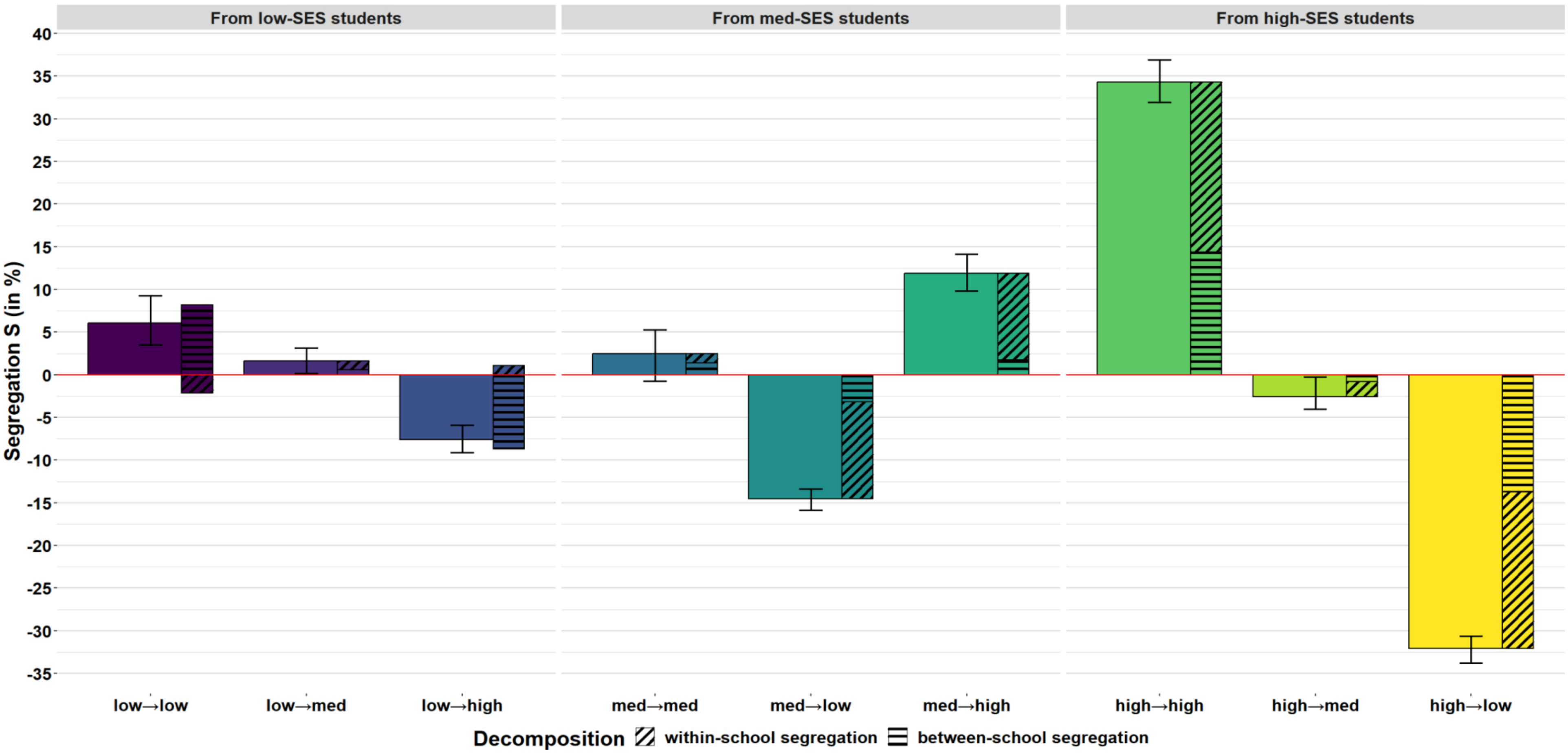


Figure 4 reveals again that the dichotomy between same- and cross-SES friendship conceals important differences across the SES distribution. The fact that low-SES students nominate other low-SES students more often and high-SES students less often than expected is largely driven by meeting opportunities. They do not select friends based on their SES background. The opposite is the case for medium-SES students. Since they attend schools both with low- and high-SES students, medium-SES students have the most opportunities to befriend students with different SES backgrounds. Their friendship choices thus largely explain why high-SES peers are overrepresented and low-SES peers are underrepresented among their friends. The disconnect of high-SES students from their low-SES peers is driven both by meeting opportunities (40%) and their friendship choices (60%).

Figure 4 also shows that the identified key pattern of segregation – the exclusion of students at the bottom and closure among students in the upper part of the SES distribution – is pronounced and determined in important ways by students' friendship choices: med→low and high→low ties are 11 and 18 percent less likely and med→high and high→high ties are 10 and 19 percent more likely than if students selected friends at random within schools. The fact that high→low ties are about as unlikely (-18%) as interracial ties on average (-20%) speaks to the strength of this pattern within schools. There is an important difference between racial and socioeconomic segregation, however. In Figure 4, the difference in likelihood between higher→lower ties and lower→higher ties equals 15.7 percentage points on average across combinations. The rejection of cross-SES ties is therefore asymmetric. The same differences in the racial mixing matrix (e.g., White→Black and Black→White ties), in contrast, equals less than 1 percent on average across tie combinations. Therefore, while racial segregation is marked by mutual exclusion, socioeconomic segregation is characterized by unilateral exclusion of low-SES students from higher-SES cliques.

Figure 5: ERGM coefficient plot. *The figure provides average marginal effects (AMEs) (Duxbury and Wertsching 2023). AMEs measure the average change in tie probability as a variable increases by one unit. Both point estimates and one (inner bar) and two times (outer bar) the standard error are displayed. AMEs are on a probability scale, which allows for effect size comparisons. Effects should be interpreted relative to the baseline tie probability, which is 0.01% on average across networks. Coefficients have been squished into a -0.5-0.5 range (AME of course overlap and reciprocity are greater than 0.5). See Appendix C4 for full results. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, + p<0.01.*
<u>*Interpretation example*</u>*: The popularity of high-SES students increases the baseline tie probability of ties to them by 0.1 percentage points (from 0.01% to 0.02%).*

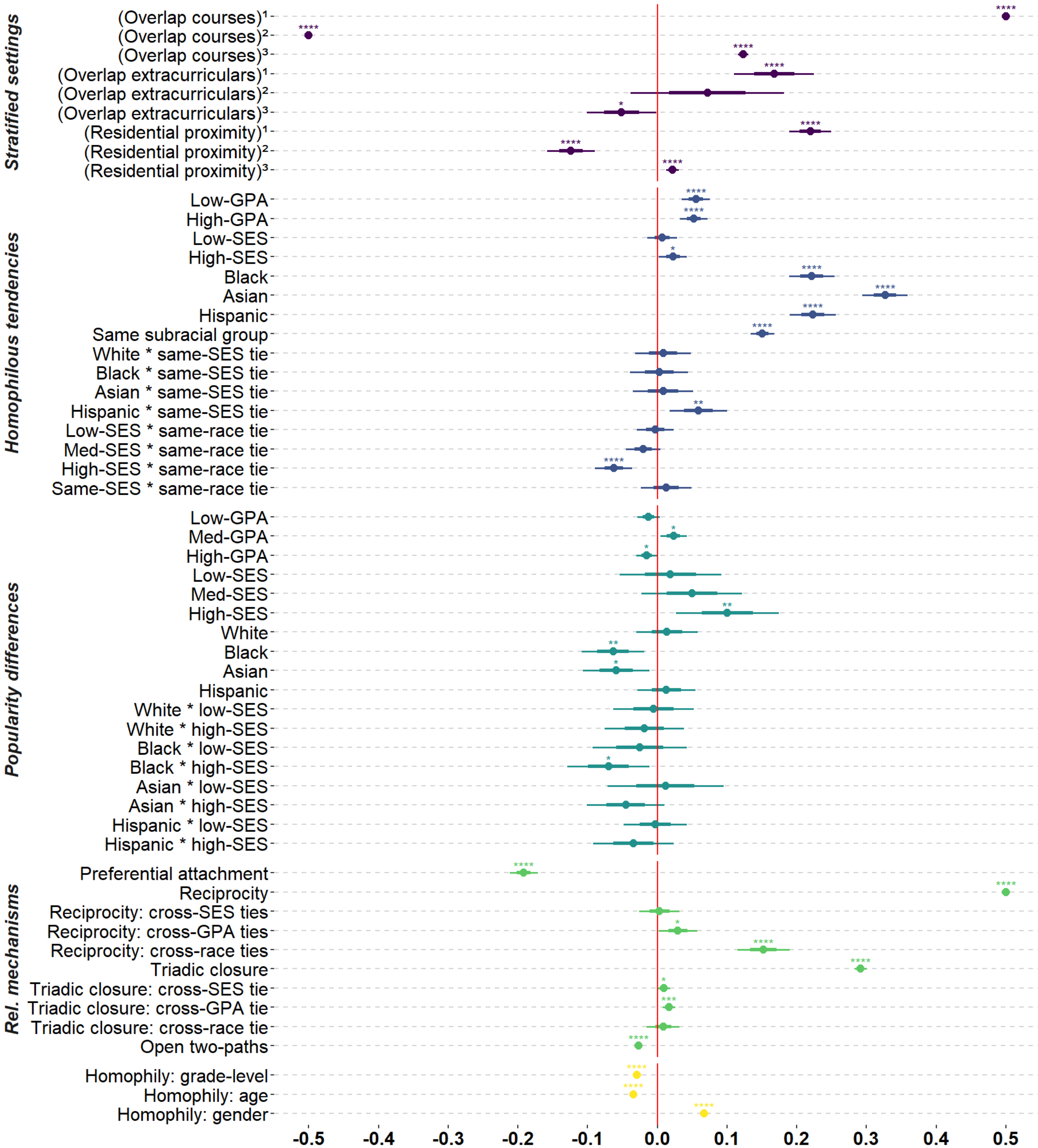
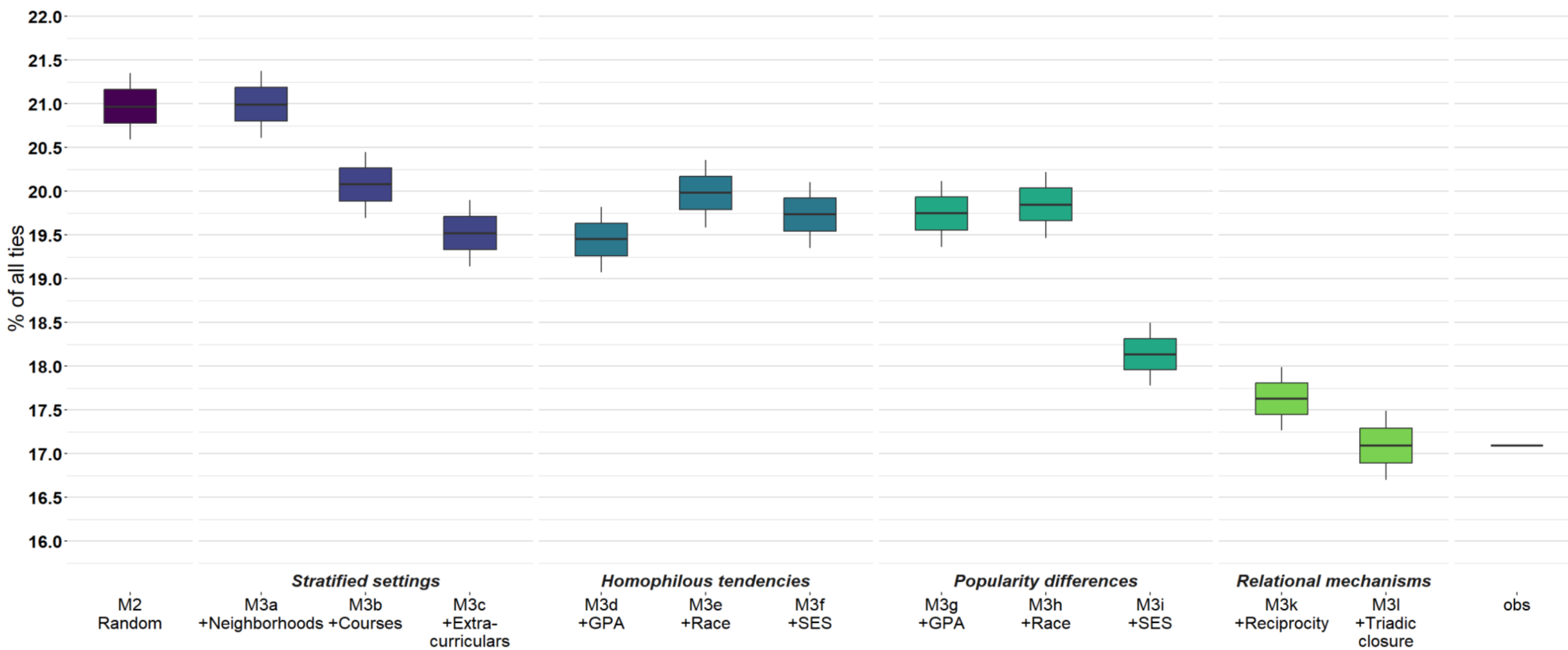


Figure 6: Determinants of the exclusion of low-SES students. *The figure displays the predicted proportion of ties to low-SES students when key determinants of socioeconomic segregation are turned on one by one. The predicted proportions are presented with boxplots that display prediction uncertainty due to sampling, estimation, imputation, and the inherent randomness in friendship formation. M2 predicts the proportion of high→low ties and med→low ties if students selected friends at random within schools. M3a-l then predict how key determinants of socioeconomic segregation contribute to the absence of ties to low-SES students. Below the graph, τ describes the impact of each mechanism on the predicted proportion (in percentage points), and % reports the impact of each mechanism as a share of the total gap between expected and observed proportion (in %).*

<u>Interpretation example:</u> The predicted proportion of ties to low-SES students if students selected friends at random equals 21%. Turning on the course effect on friendship formation decreases the predicted proportion by 0.9 percentage points to 20.1%. The proximity effect of taking the same courses thus explains 23.5% of the total gap between expected and observed proportion (0.9/(17.1-21)100≈-23.5).

<u>Calculation notes</u>: The 95% predictive interval reflects the inherent randomness of friendship formation as well as sampling, estimation, and imputation uncertainty. To differentiate tie combinations, students are categorized into low-, med-, and high-SES at the tercile cut-off points of the continuous SES score.



	M2	M3a	M3b	M3c	M3d	M3e	M3f	M3g	M3h	M3i	M3k	M3l	obs
τ		-0.02	0.91	0.56	0.07	-0.54	0.25	-0.01	-0.1	1.71	0.51	0.54	0
%		-0.6	23.5	14.4	1.9	-13.8	6.4	-0.3	-2.6	44.2	13.1	13.8	0

Figure 7: Determinants of socioeconomic segregation overall. *The figure displays how key determinants shape socioeconomic segregation overall by summing up the effect of each mechanism across tie combinations in the socioeconomic mixing matrix. Socioeconomic segregation in friendship networks is characterized, on the one hand, by an abundance of ties between students with the same SES background, and, on the other hand, by an absence of ties between students with different SES backgrounds. Accordingly, in the figure, the effect of each mechanism is displayed as segregation-increasing if it adds to the proportion of same-SES ties or takes from the proportion of cross-SES ties and vice versa.*
Interpretation example: The observed friendship networks include more same-SES ties and fewer cross-SES ties than random networks (+7.6pp in total). M3 predicts that students' course selections contribute 2.03pp to the total socioeconomic segregation by increasing same-SES ties (e.g., high→high: +0.6pp) and reducing some cross-SES ties (e.g., high→low:-0.6pp). Courses also integrate networks by increasing high→med ties but this integrative effect amounts to less than 0.3pp. Overall, students' course-taking patterns explain 26.8% of the total socioeconomic segregation in students' friendship networks.
Calculation notes: The 95% predictive interval reflects the inherent randomness of friendship formation as well as sampling, estimation, and imputation uncertainty. To differentiate tie combinations, students are categorized into low-, med-, and high-SES at the tercile cut-off points of the continuous SES score.

Determinants: Stratified settings, homophilous tendencies, popularity differences, and relational mechanisms within schools

The preceding sections made clear that socioeconomic segregation in friendship networks is not only determined by disparities in the SES composition between schools but also by students' friendship choices within schools. While these choices induce exclusion of students at the bottom and closure among students in upper part of the SES distribution, they do not have to reflect preferences for exclusion and closure. Socioeconomic segregation in friendship networks may just be an unintended consequence of mechanisms correlated with such preferences. Next, I therefore disentangle four key mechanisms that shape socioeconomic segregation in friendship networks within schools: stratified settings, homophilous tendencies, popularity differences, and relational mechanisms.

The results are presented in three figures. Figure 5 shows the ERGM coefficients – the conditional effects of the four key mechanisms on friendship formation. Figure 6 depicts the consequences of these tie-formation mechanisms for the exclusion of low-SES students. I focus on the exclusion of students at the bottom rather than closure among students in upper part of the SES distribution because the results largely mirror each other. Finally, Figure 7 displays the impact of each mechanism on socioeconomic segregation overall by aggregating the results across all tie combinations in the socioeconomic mixing matrix.

Exclusion of low-SES students

Figure 6 presents why low-SES students are underrepresented in the friendship networks of their peers with higher SES backgrounds. M2 predicts that we would expect 21 percent of all friendships to be med→low and high→low ties if students selected friends at random. Since we only observe 17.1 percent, a fifth of the possible ties of this type are unrealized (3.9 percentage points).[10]

Models M3a-l measure the extent to which stratified settings, homophilous tendencies, popularity differences, and relational mechanisms contribute to the exclusion of low-SES students. M3a-c indicates that stratified settings together explain about 38 percent of the gap between expected and observed proportions. This impact is driven by stratified courses and extracurricular activities. The neighborhood effect on friendship formation (M3a) has no appreciable impact on the exclusion of low-SES students as neighborhoods are overall as segregated as the schools they feed. In contrast, the separation of students into different courses (M3b) and extracurricular activities accounts for 24 percent and 14 percent, respectively.

M3d-f indicate that homophilous tendencies together *reduce* the exclusion of low-SES students by about 6 percent. This impact is driven primarily by racial and subracial[11] homophily (M3e). The friendship formation model (see coefficient plot in Figure 5) indicates that minority students display more racial homophily while high-SES students display less racial homophily. Together these effects increase the number of med→low and high→low ties, reducing the gap between expected and observed proportions by 14 percent. In contrast, GPA homophily (M3d) contributes to the exclusion of low-SES students but its impact is small. SES homophily (M3f) too contributes only marginally to the exclusion of low-SES students. Therefore, the exclusion of low-SES students is *not* driven by medium- and high-SES peers wanting to be among their own kind and rejecting all other students equally.

Instead, M3g-i indicate that popularity differences explain the exclusion of low-SES students, which together account for about 41 percent of the gap between expected and observed proportions. Rather than popularity differences by academic performance (M3g) or race (M3h), however, it is popularity differences by SES background that explain the effect. The friendship formation model reveals that high-SES students are more popular than other students, which nudges medium- and high-SES students to befriend high-SES instead of low-SES peers.

Lastly, M3j-l indicate that relational mechanisms also contribute to the exclusion of low-SES students, together accounting for about 27 percent. Both students' tendency to reciprocate friendship ties (M3k) and their tendency to close open triads (M3l) produce fewer ties to low-SES students.

To conclude, low-SES students are underrepresented in the friendship networks of their higher-SES peers because of basic clique formation mechanisms (+27%), stratified courses and extracurricular activities (+38%), and because both medium- and high-SES students prefer high-SES peers as friends (+44%). In contrast, racial homophily reduces the exclusion of low-SES students (-14%).

Socioeconomic segregation overall

Med→low and high→low ties constitute only a subset of tie combinations in the socio-economic mixing matrix. In a final step, I therefore aggregate the effect of each mechanism across all tie combinations to examine their overall impact on socioeconomic segregation within schools. In Figure 7, the overall impact of each mechanism is plotted as a stacked bar chart. Socioeconomic segregation in friendship networks is characterized by an overrepresentation of ties between students with the same SES background and an underrepresentation of ties between students with a different SES background. Accordingly, in the figure, the effect of each mechanism on the predicted proportion by tie type is displayed to increase segregation if it increases the proportion of same-SES ties or reduces the proportion of cross-SES ties and vice versa. Compared to analyzing aggregate segregation indices, this approach displays the overall impact of each mechanism without obscuring effect heterogeneities across the SES distribution.

The aggregate analysis shows that stratified settings account for about 27 percent of the total socioeconomic segregation within schools. The main structural constraint to integrated friendship networks are stratified courses. In contrast, extracurricular activities have an ambivalent impact on socioeconomic segregation. They nudge participating low-SES students to befriend high-SES students. At the same time, however, they prevent medium- and high-SES students from befriending nonparticipating low-SES peers. These countervailing effects largely offset each other.

Student preferences account for about 75 percent of the total socioeconomic segregation within schools. Students' preferences regarding peers' racial and socioeconomic background rather than their academic performance explain this effect. Despite contributing to socioeconomic segregation overall, students' preferences have countervailing effects on segregation, however. For example, students' racial preferences induce closure among low-SES students and among high-SES students, which explains about 68 percent of the socioeconomic segregation in friendship networks. At the same time, however, racial homophily increases med→low ties within racial groups. Students' socioeconomic preferences have an even more ambivalent impact on socioeconomic segregation. Residual SES homophily induces closure among high-

SES students, which reduces in particular high→med ties. However, the relative unpopularity of low-SES peers offsets this reduction again as high-SES students want to befriend medium-SES peers rather than low-SES peers.

In general, the segregating impact of SES homophily is largely offset by the integrating impact of popularity differences by SES background. Students' socioeconomic preferences still contribute to socioeconomic segregation (5%) because popularity differences by SES background increase some cross-SES ties (low→med, low→high, and med→high) but not others (med→low and high→low). In fact, the ambivalent impact of popularity differences by SES background is what creates the asymmetry that is characteristic of socioeconomic segregation in high school: many ties from low-SES students to peers with higher SES backgrounds are unreciprocated. The fact that both medium-SES and high-SES students want to befriend high-SES rather than low-SES peers reflects aspirational and homophilous tendencies, respectively. Students' socioeconomic preferences differ in this way from their racial preferences since racial preferences are dominated by homophilous tendencies (see Figure 5).

Finally, relational mechanisms together do not contribute much to socioeconomic segregation within schools as their effects are small and both in- and decrease cross-SES friendships. The underlying reason is that unreciprocated friendships and open triads are more likely between students with attribute combinations that inhibit tie formation. Reciprocation and balance norms can bridge such barriers, but the results show that they succeed more readily from below: reciprocity and triadic closure increase low→high and med→high ties but reduce high→low and high→med ties. While, for reciprocity, the segregating effect dominates (+4%), for triadic closure, the integrating effect prevails (-5%) because open triads close more frequently between dissimilar students (see Figure 5). Because these effects offset each other, relational mechanisms barely affect socioeconomic segregation in friendship networks.

To conclude, socioeconomic segregation in high school friendship networks is largely driven by stratified courses (+27%) and racial homophily (64%). While other mechanisms also affect socioeconomic segregation, their impacts are small, ambivalent, and offset each other.

LIMITATIONS

This study is not without limitations. First, I had to limit myself to the first wave of Add Health to analyze complete friendship networks. Only at this wave were all students in each school asked to nominate their closest friends. While this cross-sectional snapshot is rich thanks to Add Health's unparalleled data collection, it cannot capture dynamic aspects of friendship formation. The cross-sectional and observational nature of the data makes it further impossible to establish a causal direction, rendering the analysis entirely descriptive (An, Beauvile, and Rosche 2022). To give an example, the data does not allow me to discriminate whether students became friends because they took the same courses or whether they took the same courses because they are friends. Forward-looking students may even take specific courses to befriend specific peers. Therefore, the association between students' course selections and socioeconomic segregation in their friendship networks measures the degree to which segregation is channeled through courses. It cannot fully disentangle whether courses constitute exogenous structural constraints or endogenous paths through which students realize a preference to be among socioeconomically similar peers. Such a preference could drive the impact of extracurricular activities on socioeconomic segregation to some extent since students choose sports teams and clubs at their discretion. In contrast, the impact of course selections and neighborhoods is more likely to reflect how structural barriers shape socioeconomic segregation since students have less control over the courses they take and little control over where they live. Overall, however, the measured impact of stratified settings in high school likely underestimates the true impact of structural constraints on socioeconomic segregation. The reason is that this analysis does not capture how stratified settings in primary and middle school shape socioeconomic segregation in friendship networks that then carries over to high school. What the analysis does capture is the degree to which socioeconomic segregation in high school is associated with socioeconomic segregation in neighborhoods, courses, and extracurricular activities.

Second, SES homophily and aspiration are residual tendencies and thus inferred, not stated preferences. They measure the excess probabilities of same- and cross-SES ties, net of the effect of residential proximity, course and extracurricular selections, homophily and aspiration on correlates of SES, and relational mechanisms. Even though I took great care to include the most important drivers of socioeconomic segregation in the analysis, correlated unobserved mechanisms will also be captured by these terms. For instance, friends who no longer take the same courses and extracurricular activities but are friends because they shared social settings before will register as SES homophily. Further, SES homophily may also capture, for instance, parental preferences (Zwier and Geven 2023) or homophily based on personality traits that are correlated with SES background (Hughes et al. 2021). It depends on researchers' definition of SES whether to consider this a confounding or mediating relationship.

Third, Add Health contains many missing data points (but some attributes are more affected than others). I employ multiple imputation to address the issue. While this approach reduces bias and increases power vis-à-vis listwise deletion, it cannot be ruled out that measurement error attenuates effect estimates to some extent. It seems unlikely, however, that the results are significantly impacted by measurement error because missingness is partly design-induced (i.e., mechanisms are known), nonmissing data are measured precisely (e.g., residential proximity is based on exact locations, course selections are sourced from students' transcripts), and because both variables with fewer (e.g., extracurricular activities) and with more missing values (e.g., courses) display robust effects on friendship formation.

Fourth, the Add Health survey dates back to 1995. While more recent high school friendship network data is available (e.g., PROPSER), they are regional samples that preclude demographic analyses of high generalizability like this one. The truth is that Add Health remains the only nationally representative study in the United States that is publicly available and contains sociometric information. Similarities between the results from this study and Chetty et al. (2022b) suggest that socioeconomic segregation in friendship networks may not have changed much over time. Smith et al. (2014) examine change in network segregation between 1985 and 2004 using egocentric network data and find little change over time.

Fifth, this study averages across heterogeneities by school type, composition, and region, to name just a few contextual variables (McFarland et al. 2014). Prior studies show that segregation patterns depend on contextual variables like schools' demographic composition (Moody 2001) or the consolidation among demographic attributes (Zhao 2023). Due to the small sample sizes of 132 schools, however, Add Health is not particularly suited to examine how network segregation depends on the ecological context. As better data become available, the scope conditions of these results should be examined.

DISCUSSION AND CONCLUSION

A growing body of research in sociology and adjacent fields focuses on how networks and embedded social capital affect social inequality (Blau 1977; Lenkewitz 2023; Lin 2000; Simmel 1908; van Tubergen and Volker 2014; Young, Park, and Feng 2024). A core finding of this line of research is that social embeddedness consolidates rather than alleviates inequality when networks are socioeconomically segregated. The determinants of socioeconomic segregation in networks are therefore important antecedents of the link between networks and inequality.

This study examines the mechanisms underlying socioeconomic segregation in friendship networks in high school. It addresses a key puzzle raised in Chetty et al. (2022b), who identify greater SES homophily in high school friendship networks than prior studies: What is behind this friending bias? A well-developed literature in sociology on the determinants of racial segregation provides answers. This research highlights the importance of structural barriers, such as neighborhoods (Mouw and Entwisle 2006), courses (Frank et al. 2013), and extracurricular activities (Schaefer et al. 2018), as well as intentional and unintentional boundary-making between groups with intersecting attributes (Alba 2005; Moody 2001; Wimmer 2013; Zhao 2023).

I integrate these disparate perspectives in this study to disentangle their relative contributions to socioeconomic segregation. Using data from the National Study of Adolescent Health and a new exponential random graph modeling approach, this study provides a deeper understanding of the degree and patterns of socioeconomic segregation and the processes that create it.

In the following, I summarize the results to highlight contributions to the literature and to inform interventions to reduce socioeconomic segregation in friendship networks.

Degree and patterns of socioeconomic segregation

The results of this study show that socioeconomic segregation in high school friendship networks is characterized by unilateral exclusion of students in the bottom third and closure among students in the upper half of the SES distribution. While friendship networks are overall less socioeconomically segregated than they are racially segregated, the exclusion of low-SES students from high-SES cliques is pronounced. High-SES students are 18 percent less likely to befriend low-SES peers than if they selected friends at random, which compares to 20 percent for interracial ties on average. Moreover, while racial segregation is marked by mutual exclusion, socioeconomic segregation is asymmetric in that many friendship ties from low-SES students to peers with higher SES backgrounds remain unreciprocated. This difference between socioeconomic and racial segregation is concealed in studies that examine only reciprocated ties. These results imply that interventions to reduce socioeconomic segregation should focus on the friending behavior of students in the upper half of the SES distribution as their friendship choices drive socioeconomic segregation. Students are also more segregated with respect to parental education and occupation than with respect to parental income. Socioeconomic segregation in high school friendship networks will thus not be captured adequately when students' socioeconomic background is equated with parental income, as is increasingly common (Barone, Hertel, and Smallenbroek 2022).

This study and Chetty et al. (2022b) identify a greater degree of socioeconomic segregation than prior work. Measurement differences are the likely reason for the divergent conclusions. Prior work reduces socioeconomic mixing to dichotomous indicators (e.g., Malacarne 2017) and linear measure of SES distance between peers (e.g., McMillan 2022), which masks the unilateral and nonlinear pattern to some extent. I replicated the analyses by Chetty et al. (2022b) with Add Health and find that friendship networks in the Add Health schools are more segregated than friendship networks on Facebook. This result also points to measurement differences for the divergent conclusions and adds to research showing that offline (core) friendship networks are more segregated than online (extended) friendship networks (DiPrete et al. 2011; Hofstra et al. 2017).

Determinants of socioeconomic segregation

The decomposition analysis identifies that parents' choices on where to live and which school to enroll their children produce disparities in the SES composition between schools that account for about 50 percent of the total segregation. This result is in line with Chetty et al. (2022b), who also attribute 50 percent to exposure differences between schools (based on 80,000,000 Facebook friends). The other half is determined by processes within schools, which Chetty and colleagues subsume under "friending bias" due to their lack of data to disentangle them. In the main analysis of this paper, I move beyond friending bias as a singular concept and disentangle four key determinants of socioeconomic segregation within schools: stratified settings, homophilous tendencies, popularity differences, and relational mechanisms. The results are discussed in the following.

Stratified settings. Stratified courses, extracurricular activities, and neighborhoods account for about 25 percent of the total socioeconomic segregation within schools. Socioeconomic segregation in high school is therefore not only the result of friending bias but also institutional bias in the structures in which students are embedded. The main institutional constraint to integrated friendship networks are stratified courses. Eliminating academic tracking would thus not only promote educational equality (Rui 2009) but likely also increase socioeconomic integration in friendship networks. Prior research speculates that extracurricular activities, such as sports teams, may constitute settings in which students from diverse backgrounds can become friends (Lessard and Juvonen 2019; Malacarne 2017). Chetty et al. (2022b) show that this is not the case for recreational groups outside of school. The results presented here indicate that this is also not the case for extracurricular activities within school. Overall, sports teams and clubs are as stratified as the school itself.[12] Proximity effects on friendship formation can thus only be leveraged to reduce socioeconomic segregation if we enable and encourage disadvantaged youth to take advanced courses and participate in extracurricular activities.[13] Finally, while parents' residential choices shape differences in SES composition between schools, residential proximity does not contribute to socioeconomic segregation in friendship networks within schools.[14]

Student preferences. Students' preference to befriend similar peers (homophilous tendencies) and their preference to befriend peers with greater characteristics than themselves (aspirational tendencies) account for about 75 percent of the total socioeconomic segregation within schools. Note that this study only captures revealed preferences (i.e., residual tendencies net of other modeled mechanisms) since preferences were not directly measured. However, their large impact certainly suggests that socioeconomic segregation in high school is not only driven by structural constraints – as McFarland et al. (2014) suggest – but also by students' agency.

Students' preference to befriend same-race peers make up most of this impact. Therefore, while racial segregation is not a by-product of SES homophily (Moody 2001; Zeng and Xie 2008), socioeconomic segregation in friendship networks is in important ways a by-product of racial homophily. Despite contributing to socioeconomic segregation overall, the effect of racial homophily is ambiguous, however. It induces closure among low-SES students of the same race but, at the same time, reduces their exclusion by same-race peers in the upper half of the SES distribution. Therefore, educators aiming to implement programs to foster interracial friendships must be alert to potentially unintended consequences. If programs were to replace cross-SES ties within racial groups with same-SES ties between them, they would unintentionally strengthen socioeconomic boundaries.

This study also provides evidence for socioeconomic preferences, teasing apart homophilous and aspirational tendencies (An 2022; An and McConnell 2015; Homans 1961; Laumann 1965; Malacarne 2017). The results indicate that low- and medium-SES students prefer to befriend high-SES peers (aspiration) whereas high-SES students prefer to befriend other high-SES peers (homophily). These tendencies suggest an overlap of social and socioeconomic status in high school and contrast with students' racial preferences that largely reflect homophilous tendencies. These effects do not constitute direct evidence of intentional socioeconomic segregation, but they certainly invite future research to examine the drivers underlying these residual tendencies. Comparable analyses in European secondary schools find little evidence for socioeconomic preferences of students net of other modeled mechanisms (Chabot 2024;

Lenkewitz 2023; Zwier and Geven 2023). More so than ethnoracial homophily for which there is ample evidence in Europe, the existence of SES homophily therefore seems to depend on the institutional and cultural context.[15] Further, students' socioeconomic preferences have an ambivalent impact on socioeconomic segregation. While SES homophily among high-SES students increases closure among them, their popularity integrates friendship networks from below by fostering low→high and med→high ties. These aspirational tendencies, however, fail to truly reduce socioeconomic segregation because high-SES students leave many of these friendship overtures unreciprocated, creating the asymmetry characteristic of socioeconomic segregation in high school.

Finally, there is little evidence to suggest that socioeconomic segregation depends on students' preferences regarding peers' academic performance. The reason is that average- rather than high-performing students are most popular, breaking the link to socioeconomic segregation.

Relational mechanisms. Tie-formation mechanisms that reflect students' tendency to base friendship choices on already existing ties do not contribute much to socioeconomic segregation. Their effects on socioeconomic segregation are small, ambivalent, and offset each other, which contrasts with their sizable effects on friendship formation. Therefore, relational mechanisms must be modeled to accurately measure the effects of other modeled mechanisms. However, their effects on friendship formation do not translate into large effects on socioeconomic segregation. Unlike in Schelling's (1971) classic model of residential segregation in which modest preferences are dynamically amplified, relational mechanisms increase local clustering with only minor effects on school-wide socioeconomic segregation. The reason is that students' tendencies to reciprocate ties and to close open triads produce both same- and cross-SES ties. For reciprocity, the segregating impact predominates. In contrast, for triadic closure, the integrating impact prevails, challenging a long-held assumption that triadic closure amplifies network segregation (e.g., Goodreau, Kitts, and Morris 2009; Kossinets and Watts 2009; Tóth et al. 2021; Wimmer and Lewis 2010). Both mechanisms contribute to the exclusion of low-SES students from high-SES cliques, however. Therefore, to prevent clique-formation among high-SES students, teachers should change seating plans regularly and have students with different SES backgrounds work together. In doing so, reciprocity and balancing norms can help overcome homophilous tendencies of high-SES youth.

In conclusion, bridges across racial and socioeconomic divides are important to sustain social cohesion as the US becomes more ethnoracially diverse and economically unequal (Alba and Maggio 2022). The results from this study reveal more socioeconomic mixing in high school than many may realize. For a majority of Americans, high school is a place where they befriend peers with diverse socioeconomic backgrounds, the evidence of socioeconomic segregation in their friendship networks notwithstanding. To fully achieve the promises of SES-integrated education, however, policymakers and educators should target socioeconomic disparities between schools, SES-stratified courses within schools, and students' racial and socioeconomic preferences. The results show that racial homophily is behind much of what Chetty and colleagues identify as apparent SES homophily. Therefore, SES-integrated friendship networks in educational settings will be difficult to achieve without also addressing racial segregation.

NOTES

[1] See, among others, Bourdieu (1980); Coleman (1968); Crosnoe, Cavanagh, and Elder (2003); Davies et al. (2011); Lin (1999, 2000); Portes (1998).

[2] Studies examining the socioeconomic composition of classrooms and schools find negative effects of larger shares of low-SES students on the students' academic performance. See van Ewijk and Sleegers (2010) for an overview. This compositional effect is associated with school funding, teacher quality, and classroom climate (Sciffer, Perry, and McConney 2020). Given a school's socioeconomic composition, however, peer effect studies find no evidence that cross-SES friendships impact the academic performance of high-SES students.

[3] Homophily is the human tendency to befriend others similar to us. Note that the term is also used to describe homogeneity in networks ("induced homophily"), which is the outcome of homophily and other homogeneity-inducing mechanisms. I follow Wimmer and Lewis (2010) to define homophily as the individual preference ("choice homophily") since "love of the similar" clearly refers to a preference and use homogeneity and assortativity to describe the network-level outcome.

[4] Goodreau et al. (2009) and Chabot (2024) find that triadic closure amplifies network segregation. However, the effect of triadic closure depends on whether open triads close regardless of whether the edge crosses group boundaries. Therefore, without interacting triadic closure and homophily, triadic closure is bound to have a segregating impact (Abebe et al. 2022; Block 2015, 2018; Grund and Densley 2015). Unfortunately, neither paper models this interaction.

[5] Robustness checks in which I choose different percentiles as cutoff points, differentiate the SES score into more than three categories, and examine the continuous SES score show that the results are not sensitive to this operationalization (see Appendix B). A typical low-SES student lives with one parent, who has a high school diploma, works in the service sector, and provides $28,000 in annual income (1994 dollars). A typical medium-SES student lives with two parents who have at least some college education, one of them works professionally, and the household income is $46,000. A typical high-SES student lives with two parents who have graduate degrees, work professional jobs, and the household income is $65,000.

[6] In the same way, I measure segregation with respect to race, gender, GPA, language spoken at home, and parental education, occupation, and income. The cutoff points for ordinal and continuous variables are as follows: parental education (low: ≤high school, med: some college, high: ≥college degree), parental occupation (three-group EGP class schema), parental income and GPA (tercile cutoff points).

[7] The maximum value of S depends on the expected proportion. The maximum value could be restricted to 1 for $\pi_t^{obs} > \pi_t^{exp}$ using $S_t^* = (\pi_t^{obs} - \pi_t^{exp})/(1 - \pi_t^{exp})$. This measure, however, is less intuitively interpretable and gives more weight to variables that exhibit high expected proportions.

[8] In a robustness check (see Appendix C5), I measure the impact of each mechanism by switching it off while keeping all other mechanisms switched on. This approach gives similar results with some nuances. A disadvantage of this approach is that contributions do not sum up (i.e., $\sum \tau_{\theta_k} \neq \hat{\pi}_t^{M3,full} - \hat{\pi}_t^{M3,empty}$) because relational mechanisms amplify all but the absent mechanism and because mechanisms compensate for each other due to intercorrelations.

[9] Chetty and colleagues measure socioeconomic segregation as the absence of peers with an above-median (high) SES background among the friends of students with a below-median (low) SES background. I find that, in the Add Health networks, high-SES peers are underrepresented in friendship networks of low-SES students by 16 percent, which is more than they report (11 percent).

[10] Comparing this result to Figure 4 shows that schools in the saturated subsample (sample 2) are about as socioeconomically segregated as schools overall (sample 1).

[11] The friendship formation model shows that homophily on a subracial level reduces but does not absorb the effect of racial homophily on friendship formation.

[12] Schaefer et al. (2018) find that extracurricular activities do not reduce racial segregation in friendship networks either. In contrast, Malacarne (2017) finds extracurricular activities to be positively associated with cross-SES interaction with the same data but using logistic regression. The divergent conclusions highlight the limits of using dyadic models to parse determinants of network structure characteristics.

[13] Giancola & Kahlenberg (2016) find that low-SES students are less likely to take advanced courses even when grades would allow them. Pedersen and Seidman (2005) show that they are less likely to participate in extracurricular activities.

[14] This result is in line with Mouw and Entwisle (2006), who find that the "bus stop" effect does not affect the racial segregation in high school friendship networks very much either.

[15] The presented results contrast with evidence from European secondary schools in another way. Figure 5 indicates that high-SES students are more likely than their low-SES peers to form interracial ties net of meeting opportunities. High-SES students in the Netherlands, in contrast, are less likely than their high-SES peers to form interethnic ties net of meeting opportunities (Damen, Martinović, and Stark 2021).

REFERENCES

Abebe, Rediet, Nicole Immorlica, Jon Kleinberg, Brendan Lucier, and Ali Shirali. 2022. ‘On the Effect of Triadic Closure on Network Segregation’. *EC 2022 - Proceedings of the 23rd ACM Conference on Economics and Computation* 249–84. doi: 10.1145/3490486.3538322.

Alba, Richard. 2005. ‘Bright vs. Blurred Boundaries: Second-Generation Assimilation and Exclusion in France, Germany, and the United States’. *Ethnic and Racial Studies* 28(1):20–49. doi: 10.1080/0141987042000280003.

Alba, Richard, and Christopher Maggio. 2022. ‘Demographic Change and Assimilation in the Early 21st-Century United States’. *Proceedings of the National Academy of Sciences* 119(13). doi: 10.1073/pnas.2118678119.

An, Weihua. 2015. ‘Multilevel Meta Network Analysis with Application to Studying Network Dynamics of Network Interventions’. *Social Networks* 43:48–56. doi: 10.1016/J.SOCNET.2015.03.006.

An, Weihua. 2022. ‘Friendship Network Formation in Chinese Middle Schools: Patterns of Inequality and Homophily’. *Social Networks* 68:218–28. doi: 10.1016/J.SOCNET.2021.07.003.

An, Weihua, Roberson Beauvile, and Benjamin Rosche. 2022. ‘Causal Network Analysis’. *Annual Review of Sociology* 48:23–41. doi: 10.1146/ANNUREV-SOC-030320-102100.

An, Weihua, and Will R. McConnell. 2015. ‘The Origins of Asymmetric Ties in Friendship Networks: From Status Differential to Self-Perceived Centrality’. *Network Science* 3(2):269–92. doi: 10.1017/NWS.2015.12.

Asikainen, Aili, Gerardo Iñiguez, Javier Ureña-Carrión, Kimmo Kaski, and Mikko Kivelä. 2020. ‘Cumulative Effects of Triadic Closure and Homophily in Social Networks’. *Science Advances* 6(19). doi: 10.1126/sciadv.aax7310.

Ball, Brian, and M. E. J. Newman. 2013. ‘Friendship Networks and Social Status’. *Network Science* 1(1):16–30. doi: 10.1017/NWS.2012.4.

Barone, Carlo, Florian R. Hertel, and Oscar Smallenbroek. 2022. ‘The Rise of Income and the Demise of Class and Social Status? A Systematic Review of Measures of Socio-Economic Position in Stratification Research’. *Research in Social Stratification and Mobility* 78:100678. doi: 10.1016/J.RSSM.2022.100678.

Bischoff, Kendra, and Sean F. Reardon. 2014. ‘Residential Segregation by Income, 1970–2009’. Pp. 208–33 in *Diversity and Disparities: America Enters a New Century*, edited by J. Logan. Russel Sage Foundation.

Bischoff, Kendra, and Laura Tach. 2020. ‘School Choice, Neighborhood Change, and Racial Imbalance Between Public Elementary Schools and Surrounding Neighborhoods’. *Sociological Science* 7:75–99. doi: 10.15195/V7.A4.

Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.

Block, Per. 2015. ‘Reciprocity, Transitivity, and the Mysterious Three-Cycle’. *Social Networks* 40:163–73. doi: 10.1016/J.SOCNET.2014.10.005.

Block, Per. 2018. ‘Network Evolution and Social Situations’. *Sociological Science* 5:402–31. doi: 10.15195/v5.a18.

Bojanowski, Michał, and Rense Corten. 2014. ‘Measuring Segregation in Social Networks’. *Social Networks* 39(1):14–32. doi: 10.1016/J.SOCNET.2014.04.001.

Bourdieu, Pierre. 1980. ‘Le Capital Social. Notes Provisoires’. *Actes de La Recherche En Sciences Sociales* 31(1):2–3. doi: 10.1086/226505.

Bourdieu, Pierre. 1984. *Distinction*. Routledge.

Caetano, Gregorio, and Hugh Macartney. 2021. 'What Determines School Segregation? The Crucial Role of Neighborhood Factors'. *Journal of Public Economics* 194:104335. doi: 10.1016/J.JPUBECO.2020.104335.

Chabot, Timothée. 2024. 'How Does Socioeconomic Homophily Emerge? Testing for the Contribution of Different Processes to Socioeconomic Segregation in Adolescent Friendships'. *Social Networks* 76:160–73. doi: 10.1016/J.SOCNET.2023.09.002.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022a. 'Social Capital I: Measurement and Associations with Economic Mobility'. *Nature 2022 608:7921* 608(7921):108–21.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022b. 'Social Capital II: Determinants of Economic Connectedness'. *Nature 2022 608:7921* 608(7921):122–34.

Chmielewski, Anna K. 2014. 'An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems'. *American Journal of Education* 120(3):293–324. doi: 10.1086/675529.

Chung, Bobby W. 2020. 'Peers' Parents and Educational Attainment: The Exposure Effect'. *Labour Economics* 64:101812. doi: 10.1016/J.LABECO.2020.101812.

Cohen, Jere. 1979. 'Socio-Economic Status and High-School Friendship Choice: Elmtown's Youth Revisited'. *Social Networks* 2(1):65–74. doi: 10.1016/0378-8733(79)90011-X.

Coleman, J. S. 1968. *The Evaluation of Equality of Educational Opportunity*.

Coleman, James. 1961. *The Adolescent Society*. New York: Free Press.

Cotterell, John. 2007. *Social Networks in Youth and Adolescence*. Routledge.

Crosnoe, Robert, Shannon Cavanagh, and Glen H. Elder. 2003. 'Adolescent Friendships as Academic Resources: The Intersection of Friendship, Race, and School Disadvantage'. *Sociological Perspectives* 46(3):331–52. doi: 10.1525/SOP.2003.46.3.331.

Davies, Kristin, Linda R. Tropp, Arthur Aron, Thomas F. Pettigrew, and Stephen C. Wright. 2011. 'Cross-Group Friendships and Intergroup Attitudes: A Meta-Analytic Review'. *Personality and Social Psychology Review* 15(4):332–51. doi: 10.1177/1088868311411103.

Dijkstra, Jan Kornelis, Antonius H. N. Cillessen, and Casey Borch. 2013. 'Popularity and Adolescent Friendship Networks: Selection and Influence Dynamics'. *Developmental Psychology* 49(7):1242–52. doi: 10.1037/A0030098.

DiPrete, Thomas A., Andrew Gelman, Tyler Mccormick, Julien Teitler, and Tian Zheng. 2011. 'Segregation in Social Networks Based on Acquaintanceship and Trust'. *American Journal of Sociology* 116(4):1234–83. doi: 10.1086/659100.

Dunbar, R. I. M. 2016. 'Do Online Social Media Cut through the Constraints That Limit the Size of Offline Social Networks?' *Royal Society Open Science* 3(1):150292. doi: 10.1098/rsos.150292.

Duxbury, Scott W. 2021. 'The Problem of Scaling in Exponential Random Graph Models': *Sociological Methods & Research*. doi: 10.1177/0049124120986178.

Duxbury, Scott W. 2023. 'Micro Effects on Macro Structure in Social Networks'. *Sociological Methods and Research*.

Duxbury, Scott W., and Jenna Wertsching. 2023. 'Scaling Bias in Pooled Exponential Random Graph Models'. *Social Networks* 74:19–30. doi: 10.1016/J.SOCNET.2023.02.003.

Edelmann, Achim, and Stephen Vaisey. 2014. 'Cultural Resources and Cultural Distinction in Networks'. *Poetics* 46:22–37. doi: 10.1016/J.POETIC.2014.08.001.

van Ewijk, Reyn, and Peter Sleegers. 2010. 'The Effect of Peer Socioeconomic Status on Student Achievement: A Meta-Analysis'. *Educational Research Review* 5(2):134–50.

Feld, Scott L. 1981. 'The Focused Organization of Social Ties'. *American Journal of Sociology* 86(5):1015–35. doi: 10.1086/227352.

Ferguson, Sherelle, and Annette Lareau. 2021. 'Hostile Ignorance, Class, and Same-Race Friendships: Perspectives of Working-Class College Students'. *Socius: Sociological Research for a Dynamic World* 7:237802312110483. doi: 10.1177/23780231211048305.

Flashman, Jennifer. 2012. 'Academic Achievement and Its Impact on Friend Dynamics'. *Sociology of Education* 85(1):61–80.

Frank, Kenneth A., Chandra Muller, and Anna S. Mueller. 2013. 'The Embeddedness of Adolescent Friendship Nominations: The Formation of Social Capital in Emergent Network Structures'. *American Journal of Sociology* 119(1):216–53. doi: 10.1086/672081.

Giancola, Jennifer, and Richard D. Kahlenberg. 2016. 'True Merit: Ensuring Our Brightest Students Have Access to Our Best Colleges and Universities.' *Jack Kent Cooke Foundation*.

Gidron, Noam, and Peter A. Hall. 2020. 'Populism as a Problem of Social Integration'. *Comparative Political Studies* 53(7):1027–59. doi: 10.1177/0010414019879947.

Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. 'Birds of a Feather, Or Friend of a Friend?: Using Exponential Random Graph Models to Investigate Adolescent Social Networks'. *Demography* 46(1):103–25. doi: 10.1353/dem.0.0045.

Gould, Roger V. 2002. 'The Origins of Status Hierarchies: A Formal Theory and Empirical Test1'. *American Journal of Sociology* 107(5):1143-1178+I. doi: 10.1086/341744.

Grund, Thomas. 2014. 'Network Size and Network Homophily: Same-Sex Friendships in 595 Scandinavian Schools'. Pp. 297–314 in *Analytical Sociology: Actions and Networks*, edited by G. Manzo. John Wiley & Sons, Ltd.

Grund, Thomas U., and James A. Densley. 2015. 'Ethnic Homophily and Triad Closure: Mapping Internal Gang Structure Using Exponential Random Graph Models'. *Journal of Contemporary Criminal Justice* 31(3):354–70. doi: 10.1177/1043986214553377.

Hanushek, Eric A., Jacob D. Light, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. 2022. 'Long-Run Trends in the U.S. SES Achievement Gap'. *Education Finance and Policy* 17(4):608–40. doi: 10.1162/EDFP_A_00383.

Harris, Kathleen Mullan. 2018. 'The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009; Wave V, 2016-2018'.

Hjalmarsson, Simon, and Carina Mood. 2015. 'Do Poorer Youth Have Fewer Friends? The Role of Household and Child Economic Resources in Adolescent School-Class Friendships'. *Children and Youth Services Review* 57:201–11. doi: 10.1016/J.CHILDYOUTH.2015.08.013.

Hofstra, Bas, Rense Corten, Frank van Tubergen, and Nicole B. Ellison. 2017. 'Sources of Segregation in Social Networks: A Novel Approach Using Facebook'. *American Sociological Review* 82(3):625–56. doi: 10.1177/0003122417705656.

Holland, Paul W., and Samuel Leinhardt. 1981. 'An Exponential Family of Probability Distributions for Directed Graphs'. *Journal of the American Statistical Association* 76(373):33–50. doi: 10.1080/01621459.1981.10477598.

Hollingshead, A. 1975. *Elmtown's Youth and Elmtown Revisited*. New York: Wiley.

Hollingshead, A. B. 1949. *Elmstown's Youth; the Impact of Social Classes on Adolescents*. Wiley.

Homans, G. C. 1961. *Social Behavior: Its Elementary Forms*. Harcourt, Brace.

Hughes, Bradley T., Cory K. Costello, Joshua Pearman, Pooya Razavi, Cianna Bedford-Petersen, Rita M. Ludwig, and Sanjay Srivastava. 2021. 'The Big Five Across Socioeconomic Status: Measurement Invariance, Relationships, and Age Trends'. *Collabra: Psychology* 7(1). doi: 10.1525/COLLABRA.24431/117346.

Kalogrides, Demetra, and Susanna Loeb. 2013. 'Different Teachers, Different Peers: The Magnitude of Student Sorting Within Schools'. *Educational Researcher* 42(6).

Kao, Grace, Kara Joyner, and Kelly Stamper Balistreri. 2019. *The Company We Keep: Interracial Friendships and Romantic Relationships from Adolescence to Adulthood*. Russel Sage Foundation.

Khan, Shamus. 2011. *Privilege: The Making of an Adolescent Elite at St. Paul's School*. Princeton University Press.

Kitts, James A., and Diego F. Leal. 2021. 'What Is(n't) a Friend? Dimensions of the Friendship Concept among Adolescents'. *Social Networks* 66:161–70. doi: 10.1016/j.socnet.2021.01.004.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. 'Origins of Homophily in an Evolving Social Network'. *American Journal of Sociology* 115(2):405–50. doi: 10.1086/599247.

Krause, Robert W., Mark Huisman, Christian Steglich, and Tom Snijders. 2020. 'Missing Data in Cross-Sectional Networks – An Extensive Comparison of Missing Data Treatment Methods'. *Social Networks* 62:99–112. doi: 10.1016/J.SOCNET.2020.02.004.

Krivitsky, Pavel N., Pietro Coletti, Niel Hens, and South Wales. 2023. 'A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks'. *Journal of the American Statistical Association* 1–22. doi: 10.1080/01621459.2023.2242627.

Lamont, Michèle, and Marcel Fournier. 1992. *Cultivating Differences: Symbolic Boundaries and the Making of Inequality*. edited by M. Lamont and M. Fournier. University of Chicago Press.

Lamont, Michele, and Annette Lareau. 1988. 'Cultural Capital: Allusions, Gaps and Glissandos in Recent Theoretical Developments'. *Sociological Theory* 6(2):153. doi: 10.2307/202113.

Laumann, Edward O. 1965. 'Subjective Social Distance and Urban Occupational Stratification'. *American Journal of Sociology* 71(1):26–36. doi: 10.1086/223990.

Lazega, Emmanuel, and Tom A. B. Snijders. 2016. *Multilevel Network Analysis for the Social Sciences*. Springer .

Lee, Francis, and Carter T. Butts. 2018. 'Mutual Assent or Unilateral Nomination? A Performance Comparison of Intersection and Union Rules for Integrating Self-Reports of Social Relationships'. *Social Networks* 55:55–62. doi: 10.1016/j.socnet.2018.05.005.

Lenkewitz, Sven. 2023. 'Limited Opportunities: Adolescents' Access to Social Capital in Secondary Schools in Three European Countries'. *Social Networks* 74:245–58. doi: 10.1016/J.SOCNET.2023.05.001.

Lessard, Leah M., and Jaana Juvonen. 2019. 'Cross-Class Friendship and Academic Achievement in Middle School'. 55(8):1666–79.

Lewis, Kevin, and Jason Kaufman. 2018. 'The Conversion of Cultural Tastes into Social Network Ties'. *American Journal of Sociology* 123(6):1684–1742. doi: 10.1086/697525.

Lin, Nan. 1999. 'Social Networks and Status Attainment'. *Annual Review of Sociology* 25(Weber 1946):467–87. doi: 10.1146/annurev.soc.25.1.467.

Lin, Nan. 2000. 'Inequality in Social Capital'. *Contemporary Sociology* 29(6):785–95.

Lizardo, Omar. 2006. 'How Cultural Tastes Shape Personal Networks'. *American Sociological Review* 71(5):778–807. doi: 10.1177/000312240607100504.

Lubbers, Miranda J. 2003. 'Group Composition and Network Structure in School Classes: A Multilevel Application of the P∗ Model'. *Social Networks* 25(4):309–32. doi: 10.1016/S0378-8733(03)00013-3.

Lusher, D., J. Koskinen, and G. Robins. 2012. *Exponential Random Graph Models for Social Networks. Theory, Methods, and Applications*. Cambridge: Cambridge University Press.

Malacarne, Timothy. 2017. 'Rich Friends, Poor Friends: Inter–Socioeconomic Status Friendships in Secondary School'. *Socius: Sociological Research for a Dynamic World* 3:237802311773699. doi: 10.1177/2378023117736994.

Mark, Noah P. 2003. 'Culture and Competition: Homophily and Distancing Explanations for Cultural Niches'. *American Sociological Review* 68(3):319–45. doi: 10.2307/1519727.

Marsden, Peter V. 1987. 'Core Discussion Networks of Americans'. *American Sociological Review* 52(1):122. doi: 10.2307/2095397.

McFarland, Daniel A., James Moody, David Diehl, Jeffrey A. Smith, and Reuben J. Thomas. 2014. 'Network Ecology and Adolescent Social Structure'. *American Sociological Review* 79(6):1088–1121. doi: 10.1177/0003122414554001.

McMillan, Cassie. 2022. 'Strong and Weak Tie Homophily in Adolescent Friendship Networks: An Analysis of Same-Race and Same-Gender Ties'. *Network Science* 10(3):283–300. doi: 10.1017/NWS.2022.24.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. 'Birds of a Feather: Homophily in Social Networks'. *Annual Review of Sociology* 27:415–44. doi: 10.1146/annurev.soc.27.1.415.

Mijs, Jonathan J. B., and Elizabeth L. Roe. 2021. 'Is America Coming Apart? Socioeconomic Segregation in Neighborhoods, Schools, Workplaces, and Social Networks, 1970–2020'. *Sociology Compass* 15(6):e12884. doi: 10.1111/SOC4.12884.

Moody, James. 2001. 'Race, School Integration, and Friendship Segregation in America'. *American Journal of Sociology* 107(3):679–716. doi: 10.1086/338954.

Mouw, Ted, and Barbara Entwisle. 2006. 'Residential Segregation and Interracial Friendship in Schools'. *American Journal of Sociology* 112(2):394–441. doi: 10.1086/506415.

Muller, Chandra, Jennifer Pearson, Catherine Riegle-Crumb, Jennifer Harris Requejo, Kenneth A. Frank, Kathryn S. Schiller, R. Kelly Raley, Amy G. Langenkamp, Sarah Crissey, Anna Strassmann Mueller, Rebecca Callahan, Lindsey Wilkinson, and Samuel Field. 2007. *National Longitudinal Study of Adolescent Health: Wave III Education Data: Design and Implementation of the Adolescent Health and Academic Achievement Study*. Chapel Hill.

Neugarten, Bernice L. 1946. 'Social Class and Friendship Among School Children'. *American Journal of Sociology* 51(4):305–13. doi: 10.1086/219816.

Owens, Ann, Sean F. Reardon, and Christopher Jencks. 2016. 'Income Segregation Between Schools and School Districts'. *American Educational Research Journal* 53(4):1159–97. doi: 10.3102/0002831216652722.

Pedersen, S., and E. Seidman. 2005. 'Contexts and Correlates of Out-of-School Activity Participation among Low-Income Urban Adolescents'. Pp. 97–122 in *Organized activities as contexts of development*. Psychology Press.

Portes, Alejandro. 1998. 'Social Capital: Its Origins and Applications in Modern Sociology'. *Annual Review of Sociology* 24:1–24. doi: 10.1146/ANNUREV.SOC.24.1.1.

Putnam, Robert D. 2016. *Our Kids: The American Dream in Crisis*. Simon & Schuster.

Quillian, Lincoln, and Mary E. Campbell. 2003. 'Beyond Black and White: The Present and Future of Multiracial Friendship Segregation'. *American Sociological Review* 68(4):540–66. doi: 10.2307/1519738.

Raabe, Isabel J., Chaïm la Roi, and Stephanie Plenty. 2024. 'Down and out? The Role of Household Income in Students' Friendship Formation in School-Classes'. *Social Networks* 78:109–18. doi: 10.1016/J.SOCNET.2023.12.003.

Rich, Peter, and Ann Owens. 2023. 'Neighborhood–School Structures: A New Approach to the Joint Study of Social Contexts'. *Annual Review of Sociology* 49:297–317. doi: 10.1146/ANNUREV-SOC-031021-110311.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi. 2010. 'Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms'. *Annual Review of Sociology* 36:91–115. doi: 10.1146/ANNUREV.SOC.34.040507.134743.

Robins, Garry, Philippa Pattison, and Jodie Woolcock. 2005. 'Small and Other Worlds: Global Network Structures from Local Processes'. *American Journal of Sociology* 110(4):894–936. doi: 10.1086/427322.

Rui, Ning. 2009. 'Four Decades of Research on the Effects of Detracking Reform: Where Do We Stand? A Systematic Review of the Evidence'. *Journal of Evidence-Based Medicine* 2(3):164–83. doi: 10.1111/J.1756-5391.2009.01032.X.

Schaefer, David R., Sandra D. Simpkins, and Andrea Vest Ettekal. 2018. 'Can Extracurricular Activities Reduce Adolescent Race/Ethnic Friendship Segregation?' Pp. 315–39 in *Social Networks and the Life Course. Integrating the Development of Human Lives and Social Relational Networks*, edited by D. F. Alwin, D. H. Felmlee, and D. A. Kreager. Springer International Publishing.

Schelling, Thomas C. 1971. 'Dynamic Models of Segregation'. *Journal of Mathematical Sociology* 1(2):143–86. doi: 10.1080/0022250X.1971.9989794.

Sciffer, Michael G., Laura B. Perry, and Andrew McConney. 2020. 'Critiques of Socio-Economic School Compositional Effects: Are They Valid?' *British Journal of Sociology of Education* 41(4):462–75. doi: 10.1080/01425692.2020.1736000.

Simmel, Georg. 1908. 'Die Kreuzung Sozialer Kreise'. in *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*. Leipzig: Duncker & Humblot.

Sirin, Selcuk R. 2005. 'Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research'. *Review of Educational Research* 75(3):417–53. doi: 10.3102/00346543075003417.

Skvoretz, John. 2013. 'Diversity, Integration, and Social Ties: Attraction versus Repulsion as Drivers of Intra- and Intergroup Relations'. *American Journal of Sociology* 119(2):486–517. doi: 10.1086/674050.

Slaughter, Andrew J., and Laura M. Koehly. 2016. 'Multilevel Models for Social Networks: Hierarchical Bayesian Approaches to Exponential Random Graph Modeling'. *Social Networks* 44:334–45. doi: 10.1016/J.SOCNET.2015.11.002.

Small, Mario Luis. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Oxford University Press.

Smith, Jeffrey A., Miller McPherson, and Lynn Smith-Lovin. 2014. 'Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004'. *American Sociological Review* 79(3):432–56. doi: 10.1177/0003122414531776.

Smith, Jeffrey A., Jonathan H. Morgan, and James Moody. 2022. 'Network Sampling Coverage III: Imputation of Missing Network Data under Different Network and Missing Data Conditions'. *Social Networks* 68:148–78. doi: 10.1016/J.SOCNET.2021.05.002.

Snijders, Tom A. B., and Alessandro Lomi. 2019. 'Beyond Homophily: Incorporating Actor Variables in Statistical Network Models'. *Network Science* 7(1):1–19. doi: 10.1017/NWS.2018.30.

Snijders, Tom A. B., and Christian E. G. Steglich. 2015. *Representing Micro–Macro Linkages by Actor-Based Dynamic Network Models*. Vol. 44.

Stewart, Jonathan, Michael Schweinberger, Michal Bojanowski, and Martina Morris. 2019. 'Multilevel Network Data Facilitate Statistical Inference for Curved ERGMs with Geometrically Weighted Terms'. *Social Networks* 59:98–119. doi: 10.1016/J.SOCNET.2018.11.003.

Tolochko, Petro, and Hajo G. Boomgaarden. 2024. 'Same but Different: A Comparison of Estimation Approaches for Exponential Random Graph Models for Multiple Networks'. *Social Networks* 76:1–11. doi: 10.1016/J.SOCNET.2023.05.003.

Tóth, Gergő, Johannes Wachs, Riccardo Di Clemente, Ákos Jakobi, Bence Ságvári, János Kertész, and Balázs Lengyel. 2021. 'Inequality Is Rising Where Social Network Segregation Interacts with Urban Topology'. *Nature Communications 2021 12:1* 12(1):1–9. doi: 10.1038/s41467-021-21465-0.

van Tubergen, Frank, and Beate Volker. 2014. 'Inequality in Access to Social Capital in the Netherlands'. *Sociology* 49(3):521–38. doi: 10.1177/0038038514543294.

Tuma, N. B., and M. T. Hallinan. 1979. 'The Effects of Sex, Race, and Achievement on Schoolchildren's Friendships'. *Social Forces*.

Verbrugge, Lois M. 1977. 'The Structure of Adult Friendship Choices'. *Social Forces* 56(2):576–97. doi: 10.1093/SF/56.2.576.

Wimmer, Andreas. 2013. *Ethnic Boundary Making: Institutions, Power, Networks*. Oxford University Press.

Wimmer, Andreas, and Kevin Lewis. 2010. 'Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook'. *American Journal of Sociology* 116(2):583–642. doi: 10.1086/653658.

Wright, Erik Olin. 1997. *Class Counts: Comparative Studies in Class Analysis*. Cambridge University Press.

Young, Cristobal, Barum Park, and Nan Feng. 2024. *Inequality in Social Capital. Evidence from Big Diverse Data*.

Zeng, Zhen, and Yu Xie. 2008. 'A Preference-Opportunity-Choice Framework with Applications to Intergroup Friendship'. *American Journal of Sociology* 114(3):615–48. doi: 10.1086/592863.

Zhao, L. 2023. 'From Superdiversity to Consolidation: Implications of Structural Intersectionality for Interethnic Friendships'. *American Journal of Sociology*.

Zwier, Dieuwke, and Sara Geven. 2023. 'Knowing Me, Knowing You: Socio-Economic Status and (Segregation in) Peer and Parental Networks in Primary School'. *Social Networks* 74:127–38. doi: 10.1016/J.SOCNET.2023.03.003.

CONCLUSION

Explaining how macro-level outcomes emerge from their constituting parts at the micro-level is a complex undertaking. In empirical research, however, statistical methods that feature trivial aggregation functions dominate because methods to study more complex aggregation processes remain underdeveloped. In this thesis, I contribute to the development of empirical-statistical methods for the study of micro-macro links. The developed approaches complement analytical-theoretical methods to study aggregation processes, such as agent-based modeling and game theory. An advantage of the developed empirical approaches is that they confront modeled mechanisms with empirical data, which facilitates assessing their relevance, validity, and generalizability. Due to the niche status of computational and mathematical modeling in sociology, empirical methods may also facilitate a more widespread investigation of micro-macro links. The empirical insights derived from these methods demonstrate their potential.

SUPPLEMENTARY MATERIAL CHAPTER 1

APPENDIX 1: R LIBRARY "INEQX"

With this paper, I provide the R library *ineqx* to facilitate research using the variance decomposition approach. The library implements both descriptive and explanatory variance decomposition. Researchers can decide between decomposing the V and the CV^2. The library allows decomposing effects (i) at a single timepoint, (ii) effects across time relative to a specific time point, and (iii) across time relative to a counterfactual scenario. Scenarios can be based on counterfactual group sizes, pre-treatment inequality, and treatment effects on the mean and variance.

Treatment effects on the group-specific means and variances can be estimated either inside or outside the library. Inside the library it is possible to estimate and decompose treatment effects conveniently in one single command. The simple difference estimator and the difference-in-difference estimator are implemented using the generalized additive model for location, scale and shape (GAMLSS) as framework (Rigby and Stasinopoulos 2005). GAMLSS uses maximum penalized likelihood to estimate model coefficients, which is noticeably faster than the Bayesian Markov chain Monte Carlo estimation used in prior research, especially on larger datasets, such as the Current Population Survey. Estimating treatment effects outside the library allows researchers to draw on novel estimation strategies. The library is then simply used to decompose the estimated treatment effects into within- and between-group components.

The R-code to perform a descriptive decomposition is:

```
ineqx(y="inc", ystat = "CV2", group = "SES", time = "year", ref=1980, dat)
```

The R-code to perform an explanatory decomposition is:

```
ineqx(treat="mother", post="birth", y="inc", ystat = "CV2", decomp = "effect",
      group = "SES", time="year", controls = c("race", "edu"), ref=1980, dat)
```

Instead of a reference time, reference values can be provided, e.g., `ref=list(beta=c(0,0,0), lambda=c(0,0,0))`.

APPENDIX 2: DERIVATIONS

A2.1 Equation 2

With repeated cross-sectional or panel data, the change in the variance from t_0 (baseline) to t (any timepoint post baseline) can then be decomposed into the sum of a within-group effect (δ_W), a between-group effect (δ_B), and a compositional effect (δ_C) as detailed in equation (2) by adding and subtracting two components:

$$V_t - V_{t_0} = \sum_j \pi_{jt}\left(\mu_{jt} - \bar{\mu}_t\right)^2 + \sum_j \pi_{jt}\sigma_{jt}^2 - \left(\sum_j \pi_{jt_0}\left(\mu_{jt_0} - \bar{\mu}_{t_0}\right)^2 + \sum_j \pi_{jt_0}\sigma_{jt_0}^2\right) + \underbrace{\sum_j \pi_{jt_0}\left(\mu_{jt} - \bar{\mu}_t\right)^2 - \sum_j \pi_{jt_0}\left(\mu_{jt} - \bar{\mu}_t\right)^2}_{=0} + \underbrace{\sum_j \pi_{jt_0}\sigma_{jt}^2 - \sum_j \pi_{jt_0}\sigma_{jt}^2}_{=0}$$

$= \delta_B^T + \delta_W^T + \delta_C^T$, where $\bar{\mu}_t = \sum_j \pi_{jt}\mu_{jt}$ and

$$\delta_W^T = \sum_j \pi_{jt_0}\left(\sigma_{jt}^2 - \sigma_{jt_0}^2\right)$$

$$\delta_B^T = \sum_j \pi_{jt_0}\left(\left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 - \left(\mu_{jt_0} - \sum_j \pi_{jt_0}\mu_{jt_0}\right)^2\right)$$

$$\delta_C^T = \sum_j\left(\pi_{jt} - \pi_{jt_0}\right)\left(\left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 + \sigma_{jt}^2\right).$$

The same decomposition for the squared coefficient of variation can be better expressed in functional form:

$$CV_t^2 - CV_{t_0}^2 = \frac{\sum_j \pi_{jt}\left(\mu_{jt} - \bar{\mu}_t\right)^2 + \sum_j \pi_{jt}\sigma_{jt}^2}{\bar{\mu}_t^2} - \frac{\left(\sum_j \pi_{jt_0}\left(\mu_{jt_0} - \bar{\mu}_{t_0}\right)^2 + \sum_j \pi_{jt_0}\sigma_{jt_0}^2\right)}{\bar{\mu}_{t_0}^2} = \delta_W^T + \delta_B^T + \delta_C^T$$

where

$$\delta_W^T = W\left(\pi_{jt}, \mu_{jt}, \sigma_{jt}^2\right) - W\left(\pi_{jt_0}, \mu_{jt_0}, \sigma_{jt_0}^2\right)$$

$$\delta_B^T = B\left(\pi_{jt}, \mu_{jt}\right) - B\left(\pi_{jt_0}, \mu_{jt_0}\right)$$

$$\delta_C^T = W\left(\pi_{jt}, \mu_{jt}, \sigma_{jt}^2\right) - W\left(\pi_{jt_0}, \mu_{jt}, \sigma_{jt}^2\right) + B\left(\pi_{jt}, \mu_{jt}\right) - B\left(\pi_{jt_0}, \mu_{jt}\right)$$

and $W(\pi, \mu, \sigma^2) = \sum_j \pi_j\sigma_j^2/\bar{\mu}^2$ and $B(\pi, \mu) = \sum_j \pi_j\left(\mu_j - \sum_j \pi_j\mu_j\right)^2/\bar{\mu}^2$ are the within- and between-group equation functions of the CV^2 that take the parameters $\pi = \pi_1, \ldots, \pi_J$, $\sigma^2 = \sigma_1^2, \ldots, \sigma_J^2$, and $\mu = \mu_1, \ldots, \mu_J$ as input.

A2.2 Equation 5

The change in the variance due to a treatment effect D on the mean and variance of each group at a single timepoint can be decomposed into a between-group effect (δ_B^D) and a within-group effect (δ_W^D):

$\mathrm{V}[\mathrm{Y}(1)|\mathrm{D}] - \mathrm{V}[\mathrm{Y}(0)|\mathrm{D}] = \delta_{\mathrm{B}}^{\mathrm{D}} + \delta_{\mathrm{W}}^{\mathrm{D}}$, where

$$\delta_{\mathrm{B}}^{\mathrm{D}} = \sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(\left(\mu_{\mathrm{j}} + \beta_{\mathrm{j}} - \sum_{\mathrm{j}} \pi_{\mathrm{j}} (\mu_{\mathrm{j}} + \beta_{\mathrm{j}}) \right)^2 - \left(\mu_{\mathrm{j}} - \sum_{\mathrm{j}} \pi_{\mathrm{j}} \mu_{\mathrm{j}} \right)^2 \right)$$

$$\delta_{\mathrm{W}}^{\mathrm{D}} = \sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(\sigma_{\mathrm{j}} + \lambda_{\mathrm{j}} \right)^2 - \sum_{\mathrm{j}} \pi_{\mathrm{j}} \sigma_{\mathrm{j}}^2 = \sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(2\lambda_{\mathrm{j}} \sigma_{\mathrm{j}} + \lambda_{\mathrm{j}}^2 \right)$$

The same decomposition for the squared coefficient of variation:

$\mathrm{CV}^2[\mathrm{Y}(1)|\mathrm{D}] - \mathrm{CV}^2[\mathrm{Y}(0)|\mathrm{D}] = \delta_{\mathrm{B}}^{\mathrm{D}} + \delta_{\mathrm{W}}^{\mathrm{D}}$, where

$$\delta_{\mathrm{B}}^{\mathrm{D}} = \frac{\sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(\left(\mu_{\mathrm{j}} + \beta_{\mathrm{j}} - \sum_{\mathrm{j}} \pi_{\mathrm{j}} (\mu_{\mathrm{j}} + \beta_{\mathrm{j}}) \right)^2 - \left(\mu_{\mathrm{j}} - \sum_{\mathrm{j}} \pi_{\mathrm{j}} \mu_{\mathrm{j}} \right)^2 \right)}{\bar{\mu}^2}$$

$$\delta_{\mathrm{W}}^{\mathrm{D}} = \frac{\sum_{\mathrm{j}} \pi_{\mathrm{j}} \left(2\lambda_{\mathrm{j}} \sigma_{\mathrm{j}} + \lambda_{\mathrm{j}}^2 \right)}{\bar{\mu}^2}$$

A2.3 Equation 8

With repeated cross-sectional or panel data, the change in the total variance from t_0 (baseline) to t (any timepoint post baseline) due to change in the effect of treatment can be decomposed into the sum of a between-group ($\delta_B^{D,T}$), within-group ($\delta_W^{D,T}$) a compositional ($\delta_C^{D,T}$), and a pre-treatment effect ($\delta_P^{D,T}$):

$$(V[Y_t(1)|D_t] - V[Y_t(0)|D_t]) - \left(V\left[Y_{t_0}(1)|D_{t_0}\right] - V\left[Y_{t_0}(0)|D_{t_0}\right]\right)$$

$$= \sum_j \pi_{jt}\left(\left(\mu_{jt} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt} + \beta_{jt})\right)^2 - \left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 + \left(\sigma_{jt} + \lambda_{jt}\right)^2 - \sigma_{jt}^2\right) - \sum_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt_0} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt_0})\right)^2 - \left(\mu_{jt_0} - \sum_j \pi_{jt_0}\mu_{jt_0}\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt_0}\right)^2 - \sigma_{jt_0}^2\right) +$$

$$\underbrace{\sum_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 + \left(\mu_{jt} - \sum_j \pi_{jt_0}\mu_{jt}\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2 + \sigma_{jt}^2\right) - \sum_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 + \left(\mu_{jt} - \sum_j \pi_{jt_0}\mu_{jt}\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2 + \sigma_{jt}^2\right)}_{=0} +$$

$$\underbrace{\sum_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2\right) - \sum_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2\right)}_{=0}$$

$= \delta_B^{D,T} + \delta_W^{D,T} + \delta_C^{D,T} + \delta_D^{D,T}$, where

$$\delta_B^{D,T} = B\left(\pi_{t_0}, \mu_{t_0} + \beta_t\right) - B\left(\pi_{t_0}, \mu_{t_0} + \beta_0\right)$$

$$= \sum_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 - \left(\mu_{jt_0} + \beta_{jt_0} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt_0})\right)^2\right)$$

$$\delta_W^{D,T} = W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_t\right) - W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_{t_0}\right)$$

$$= \sum_j \pi_{jt_0}\left(\left(\sigma_{jt_0} + \lambda_{jt}\right)^2 - \left(\sigma_{jt_0} + \lambda_{jt_0}\right)^2\right) = \sum_j \pi_{jt_0}\left(\lambda_{jt}^2 - \lambda_{jt_0}^2 + 2\sigma_{jt_0}\left(\lambda_{jt} - \lambda_{jt_0}\right)\right)$$

$$\delta_C^{D,T} = \left(B\left(\pi_t, \mu_{t_0} + \beta_t\right) - B\left(\pi_{t_0}, \mu_{t_0} + \beta_t\right)\right) - \left(B(\pi_t, \mu_t) - B\left(\pi_{t_0}, \mu_t\right)\right) + \left(W\left(\pi_t, \sigma_{t_0} + \lambda_t\right) - W\left(\pi_{t_0}, \sigma_{t_0} + \lambda_t\right)\right) - \left(W(\pi_t, \sigma_t) - W\left(\pi_{t_0}, \sigma_t\right)\right)$$

$$= \sum_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 - \left(\mu_{jt} - \sum_j \pi_{jt}\mu_{jt}\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2 - \sigma_{jt}^2\right) - \sum_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 - \left(\mu_{jt} - \sum_j \pi_{jt_0}\mu_{jt}\right)^2 + \left(\sigma_{jt_0} + \lambda_{jt}\right)^2 - \sigma_{jt}^2\right)$$

$$\delta_D^{D,T} = B(\pi_t, \mu_t + \beta_t) - B\left(\pi_t, \mu_{t_0} + \beta_t\right) + W(\pi_t, \sigma_t + \lambda_t) - W\left(\pi_t, \sigma_{t_0} + \lambda_t\right) - \left(B\left(\pi_{t_0}, \mu_t\right) - B\left(\pi_{t_0}, \mu_{t_0}\right) + W\left(\pi_{t_0}, \sigma_t\right) - W\left(\pi_{t_0}, \sigma_{t_0}\right)\right)$$

$$= \sum_j \pi_{jt}\left(\left(\mu_{jt} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt} + \beta_{jt})\right)^2 - \left(\mu_{jt_0} + \beta_{jt} - \sum_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + \left(\sigma_{jt} + \lambda_{jt}\right)^2 - \left(\sigma_{jt_0} + \lambda_{jt}\right)^2\right) - \sum_j \pi_{jt_0}\left(\left(\mu_{jt} - \sum_j \pi_{j,0}\mu_{jt}\right)^2 - \left(\mu_{jt_0} - \sum_j \pi_{jt_0}\mu_{jt_0}\right)^2 + \sigma_{jt}^2 - \sigma_{jt_0}^2\right)$$

The same decomposition for the squared coefficient of variation is possible using the same approach as in A2.1. The equations are not shown to conserve space.

A2.4 Decomposition of the change in post-treatment variance induced by the change in the effect of treatment on the variance

$$
\begin{aligned}
&\mathrm{V}[\mathrm{Y_t}(1)|\mathrm{D_t}] - \mathrm{V}\big[\mathrm{Y_{t_0}}(1)|\mathrm{D_{t_0}}\big] \\
&\qquad = \sum\nolimits_j \pi_{jt}\left(\left(\mu_{jt} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt} + \beta_{jt})\right)^2 + (\sigma_{jt} + \lambda_{jt})^2\right) - \sum\nolimits_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt_0} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt_0})\right)^2 - (\sigma_{jt_0} + \lambda_{jt_0})^2\right) \\
&\qquad + \underbrace{\sum\nolimits_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right) - \sum\nolimits_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right)}_{=0} \\
&\qquad + \underbrace{\sum\nolimits_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right) - \sum\nolimits_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right)}_{=0}
\end{aligned}
$$

$= \delta_B^{D,T} + \delta_W^{D,T} + \delta_C^{D,T} + \delta_D^{D,T}$, where

$$
\begin{aligned}
\delta_B^{D,T} &= \sum\nolimits_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 - \left(\mu_{jt_0} + \beta_{jt_0} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt_0})\right)^2\right) \\
\delta_W^{D,T} &= \sum\nolimits_j \pi_{jt_0}\left((\sigma_{jt_0} + \lambda_{jt})^2 - (\sigma_{jt_0} + \lambda_{jt_0})^2\right) = \sum\nolimits_j \pi_{jt_0}\left(\lambda_{jt}^2 - \lambda_{jt_0}^2 + 2\sigma_{jt_0}(\lambda_{jt} - \lambda_{jt_0})\right) \\
\delta_C^{D,T} &= \sum\nolimits_j \pi_{jt}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right) - \sum\nolimits_j \pi_{jt_0}\left(\left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt_0}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right) \\
&\approx \sum\nolimits_j (\pi_{jt} - \pi_{jt_0})\left(\left(\mu_{jt} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt} + \beta_{jt})\right)^2 + (\sigma_{jt_0} + \lambda_{jt})^2\right) \\
\delta_P^{D,T} &= \sum\nolimits_j \pi_{jt}\left(\left(\mu_{jt} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt} + \beta_{jt})\right)^2 - \left(\mu_{jt_0} + \beta_{jt} - \sum\nolimits_j \pi_{jt}(\mu_{jt_0} + \beta_{jt})\right)^2 + (\sigma_{jt} + \lambda_{jt})^2 - (\sigma_{jt_0} + \lambda_{jt})^2\right)
\end{aligned}
$$

The same decomposition for the squared coefficient of variation is possible using the same approach as in A2.1. The equations are not shown to conserve space.

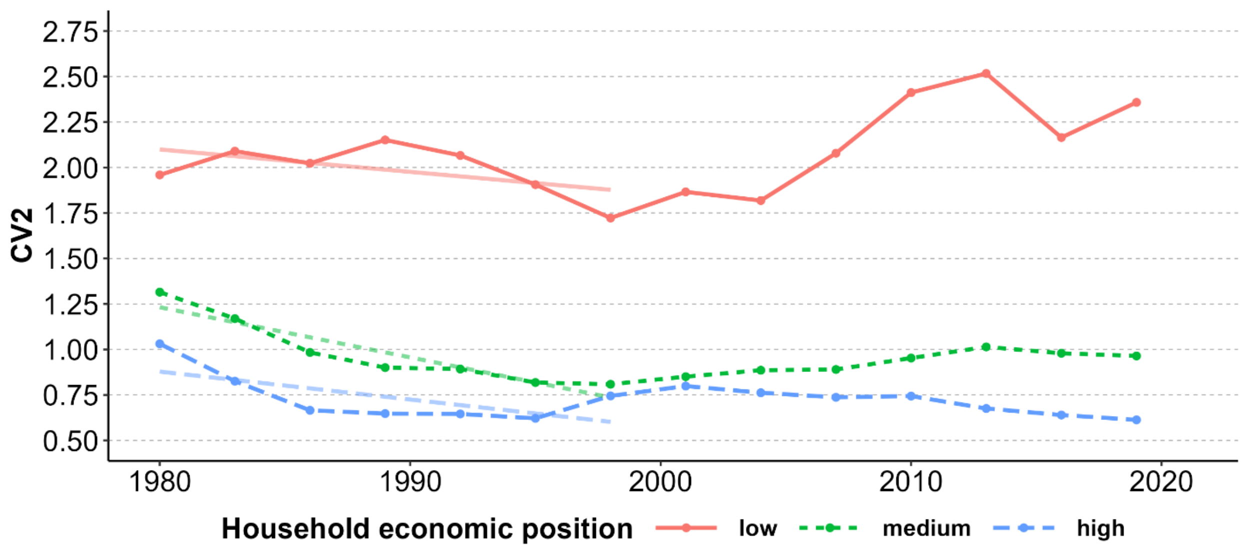
APPENDIX 3: APPLICATIONS

Application 1

Table A3.1: Decomposition of the change in the CV^2

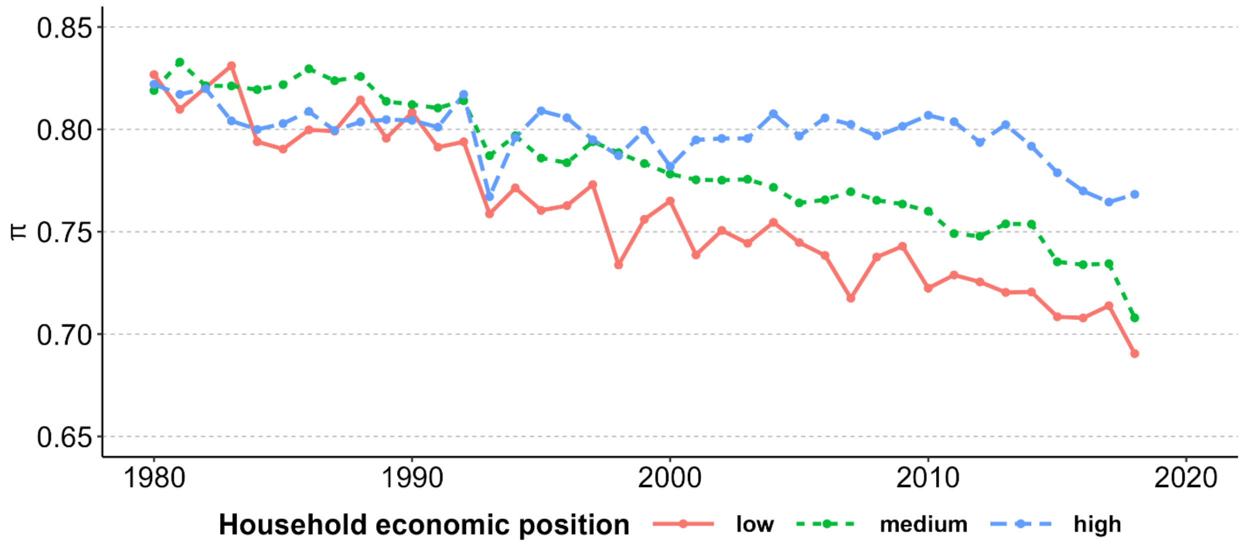
Year	$\delta_W^{D,T}$	$\delta_B^{D,T}$	$\delta_C^{D,T}$	$\delta_T^{D,T}$
1980	0	0	0	0
1985	-0.240	0.036	-0.009	-0.213
1990	-0.282	0.032	0.000	-0.250
1995	-0.297	0.039	0.001	-0.257
2000	-0.219	0.025	-0.006	-0.200
2005	-0.224	0.037	0.002	-0.185
2010	-0.201	0.048	0.035	-0.118
2015	-0.257	0.053	-0.006	-0.209
2020	-0.263	0.059	-0.047	-0.251

Figure A3.1: Trends in the group-specific CV^2



Application 2

Figure A3.2: Share of mothers in the final sample by economic strata over time



The results of the explanatory variance decomposition depicted in Figure 7 are given by Table A3.2.

Table A3.2: Decomposition of the change in the motherhood effect

Year	$\delta_W^{D,T}$	$\delta_B^{D,T}$	$\delta_C^{D,T}$	$\delta_P^{D,T}$	$\delta_T^{D,T}$
1980	0.000	0.000	0.000	0.000	0.000
1985	-0.001	0.018	0.002	-0.003	0.016
1990	-0.012	0.022	-0.011	0.008	0.008
1995	-0.061	0.014	-0.017	0.017	-0.047
2000	-0.109	0.001	-0.015	0.056	-0.068
2005	-0.098	0.025	-0.027	0.051	-0.049
2010	-0.099	0.064	0.002	0.008	-0.025
2015	-0.098	0.018	0.023	0.017	-0.041
2020	-0.113	-0.012	0.015	0.028	-0.082

To calculate how much inequality has changed over time due to the motherhood effect, I decompose the change in the post-treatment variance due to the motherhood rather than the motherhood effect itself. The results are shown by Table A3.3. Within-group inequality is about 11.5 percent higher in 2020 than in 1980 because $\left(1-\frac{0.113}{0.87+0.113}\right)*100 \approx 88.5$, where -0.113 is the within-group effect in 2020, $0.87+0.113$ is the post-treatment CV_W^2 in 2020 without the within-group effect. The inequality-reducing within-group effect, therefore, is 11.5 percent higher. The same calculation is done for $\delta_B^{D,T}$ and $\delta_T^{D,T}$.

Table A3.3: Decomposition of the change in the post-treatment CV^2 due to the motherhood effect

Year	$\delta_W^{D,T}$	$\delta_B^{D,T}$	$\delta_C^{D,T}$	$\delta_P^{D,T}$	$\delta_T^{D,T}$	CV_W^2	CV_B^2	CV_T^2
1980	0.000	0.000	0.000	0.000	0.000	1.50	0.17	1.63
1985	-0.002	0.018	-0.018	-0.002	-0.003	1.18	0.21	1.36
1990	-0.013	0.021	-0.020	0.023	0.011	1.08	0.19	1.25
1995	-0.062	0.012	-0.033	-0.005	-0.088	1.05	0.22	1.24
2000	-0.108	0.001	-0.027	0.181	0.046	1.30	0.21	1.42
2005	-0.096	0.024	-0.039	0.069	-0.042	1.13	0.26	1.33
2010	-0.099	0.063	-0.017	0.049	-0.004	1.00	0.26	1.20
2015	-0.097	0.018	-0.055	0.153	0.018	0.92	0.26	1.14
2020	-0.113	-0.011	-0.086	0.192	-0.017	0.87	0.25	1.09

The results of the decomposition in reference to a zero-effect depicted in Figure 8 are given by Table A3.4.

Table A3.4: Decomposition of the change in the post-treatment CV^2 due to the motherhood effect in reference to a zero-effect

Year	$\delta_W^{D,T}$	$\delta_B^{D,T}$	$\delta_C^{D,T}$	$\delta_P^{D,T}$	$\delta_T^{D,T}$	CV_W^2	CV_B^2	CV_T^2
1980	0.009	0.009	0.000	0.000	0.018	1.50	0.17	1.63
1985	-0.001	0.034	0.000	0.000	0.033	1.18	0.21	1.36
1990	-0.004	0.029	0.000	0.000	0.025	1.08	0.19	1.25
1995	-0.049	0.018	0.000	0.000	-0.031	1.05	0.22	1.24
2000	-0.057	0.008	0.000	0.000	-0.050	1.30	0.21	1.42
2005	-0.055	0.024	0.000	0.000	-0.031	1.13	0.26	1.33
2010	-0.051	0.044	0.000	0.000	-0.007	1.00	0.26	1.20
2015	-0.052	0.029	0.000	0.000	-0.023	0.92	0.26	1.14
2020	-0.069	0.005	0.000	0.000	-0.064	0.87	0.25	1.09

APPENDIX 4: REPLICATION OF WODTKE (2016)

Wodtke (2016) uses the 1980 to 2010 waves of the GSS. The author provided me with the data and analysis files, allowing me to use the same analytical sample. The key variables of the analysis are individual earnings as outcome and social class (proprietors, independent producers, managers, and workers) as grouping variable. I refer the reader to the original paper for further details. Figure A4.1 shows that I am able to replicate the paper's basic descriptives as the trends in the variance of log income match Figure 6 and 7 in the paper (Wodtke 2016: 1402-3). Panel A of Figure A4.2 then shows that I am able to replicate the main decomposition results of Table 1 in which the variance of log income is decomposed (Wodtke 2016: 1404). Finally, Panel B and C show the same decomposition but for the variance of income and the squared coefficient of variation. The figure indicates that the results are largely identical. Therefore, in the case of Wodtke (2016), the choice between variance, variance of logarithms, or squared coefficient of variation would have not made a difference in interpretation. As the application the present paper shows, this might not hold in other applications.

Figure A4.1: The trends in the V_L match Figures 6 and 7 in the paper (Wodtke 2016: 1402-3)

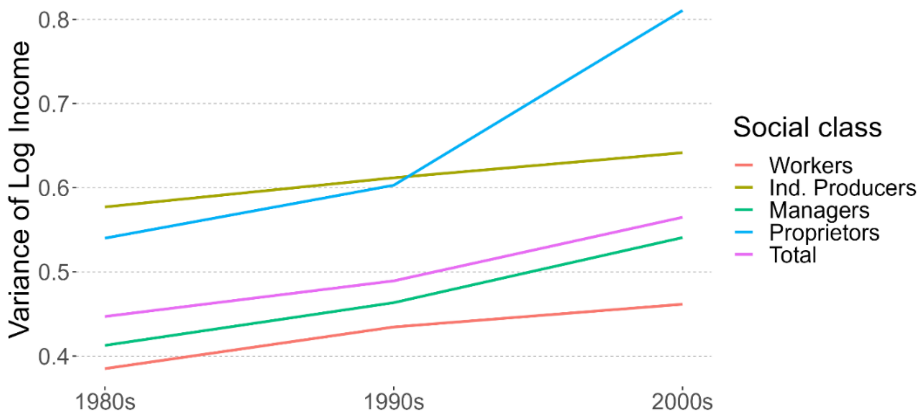
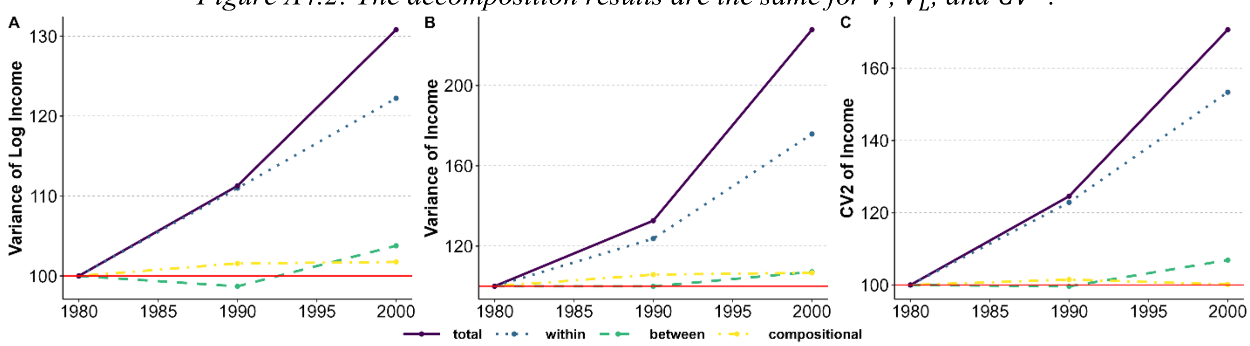


Figure A4.2: The decomposition results are the same for V, V_L, and CV^2.



REFERENCES

Rigby, R. A., and D. M. Stasinopoulos. 2005. "Generalized Additive Models for Location, Scale and Shape." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(3):507–54. doi: 10.1111/J.1467-9876.2005.00510.X.

Wodtke, Geoffrey T. 2016. "Social Class and Income Inequality in the United States: Ownership, Authority, and Personal Income Distribution from 1980 to 2010." *American Journal of Sociology* 121(5):1375–1415. doi: 10.1086/684273.

SUPPLEMENTARY MATERIAL CHAPTER 2

APPENDIX A1:
ADVANTAGES OF USING BAYESIAN ESTIMATION TO FIT MULTILEVEL MODELS

First, Bayesian estimation of multilevel models naturally accounts for uncertainty in the variance estimates while likelihood estimation treats variance estimates as fixed values. Moreover, Bayesian estimation provides variance estimates and plausible credibility intervals in cases where likelihood estimation would give zero variance estimates and no confidence intervals. The carried-out hypothesis tests are therefore more precise (Hodges 2014; Hox, Moerbeek, and van de Schoot 2017). Second, coalition government datasets cover the entire population of governments within a certain spatio-temporal frame. Frequentist statistics models sampling uncertainty even though there is none – the entire population is observed. Bayesian statistics, by contrast, is not based on sampling uncertainty but on a belief-based definition of uncertainty. Such a definition better conveys the nature of uncertainty in coalition outcomes, which is that they are realizations of a probabilistic process. Third, the use of weakly informative priors can help identify this more complex model.

APPENDIX A2: RMM PACKAGE

The three-level multiple membership multilevel version of a Weibull proportional hazard model, which is employed in the application below can specified in the rmm package as follows:

```
rmm(Surv(govdur, earlyterm) ~ 1 + majority + minimalwinning +
    mm(id(pid, gid), mmc(findep), mmw(w ~ 1/offset(n)^exp(-(seatshare+hetero)))) +
    hm(id=cid, type=FE),
    family="Weibull", data=dat)
```

In the `mm()` container, the multiple membership structure between parties and governments is specified in three steps. First, party and government ids must be given in `id()`. Second, party-level covariates can be included in `mmc()`. Third, the aggregation function can be specified in `mmw()`. The here shown functional form corresponds to equation 6. However, researchers can specify their own ideas in a flexible way. Mean aggregation, for instance, is specified by `w ~ 1/offset(n)`. Finally, in the `hm()` container it is specified that parties and governments are hierarchically nested in countries and that country fixed-effects shall be estimated.

APPENDIX A3: SIMULATION STUDY

Simulation setup

Rather than generating entirely artificial data, the simulation is based on the same datasets as the subsequent empirical application. In the simulation, I generate a linear and a survival outcome by manipulating the effect of three structural variables ($\mathrm{x}^{\mathrm{P}}, \mathrm{x}^{\mathrm{G}}, \mathrm{x}^{\mathrm{C}}$) and three weight variables ($\mathrm{z}^{\mathrm{P}}, \mathrm{z}^{\mathrm{G}}, \mathrm{z}^{\mathrm{C}}$) to those outcomes.

Generating the linear and survival outcome

The linear outcome, I generate by

$$\mathrm{Y}_{\mathrm{i}}^{\mathrm{G}} = \beta^{\mathrm{G}}\mathrm{x}_{\mathrm{i}}^{\mathrm{G}} + \mathrm{u}_{\mathrm{i}}^{\mathrm{G}} + \beta^{\mathrm{C}}\mathrm{x}_{\mathrm{C(i)}}^{\mathrm{C}} + \mathrm{u}_{\mathrm{C(i)}}^{\mathrm{C}} + \sum_{\mathrm{j}\in\mathrm{P(i)}} \frac{1}{\mathrm{n}_{\mathrm{i}}^{\exp(-\{\rho^{\mathrm{P}}\mathrm{z}_{\mathrm{ij}}^{\mathrm{P}}+\rho^{\mathrm{G}}\mathrm{z}_{\mathrm{ij}}^{\mathrm{G}}+\rho^{\mathrm{C}}\mathrm{z}_{\mathrm{ij}}^{\mathrm{C}}\})}} (\beta^{\mathrm{P}}\mathrm{x}_{\mathrm{ij}}^{\mathrm{P}} + \mathrm{u}_{\mathrm{j}}^{\mathrm{P}}) \tag{1}$$

The survival outcome is generated using the same linear predictor and the approach described by Bender et al. (2005) to sample survival times. Table 1 details the used variables and manipulated parameter values. The data generation process in equation (1) models the actual multilevel structure of coalition government data. In the simulation, I examine the consequences of ignoring this structure by comparing the performance of the SLM vis-à-vis the MMMM in recovering the true effect of the structural variables (i.e., $\hat{\beta}^{\mathrm{P}}, \hat{\beta}^{\mathrm{G}}, \hat{\beta}^{\mathrm{C}}$) when increasing:

(i) similarity of observations at the party level (i.e., variance of the random effect $\sigma^2_{u^P}$)

(ii) number of multiple memberships *m* (i.e., # of parties' government participations over time)

(iii) interdependencies in the aggregation process (i.e., effects of weight variables ρ).

I specify $\beta^{\mathrm{P}} = \beta^{\mathrm{G}} = \beta^{\mathrm{C}} = 1$ to examine bias and power and $\beta^{\mathrm{P}} = \beta^{\mathrm{G}} = \beta^{\mathrm{C}} = 0$ to assess the false positive rate. To separate the consequences of clustering and interdependencies, the weight coefficients ($\rho^{\mathrm{P}}, \rho^{\mathrm{G}}, \rho^{\mathrm{C}}$) are set to 0 when assessing the consequences of party-level clustering (i.e., increasing (i) + (ii)). Conversely, party-level clustering ($\sigma^2_{\mathrm{u}^{\mathrm{P}}}$) is set to 0 when examining the consequences of ignoring interdependencies in the aggregation process (i.e., increasing (iii)).

In total, I examine 33 simulation conditions (i.e., combinations of parameter values) and run each condition 10,000 times. The SLMs are estimated with maximum likelihood and the MMMMs are estimated with Bayesian MCMC.

Table A1.1: Parameters that are manipulated in the simulation

Level	Variable type	Variable name	Parameter values
Party	Structural	x^P: Parties' financial dependency on members	$\beta^P = \{0,1\}$
N=194	Weight	z^P: Parties' relative seat share in government	$\rho^P = \{-5,-2,0,2,5\}$
	Clustering	Similarity of obs. (variance of random effect)	$\sigma^2_{u^P} = \{0,0.1,1,2,4\}$
		Number of multiple memberships	$m = \{1,2,4, \text{act. distr.}\}$
Gov.	Structural	x^G: Majority status	$\beta^G = \{0,1\}$
N=401	Weight	z^G: Governments' ideological heterogeneity	$\rho^G = \{-5,-2,0,2,5\}$
	Disturbance		$\sigma^2_{u^G} = 0$
Country	Structural	x^C: Investiture requirement	$\beta^C = \{0,1\}$
N=18	Weight	z^C: Prime ministerial powers	$\rho^C = \{-5,-2,0,2,5\}$
	Clustering	Similarity of obs. (variance of random effect)	$\sigma^2_{u^C} = 0$

A comment on the mixture of frequentist and Bayesian estimation

I estimate the MMMM in a Bayesian way because Bayesian estimation of multilevel models offers several attractive features, which are detailed in the paper in section 3.4. By contrast, I mainly employ maximum likelihood estimation for the SLM to reduce the overall duration of the simulation. In line with objective Bayesian statistics (Berger 2006), I then examine the frequentist properties (Type-I and II error rate) of the MMMM and compare it to the SLMs, which are estimated in a frequentist way. This is useful because the Type-I error rate of Bayesian models is not known a priori (as compared to the frequentist promise of 5% over the iterations of the simulation when setting the nominal α to 0.05) (Gelman and Tuerlinckx 2000). I decided not to make the Bayesian paradigm center stage of this paper to be able to focus on the advantages of the multilevel approach. In the future, I might add other estimation algorithms to the rmm package.

Manipulating (i), (ii), and (iii)

(i) I draw party-level random effects from a normal distribution $u_j^P \sim N(0, \sigma^2_{u^P})$ and vary the variance $\sigma^2_{u^P}$ from 0 to 4 to assess the impact of increasingly similar observations at the party level. The random effects at the government and country level are set to 0.

(ii) The number of observations at the party level equals the sum of the number of times each party has been in government. I manipulate the number of multiple memberships m (i.e., parties' participations in government) by giving each party m observations. By giving all parties the same number of observations, m can be varied in a reasonably straightforward way. The only difficulty lies in keeping the number of observations at the party level constant. To achieve that, I only use as many observations as actually observed at the party level of the generated data. The approach is not perfect as it assigns each party the same number of observations and does not always use all generated data. I chose this approach nonetheless because other, more complex approaches that I considered were not successful in varying m in a controlled way. I also use the actually observed distribution of parties' participations in government as simulation condition.

(iii) To examine the of bias owing to party-level clustering (i.e., increasing (i) and (ii)), I keep $\boldsymbol{\rho} = \mathbf{0}$. To create interdependencies in the aggregation process, I vary the weight regression coefficients $\rho^{P}, \rho^{G}, \rho^{C}$ from -5 to 5.

Simulations conditions

If I would take the full factorial of parameter values, I would end up with more than 1000 simulation conditions. To keep the simulation study accessible, I therefore combine conditions where it makes sense:

- To analyze the impact of the variance of the party-level random effect on the structural regression coefficients, I create 5 x 2 = 10 simulation conditions. The conditions are just the factorial of parameter values of $\sigma^2_{u^P}$ and β if $\beta^{P}, \beta^{G}, \beta^{C}$ are varied together, $\rho^{P}, \rho^{G}, \rho^{C} = 0$, *m* equals the actual distribution, and $\sigma^2_{u^G} = \sigma^2_{u^C} = 0$.
- To analyze the impact of the number of multiple memberships *m* on the structural regression coefficients, I create 4 x 2 = 8 simulation conditions. The conditions are just the factorial of parameter values of m and β if $\beta^{P}, \beta^{G}, \beta^{C}$ are varied together, $\rho^{P}, \rho^{G}, \rho^{C} = 0$, $\sigma^2_{u^P} = 1$, and $\sigma^2_{u^G} = \sigma^2_{u^C} = 0$.
- To analyze the impact of the weight predictors, I create 5 x 3 = 15 simulation conditions. That is, I first vary ρ^{P} and keep $\beta^{P}, \beta^{G}, \beta^{C} = 1$, $\rho^{G}, \rho^{C} = 0$, and $\sigma^2_{u^P} = \sigma^2_{u^G} = \sigma^2_{u^C} = 0$. Then, I do the same with ρ^{G} and ρ^{G}. I set the variances of the random effects to zero to separate the impact of party-level clustering from the impact of party-level interdependencies.

Therefore, the total number of simulation conditions is 33.

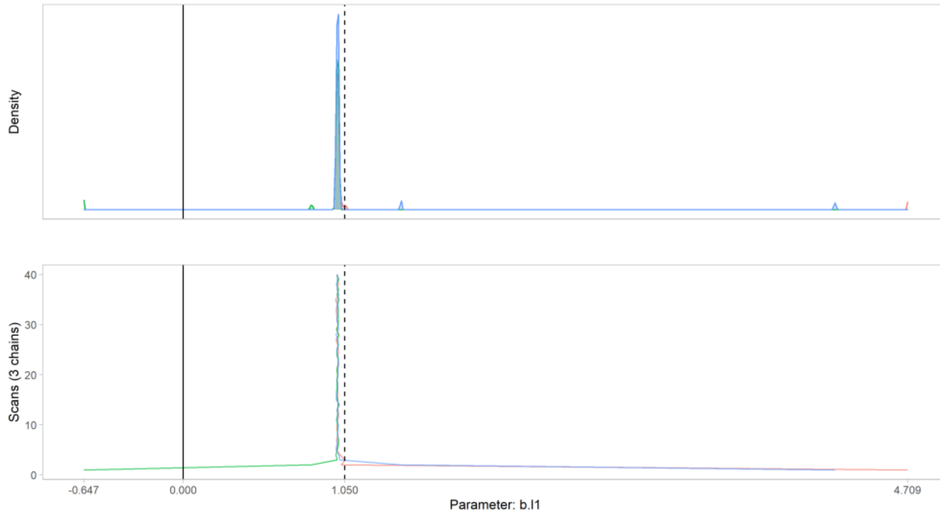
Models

- SLMs: I estimate two SLMs, one at the party level and one at the government level. I use the mean to aggregate the party-level variables to the government level. For the linear outcome, I use the lm_robust command from the estimatr package. For the survival outcome, I use the survreg(…, family="Weibull", robust=T) command from the survival package.
- MMMMs: I estimate two MMMMs using the rmm package – one in which I specify a weight function that depends on the weight covariates, and one in which I use the mean as aggregation function.

Number of iterations and estimation

I run each model in each condition 1,000 times. I use Bayesian MCMC to estimate the MMMMs and maximum likelihood to estimate the SLMs. The MMMM model is run with 2,000 MCMC iterations and 5 chains. Since the ground truth is known in the simulation, I can show that this number is more than is enough to converge to the posterior distribution. Figure A1.1 illustrates that it takes the Gibbs sampler less than 10 steps to converge to the true parameter value of 1. The entire simulation (32 conditions x 4 models x 1,000 iterations of the simulation) takes about a week on a server with 15 cores.

Figure A1.1: Convergence of the party-level regression coefficient. The true value is 1, which is approached in less than 10 steps.



Criteria of evaluation

The criteria for evaluation are (i) bias in the regression coefficient, (ii) bias in the uncertainty estimate, (iii) root mean squared error (RMSE), and (iv) percent of the RMSE that is caused by bias in the regression coefficient, (v) percent true positive (power), and, finally, (vi) percent false positive.

(i) The *bias in the regression coefficient* is defined as $\left(\bar{\hat{\beta}} - \beta\right)/\beta \times 100$, where β is the true parameter value used to generate the data sets and $\bar{\hat{\beta}}$ is the average estimate over the simulated datasets. It illustrates how much the parameter estimates deviate from their true values in percent. Note that the regression coefficients of the frequentist models are maximum likelihood estimates and the regression coefficients of the Bayesian models are means of the posterior distribution.

(ii) The *bias in the uncertainty estimate* is given by $\left(\overline{\widehat{\mathrm{sd}}(\theta)} - \mathrm{sd}\left(\hat{\theta}\right)\right)/\mathrm{sd}(\hat{\theta}) \times 100$, where $\mathrm{sd}(\hat{\theta})$ is the actually observed standard deviation of the estimated regression coefficients across the simulated datasets and $\overline{\widehat{\mathrm{sd}}(\theta)}$ is the average standard deviation of the estimated posterior distributions for the Bayesian models and the average estimated standard errors for the frequentist models. Note that this analysis does not require the standard deviation of the posterior to equal the standard error. This statistic measures whether the models correctly estimate uncertainty. For the frequentist models, the variation across the simulated datasets should equal the average standard error by design. For the Bayesian models, the standard deviation of the posterior should reflect the variation across the simulated datasets to have good frequentist properties.

(iii) In addition to unbiasedness, accuracy is another desirable feature of estimators since it results in narrower confidence intervals and consequently higher power in statistical tests. The mean squared error is an overall measure of accuracy which considers both bias and standard deviation of the estimated parameters. The root mean squared error (RMSE) returns to the original scale of the parameter and therefore measures the spread of the estimates around the true value. The *RMSE* is defined as RMSE $= \sqrt{(\bar{\hat{\theta}} - \theta)^2 + sd(\hat{\theta})^2}$.

(iv) I also report the percent of the RMSE that is caused by bias in the regression coefficient. That is, the degree to which the RMSE is caused by its first element $(\bar{\hat{\theta}} - \theta)^2$. I do this because the simulation indicates that SLM and MMMM sometimes exhibit similar levels of RMSE. Using this measure, I can show that this is because, in some cases, the MMMM is less accurate than the SLM but the SLM is more biased.

(v) The *power* is estimated as the proportion of simulated datasets for which the 95% confidence interval (frequentist models) / 95% posterior credible interval (Bayesian models) does not include zero and $\beta^P = \beta^G = \beta^C = 1$.

(vi) The false positive rate is estimated as the proportion of times the 95% confidence interval (frequentist models) / 95% posterior credible interval (Bayesian models) does not include zero even though $\beta^P = \beta^G = \beta^C = 0$.

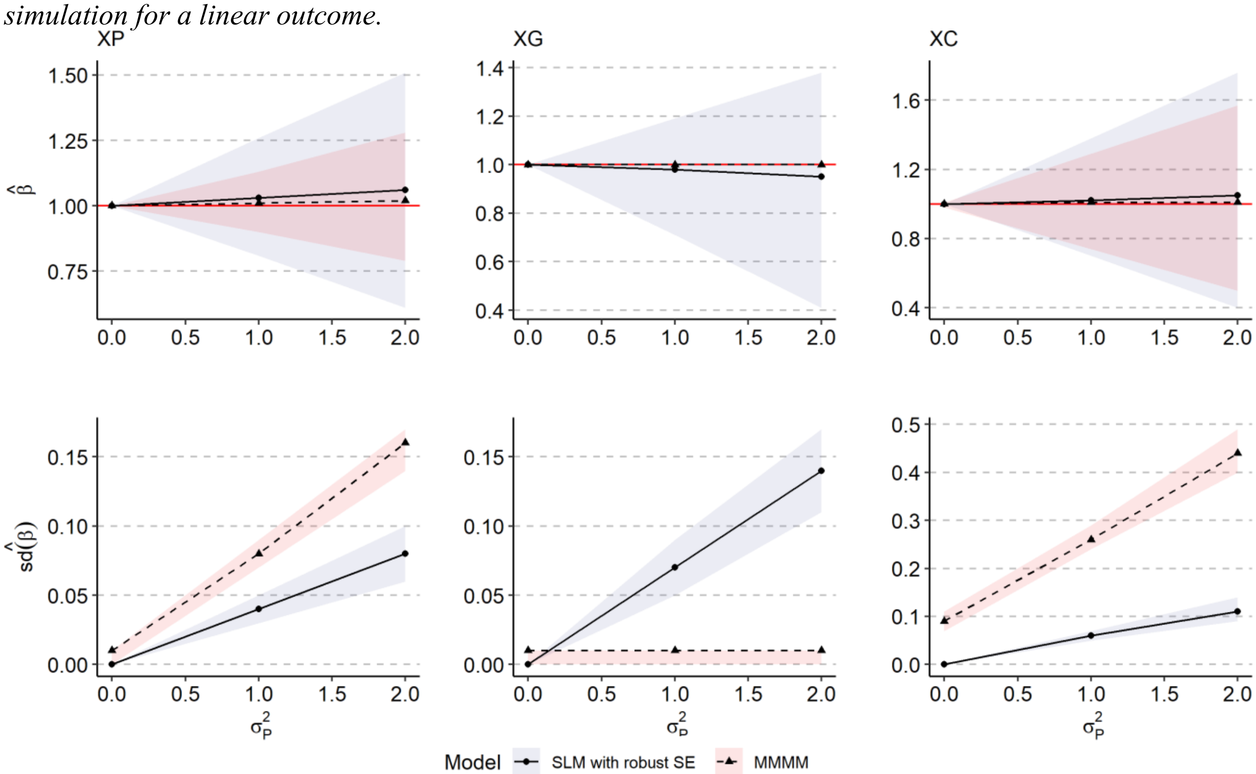
Results: Party-level clustering

Figure A1.2 illustrates the simulation output for a linear outcome. The figure displays the distribution of regression coefficient estimates $\hat{\beta}$ and their uncertainty estimates $\widehat{sd}(\theta)$ under increasing party-level clustering $\sigma^2_{u^P}$ by model type. The first row depicts the mean and standard deviation of the estimated regression coefficients across simulations at the party, government, and country level. The second row shows the estimated uncertainty – the average standard error of the SLMs, and the average standard deviation of the estimated posterior distribution of the MMMM. Ideally, the estimated uncertainty in the second row equals the actual uncertainty displayed in the first row.

The figure indicates that both SLM and MMMM estimate regression coefficients without bias in the linear case. The SLM, however, is less accurate as the standard deviation of regression coefficient estimates across simulation runs (blue shades in the first row) is wider. Moreover, the SLM underestimates this uncertainty, which is visible in the second row. Even though it is in fact less accurate, the SLM estimates that it is more accurate than the MMMM. That is, it the uncertainty estimates are biased towards zero.

Rather than presenting the raw output for all simulation conditions, Table A1.2 and Table A1.3 presents the results in evaluated form. Table A1.2 presents the results when increasing party-level clustering and Table A1.3 presents the results when increasing the number of multiple memberships.

Figure A1.2: Distribution of regression coefficients and uncertainty estimates across the iterations of the simulation for a linear outcome.



Linear outcome. As mentioned in the paper, the regression coefficients ($\bar{\hat{\beta}}$) are unaffected by the presence of party-level clustering ($\sigma^2_{u^P}$) and multiple membership (m) in the linear case. The uncertainty estimates $\widehat{sd}(\theta)$, however, are affected. The SLM underestimates the standard errors considerably. The pattern in Table A1.2 and Table A1.3 indicates that the degree of similarity at the party level is less important than the number of multiple memberships. This is the reason why the bias in SLM is so pronounced – the number of multiple memberships in coalition government data is high (see Figure 3 in the paper). The MMMM estimates the posterior standard deviation at the party level without bias and has a small positive bias at the government and country level. For very small $sd(\hat{\theta})$, the percent bias will be large even though the absolute bias is negligible. Therefore, whenever the bias is very small in absolute magnitudes and the associated percent bias would be misleading, I report sign of the bias $(-/+)$ rather than the percent bias. The RMSE shows that the MMMM is more accurate than the SLM, particularly for government-level variables.

Survival outcome. In the survival case, both regression coefficients and standard errors are affected by the party-level clustering. Tables A1.2 and A1.3 show that the SLM *underestimates* both. Some SLMs had convergence issues during the simulation. I mark cells for which I did not obtain a result by a "-". Looking at the RMSE, we can see that the MMMM is less accurate for country-level effects on the employed dataset (which features only 18 countries). The RMSE considers both bias and standard deviation of the estimated parameters and can be interpreted as the spread of the estimates around the true value. The results reveal that even though SLM and MMMM in some cases (country-level effects!) are equally imprecise, the source of this imprecision is different for the two models. The percentage of the RMSE due to the bias in the regression coefficient indicates that the SLM is imprecise because it is biased while the MMMM is imprecise when the sample size is small and thus the variance of the estimator surges.

Interpreting coefficients at the government level that were estimated at the party level

Figures A1.3- A.15 show that interpreting coefficients at the government level that were estimated at the party level leads to incorrect inference. This is the case irrespective of whether there is party-level clustering and multiple memberships. On the x-axis, the figures show the true regression coefficient, and on the y-axis, they show the estimated effect. The solid line gives the coefficient when estimated at the party level and the dashed line gives the coefficient when estimated at the government level. Figure A1.3 varies the true party-level regression coefficient and keeps the government- and country-level coefficient equal to 1. Figure A1.4 varies the true government-level regression coefficients. Finally, Figure A1.5 varies the true country-level coefficient.

The bias follows a clear pattern. Figure A1.3 shows that party-level coefficients will be overestimated if the true effect is positive and underestimated when the true effect is negative. Importantly, not only the party-level coefficients are affected when a model including variables at all three levels. Government-level coefficients will also be overestimated if the party-level effect is positive and underestimated when the party-level effect is negative. Country-level coefficients show the opposite behavior. As mentioned in the paper, the standard errors will generally be too low when using party-level data.

Figure A1.3 and Figure A1.4 show that estimating government- and country-level effects on party-level data induces less bias. However, standard errors are likewise underestimated. The imprecision visible in Panel A and B of Figure A1.4 are probably because the dataset only includes 18 countries.

Table A1.2: The consequences of increasing party-level clustering

	Level	Model	% Bias $\bar{\bar{\beta}}$				% Bias in $\widehat{sd}(\theta)$				RMSE (same scale as β)				Bias/RMSE (in %)			
			Party-level clustering $\sigma^2_{u^P}$				Party-level clustering $\sigma^2_{u^P}$				Party-level clustering $\sigma^2_{u^P}$				Party-level clustering $\sigma^2_{u^P}$			
			0	1	2	4	0	1	2	4	0	1	2	4	0	1	2	4
Linear	Party	SLM	0	0	0	0	0	-74	-74	-74	0	0.01	0.02	0.04	0	0	0	0
		MMMM	0	0	0	0	0	0	0	0	0	0.01	0.01	0.03	0	0	0	0
	Government	SLM	0	0	0	0	0	-58	-58	-58	0.01	0.16	0.32	0.65	0	0	0	0
		MMMM	0	0	0	0	0	+	+	+	0.02	0.02	0.03	0.03	0	0	0	0
	Country	SLM	0	0	0	0	0	-76	-76	-76	0.01	0.24	0.47	0.95	0	0	0	0
		MMMM	0	0	0	0	0	+	+	+	0.04	0.21	0.40	0.77	0	0	0	0
Survival	Party	SLM	0	-11	-35	-56	0	-43	-46	.	0.04	0.18	0.35	0.59	3	40	98	99
		MMMM	0	0	0	0	0	+	+	+	0.06	0.06	0.06	0.07	0	0	0	0
	Government	SLM	0	-15	-33	-56	0	-22	-13	.	0.13	0.24	0.42	0.64	0	36	61	76
		MMMM	0	0	0	0	0	+	+	+	0.16	0.19	0.22	0.25	0	0	0	0
	Country	SLM	0	-11	-33	-57	0	-41	-44	.	0.11	0.24	0.45	0.69	0	22	54	69
		MMMM	0	0	0	0	0	+	+	+	0.99	0.97	1.05	1.32	0	0	0	0

Table A1.3: The consequences of increasing multiple memberships

	Level	Model	% Bias $\bar{\bar{\beta}}$				% Bias in $\widehat{sd}(\theta)$				RMSE				Bias/RMSE			
			Multiple memberships m				Multiple memberships m				Multiple memberships m				Multiple memberships m			
			1	2	4	A	1	2	4	A	1	2	4	A	1	2	4	A
Linear	Party	SLM	0	0	0	0	-3	-16	-26	-74	0	0	0.01	0.01	0	0	0	0
		MMMM	0	0	0	0	+	+	+	+	0	0	0	0.01	0	0	0	0
	Government	SLM	0	0	0	0	3	-1	-13	-58	0.07	0.09	0.12	0.16	0	0	0	0
		MMMM	0	0	0	0	+	+	+	+	0.07	0.09	0.02	0.02	0	0	0	0
	Country	SLM	0	0	0	0	-1	-7	-24	-76	0.06	0.08	0.12	0.24	0	0	0	0
		MMMM	0	0	0	0	+	+	+	+	0.07	0.09	0.09	0.21	0	0	0	0
Survival	Party	SLM	-15	-21	-26	-11	.	.	-79	-43	.	.	0.27	0.18	.	.	98	39
		MMMM	0	0	0	0	0	+	+	+	0.12	0.13	0.10	0.06	0	0	0	0
	Government	SLM	-14	-22	-24	-15	.	.	-21	-22	.	.	0.31	0.24	.	.	77	35
		MMMM	0	0	0	0	0	+	+	+	0.26	0.29	0.25	0.19	0	0	0	0
	Country	SLM	-12	-19	-24	-11	.	.	-20	-41	.	.	0.29	0.24	.	.	80	22
		MMMM	0	0	0	0	0	+	+	+	1.21	1.30	1.25	0.97	0	0	0	0

Figure A1.3: Estimating party-level effects on government outcomes at the party level.

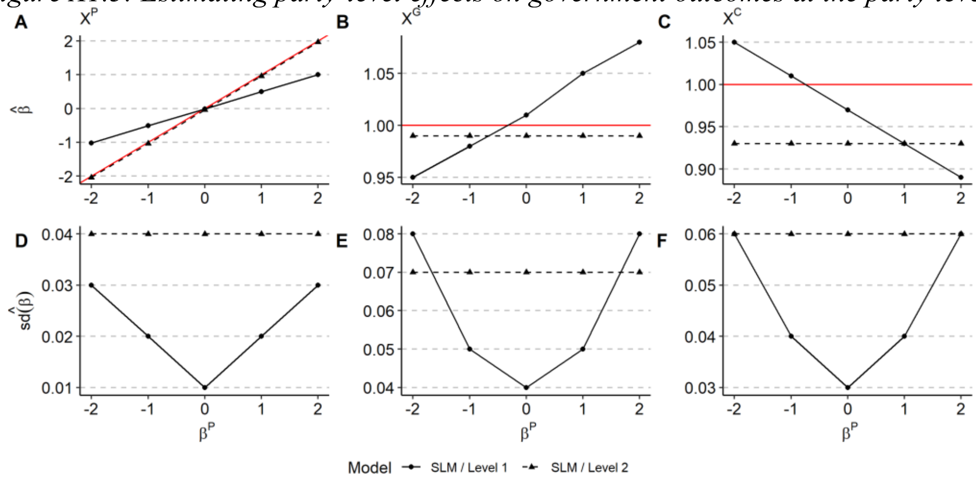


Figure A1.4: Estimating government-level effects on governments at the party level.

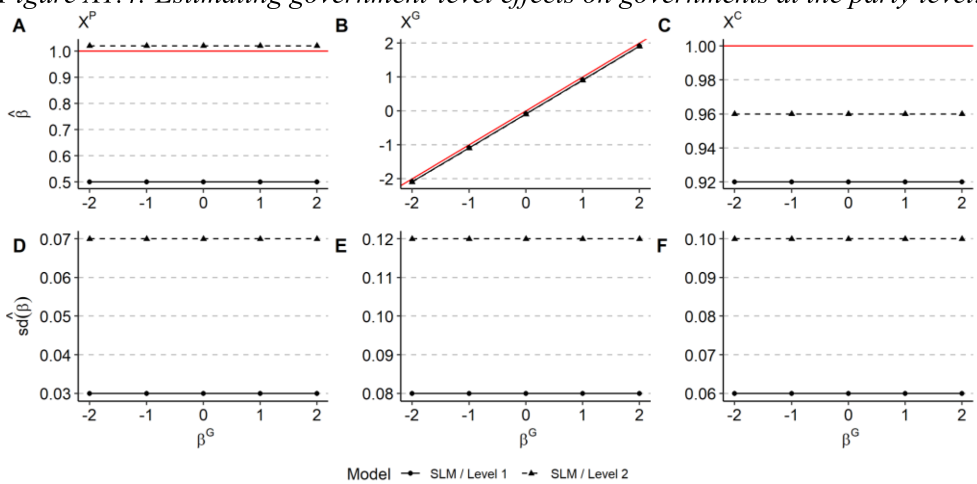
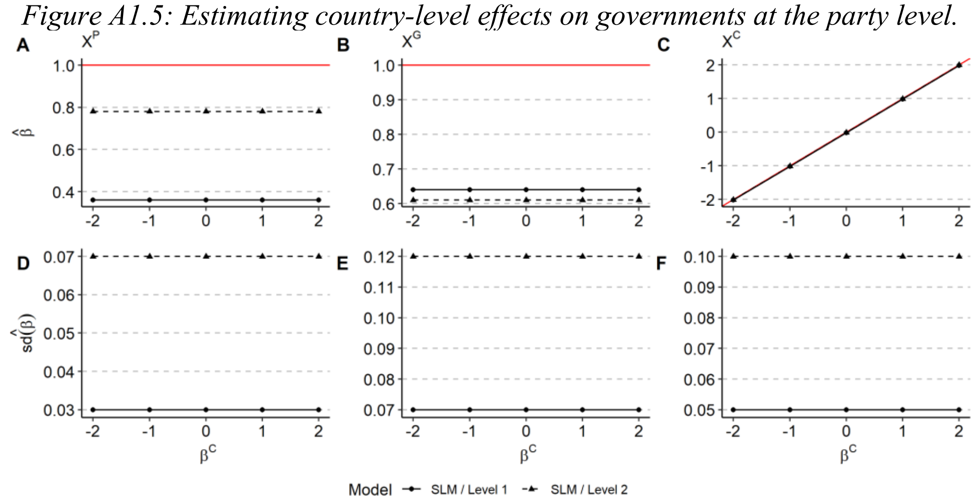


Figure A1.5: Estimating country-level effects on governments at the party level.



Results: Interdependencies in the aggregation process

I present the consequences of ignoring interdependencies in the aggregation process graphically as they are rather nonlinear. Figures A1.6 – A1.8 show the bias of the structural regression coefficient estimates $\hat{\beta}$ and associated uncertainty estimates $\widehat{\mathrm{sd}}(\theta)$ caused by the effect of weight variables ρ in the aggregation process. The considered weight predictors are (i) parties' seat share relative to the other coalition partners (party level), (ii) governments' ideological heterogeneity (government level), and (iii) prime ministerial powers (country level).

Linear outcome. Figure A1.6 shows that weight predictors at the party level barely affect regression coefficients at the party level. Instead, they cause a positive bias in the regression coefficients at the government and country level when the effect of the weight regressor is positive. Figure A1.7 indicates that weight predictors at the government level affect regression coefficients at all three levels. If the effect of the weight regressor is negative, party-level coefficients are overestimated, and government- and country-level coefficients are underestimated. If the effect is positive, party-level coefficients are underestimated, and government- and country-level coefficients are overestimated. Figure A1.8 shows that weight predictors at the country level also affect regression coefficients at all three levels in similar ways.

The impact on the standard error estimates is very similar across conditions: they are overestimated regardless of the effect direction of the weight predictors.

Figure A1.9 shows the impact on percent true positive (power). In the linear case, power is only affected when the effect of weight predictors is negative.

Survival outcome. The effect of weight predictors is different for survival outcomes. Figures A1.6 – A1.8 display a consistent pattern – weight predictors at all three levels cause a negative bias in the regression coefficients and a positive bias in the standard errors at all three levels. This combination, of course, has a detrimental impact on power as Figure A1.9 shows. In particular weight predictors at the party and government level drastically reduce the power to detect effects at the government and country level.

Figure A1.6: Interdependencies arising at the party level.

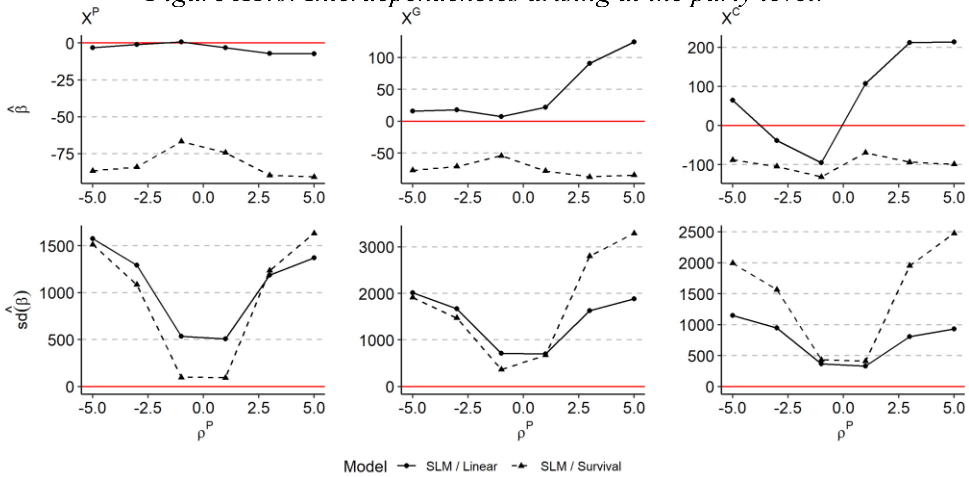


Figure A1.7: Interdependencies arising at the government level.

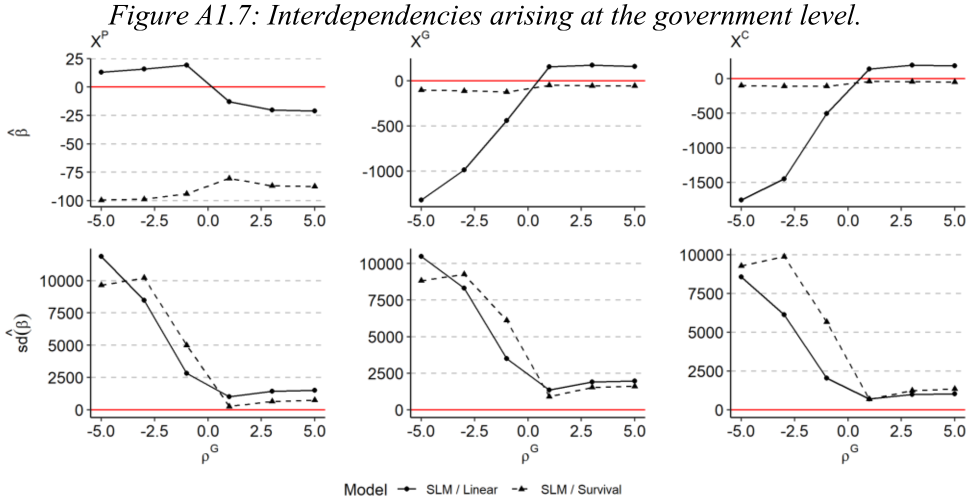


Figure A1.8: Interdependencies arising at the country level.

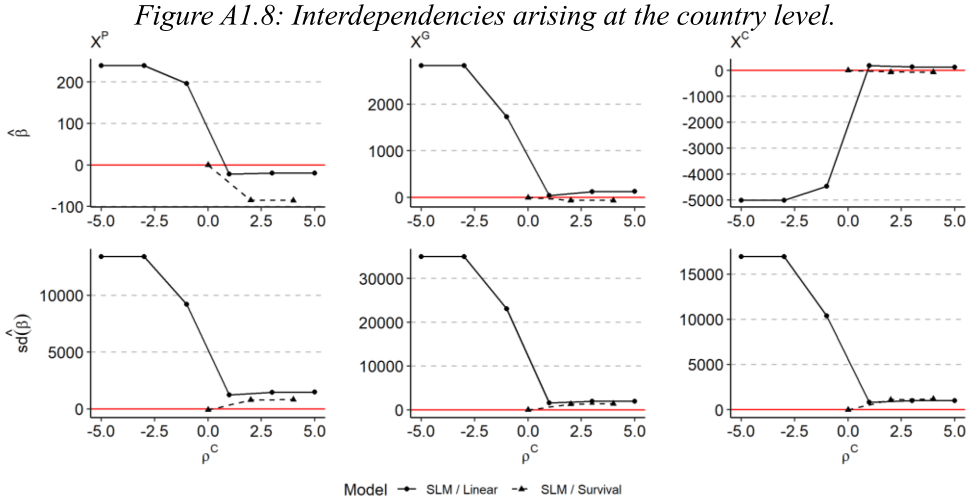


Figure A1.9: Consequences for the power to detect effects (linear outcome)

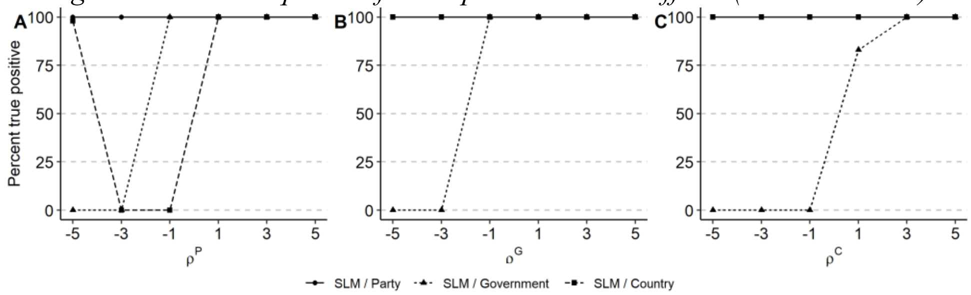
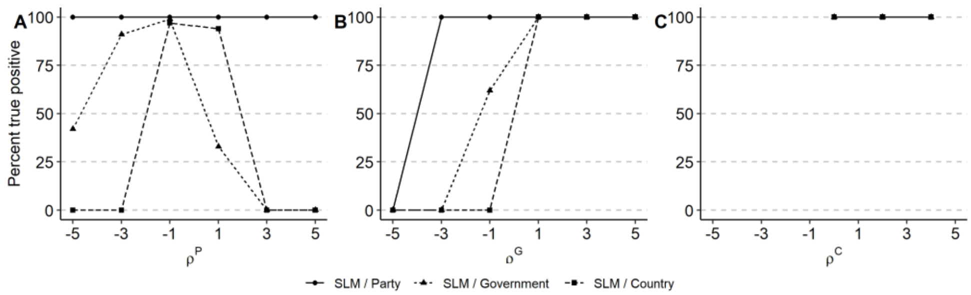


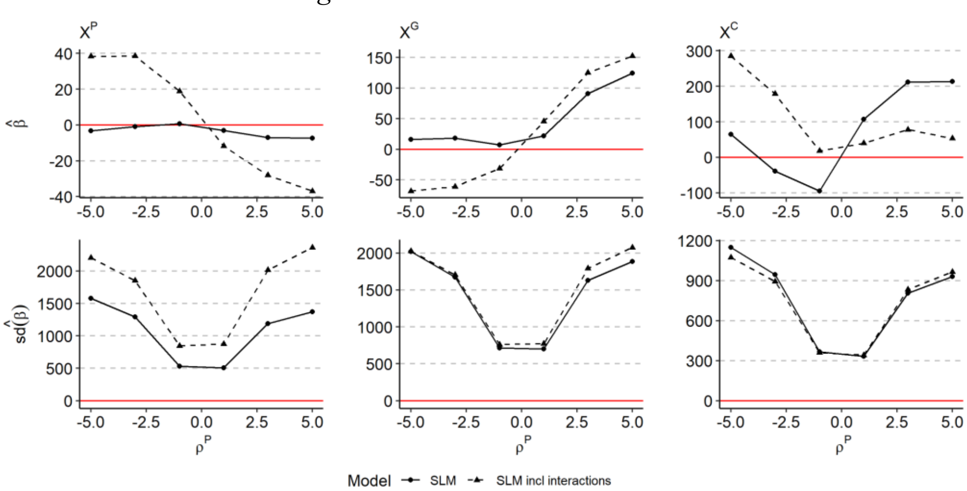
Figure A1.10: Consequences for the power to detect effects (survival outcome)



Interacting party-level variables with the variables causing the interdependencies

One might think that including interactions between structural and weight variables would ameliorate the bias of the SLM. This is, unfortunately, not the case. First, a SLM including interactions between weight and structural variables and a MMMM modeling the weight function are not the same model. The SLM models the effect of weight variables on the outcome and the MMMM models the weight variables on the weights, which are scaled differently than the outcome. Second, the results of such an interaction model in Figure A1.11 show that the structural regression coefficients are actually more heavily biased than without specifying those interactions. The specified interaction model on centered data: *survreg(Surv(sim.st, sim.e) ~ 1 + fdep + majority + investiture +* ***fdep*pseatrel + fdep*hetero + fdep*pmpower****),…).*

Figure A1.11: Interaction model



APPENDIX A4: EMPIRICAL APPLICATION

To illustrate the advantages of modeling the crisscrossing relationship between parties and government, I apply the MMMM to study the effect of parties' financial dependency on their members on the survival of coalition governments. Here, I provide more details on theoretical considerations, data, measurement, and model estimation.

Theory

The effect of parties' financial dependency on their rank and file

There is a latent power distribution within each party apart from the conspicuous set of rules. In fact, the latent power distribution is often considered more informative than the 'official story' captured by the party statutes. Parties need financial capital and labor inputs to be financed, to campaign, to develop and implement party policies, and to get information about the electorate. The less financial capital parties receive from sources other than their members, the more they depend on them. Most members cannot be rewarded with office spoils but are reimbursed with promises about future public policy. Consequently, the more party leaders depend on their rank-and-file, the more they are constrained in departing from their parties' policy position. This policy inflexibility thwarts reaching and maintaining inter-party agreement (Müller and Strøm 1999; Strøm 1990).

Interdependencies in the aggregation process

The distinct features of each coalition party add complexity to coalition governance. Yet, whether parties' financial dependency and other party features will influence government survival likely depends on coalitions' structural situation. Specifically, I argue, *the relationship of coalition parties to each other* will determine the influence of party features because intra- and inter-party politics condition each other. In the following, two aspects of coalitions' interdependence structure are considered, their ideological relationship, and power distribution. As I show in section 3.3 of the paper, these hypotheses suggest a nonlinear aggregation of party effects.

First, the effect of each party on government survival should depend on its clout within coalition. Gamson (1961) famously argued that party impact is proportional to legislative seat share. However, as any party, regardless of seat share, may pull out of the coalition, it is unclear whether this idea holds in the context of government survival.

Second, the impact of de/stabilizing party features should be amplified when the potential for inter-party disagreement is high, which is the case when ideologically heterogenous parties work together. Thus, I hypothesize that party effects increase with the ideological heterogeneity of coalition governments.

There is no empirical record on the weight of parties in their effect on government survival because the aggregation of party effects has not been explicitly modeled yet. Available evidence on the impact of parties' relative seat share is in line with Gamson's law (Bäck, Debus, and Dumont 2011). However, as those studies treat government policy as an independently realized outcome of each party in government, they likely suffer from disaggregation bias (causing excessive Type-I error). A breadth of work shows that larger and ideologically heterogenous coalitions exhibit higher termination hazards (Warwick 1994). These studies, however, do not disentangle whether the increased termination hazard is a direct effect of coalition size and ideological heterogeneity or an indirect effect because *party effects* in larger-sized and ideologically heterogeneous coalitions are amplified.

Data

Offering the most comprehensive information on coalition governments in Western democracies since WWII, I use the Woldendorp, Keman and Budge dataset (2000) (WKB) to source start and end date of each government, reason for termination, government parties, and their distribution of seats in parliament. The recent update (Seki and Williams 2014) covers 1,342 governments in 48 countries between 1944 and 2014.

I use two datasets to source party features. First, the political position of government parties on a right-left scale is taken from the Comparative Manifestos Project (Volkens et al. 2016). Second, Round 1a of the Political Party Database (Scarrow, Poguntke, and Webb 2017) (PPDP) is used to measure parties' financial dependency on their members. Released in 2017, the dataset provides the most detailed information on political party organization available. It is coded by country experts – for the most part directly from party statutes – and describes the extra-parliamentary organization of 122 parties from 20 countries in 2011. Parties' financial dependency is therefore assumed to be time constant. A desideratum for future research is to relax this assumption as more time points become available or party features of past governments are coded. Although the PPDB is the most comprehensive effort to collect comparative information on intra-party features, it covers only 122 parties, which is why I make use of imputation by chained equations with predictive mean matching. This approach is preferable over listwise deletion, which induces strong bias in the presence of nonrandom missingness and small groups (Newman and Sin 2009). To assess the robustness of the approach, I also employed a multivariate normal imputation model and tried multiple combinations of variables in the imputation models. The sensitivity checks confirm that the obtained results are robust to a wide array of specifications and are available on request.

Excluding parties from countries not part of the PPDB, 199 single-party governments, and 21 caretaker governments, I end up with a sample of 401 governments from 18 countries[23], and a total of 194 parties. Table A2.2 presents the univariate descriptives of the final sample.

---

[23] The countries are Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Hungary, Ireland, Israel, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden, United Kingdom.

Table A2.2: Univariate statistics of the employed variables.

Level	Variable	Final sample	
		Mean (SD)	Min / Max
Party	Parties' financial dependency (in percent)	15.923 (14.104)	0.018 / 75
N=194	Parties' share of seats relative to the other coalition partners	-0.053 (0.331)	-0.727 / 0.849
	Parties' position on right-left scale	2.250 (22.023)	-44.500 / 68.966
Government	Premature termination	0.462 (0.499)	0 / 1
	Government duration (in days)	579 (460)	8 / 1840
	Ideological heterogeneity	0.768 (0.416)	0 / 4.444
N=401	Majority	0.786 (0.410)	0 / 1
	Minimal winningness	0.313 (0.464)	0 / 1
	Coalition size	3.820 (1.674)	2 / 9
Country	Investiture vote	0.556 (0.511)	0 / 1
	Prime ministerial powers	4.056 (1.955)	1 / 7

Statistics are calculated on the final sample after imputation. Variables are unstandardized.

Measurement

Government duration. While in WKB, government duration is measured as the time between investiture of a government and investiture of the succeeding government, the duration measure of the European Representative Democracy dataset (Andersson, Bergman, and Ersson 2014) excludes the government formation period. Therefore, I use their duration measure where available. I also exclude crisis/caretaking periods and country differences in term length will be absorbed by country fixed effects.

Government termination. WKB records seven reasons for termination: (1) elections, (2) voluntary resignation of the prime minister, (3) resignation of the prime minister due to health reasons, (4) dissension within government, (5) lack of parliamentary support, (6) intervention by the head of state, and (7) broadening of the coalition. Focusing on termination decisions to escape political deadlock, I coded reason 2, 4, 5, and 6 as conflictual termination and censored all other reasons. I conducted multiple robustness checks for this decision, which are available on request. Lupia and Strøm (1995) differentiate government terminations with respect to the employed constitutional mechanism. This approach is not ideal for the purposes of this paper as the categorization (nonelectoral replacements/early elections) crosses the underlying motivations (gridlock/opportunism). Whether coalitional gridlock leads to nonelectoral replacements or early election depends on the availability of outside offers. Focusing on government durability, it is therefore more expedient to differentiate terminations with respect to their underlying motivations. Finally, I adopt the approach of King et al. (1990) to censor all governments that reach the year preceding regular elections to remove terminations due to strategic election timing.

Parties' financial dependency. The latent dependency of party leaders on their rank-and-file is measured by the share of financial contributions from party members in the total funding of parties, which the PPDB obtained from parties' financial reports.

Ideological heterogeneity. Following Warwick (1994), I compute the standard deviation of coalition parties' right-left score to measure governments' ideological heterogeneity. Benoit and Laver (2007) point to the incomparability of the left-right score across countries and time, which is why I divide the coalitions' standard deviation by the left-right standard deviation of all parties in the parliament that year. Our measure of ideological heterogeneity is therefore relative to the ideological distribution at time and place.

Relative seat share. I define the relative seat share of party j in government i as $S_{ij} = \left(\frac{seats_{ij}}{\sum_j seats_{ij}} - \frac{1}{N_i}\right)\left(\frac{N_i}{N_i-1}\right)$. When parties' seat shares are equally distributed within the coalition, $S = 0$. By contrast, when one party holds all seats (which is never the case), $S = 1$ for the party holding all seats and $S = -1$ for all parties holding no seats. Further explanations for this coding approach are presented in Appendix A4.

Control variables. The following variables are considered: Whether the government holds a majority in parliament, the number of government parties, whether an investiture vote was required, and country dummies.

Model estimation

The model is fitted in JAGS (Plummer 2015; v.4.3) from within R using the rmm package provided with this paper. The approach is employed because popular software packages such as SPSS, Stata, and R are not yet able to estimate continuous-time multiple membership multilevel survival models within endogenized weights. Parameters to be estimated are the vectors of regression coefficients $\beta^{G}, \beta^{P}, \beta^{C}, \rho^{P}, \rho^{G}$, the variance of the random terms $\sigma^2_{u^P}, \sigma^2_{u^C}$, and the shape parameter of the Weibull distribution p. The identification of the variance of party random effects rests on parties having been in different coalitions over time, which is the case. I specified 5 chains, a chain-length of 100,000 with a burn-in of 10,000, and weakly informative priors because I ran in convergence issues with too uninformative priors. The priors are normal distributions for regression coefficients $\beta \sim N(0{,}000.1)$ and $\rho \sim N(0{,}0.1)$, scaled half-t distributions for the standard deviation of random effects (Gelman 2006) $\sigma_U \sim$Half-T$(S = 25, df = 1)$, and an exponential distribution for the shape parameter: $p \sim \exp(0.01)$.

Implementing the weight constraint in JAGS

JAGS does not easily accommodate constraints. The weight function below is equivalent to equation (6) in section 3.3 and programmable in JAGS:

Table A2.4: Implementing the weight constraint in JAGS

Equation (6) in section 3.3	**Equivalent function that is estimable in JAGS**
$\boldsymbol{w_{ij}} = \dfrac{\mathbf{1}}{\boldsymbol{n_i}^{\boldsymbol{exp(-\{z'_{ij}\rho\})}}}$ s.t. $\sum_i \sum_j \boldsymbol{w_{ij}} = \boldsymbol{N}$	$w_{ij} = \dfrac{1}{n_i^{exp(-\{\boldsymbol{z'_{ij}\rho}\})}} \cdot \dfrac{N}{\sum_i \sum_j w_{ij}}$

A comment on the use of multilevel modeling

The prevalent method in the government survival literature is the Cox proportional hazard model, a continuous-time survival model, which accounts for the perfectly hierarchical nesting of governments in countries with the use of *heteroscedasticity-consistent standard errors* (Lin and Wei 1989) or with the use of *shared frailty modeling* (Hougaard 2000). Robust standard errors correct for the misspecification of the error-covariance structure so that the estimates are unbiased even though the nesting structure is ignored. Such linearization methods, however, impose a population-averaged rather than subject-specific interpretation of regression coefficients (Neuhaus, Kalbfleisch, and Hauck 1991). Moreover, while robust standard errors to account for multiple membership structures have been proposed for linear regression (see Aronow, Samii, and Assenova 2015), they do not yet exist for survival models. Shared frailty modeling, which is the approach I take, explicitly models multilevel structure using random terms.

While micro-macro applications are quite rare in the multilevel literature, as this application shows, it is possible to identify the individual contributions to group outcomes within the multilevel framework (Goldstein 2011a, 255-65; Snijders 2016). There are alternatives, such as a latent variable approach by Croon and van Veldhoven 2007 and a network autocorrelation approach by Leenders 2002. The advantage of using multilevel modeling is that the framework is well-developed.

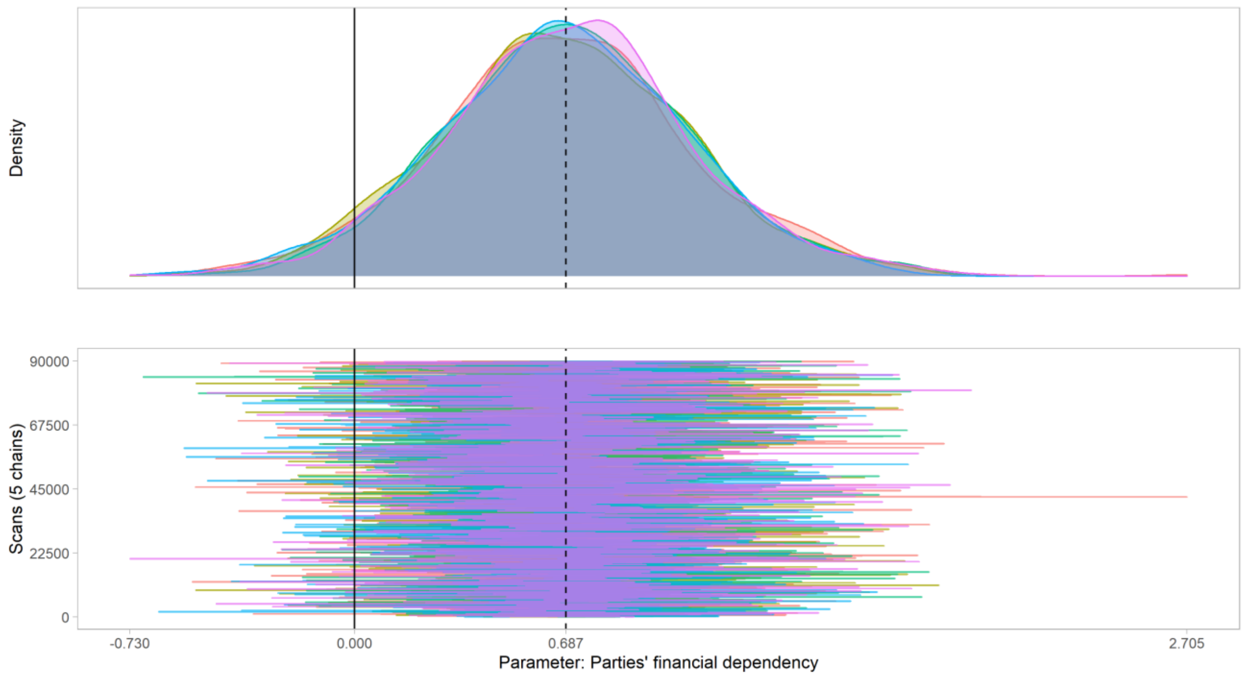
MCMC diagnostics

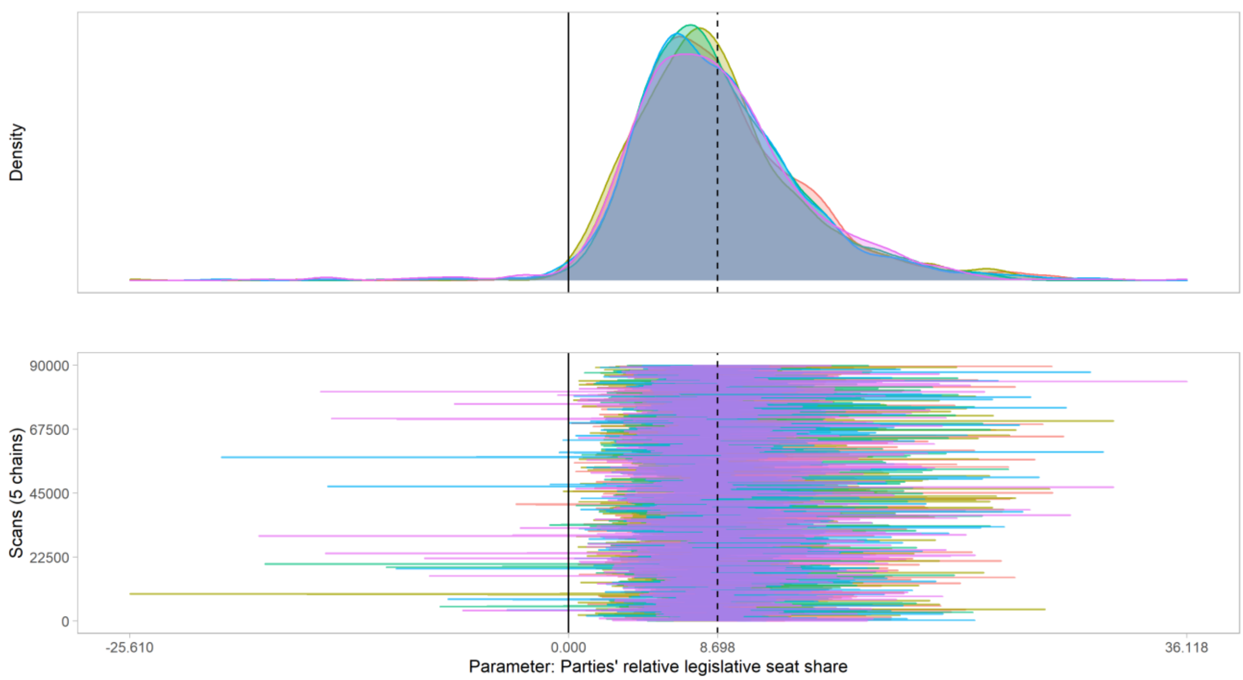
The carried-out MCMC convergence and autocorrelation diagnostics are acceptable and indicate the convergence of all models:

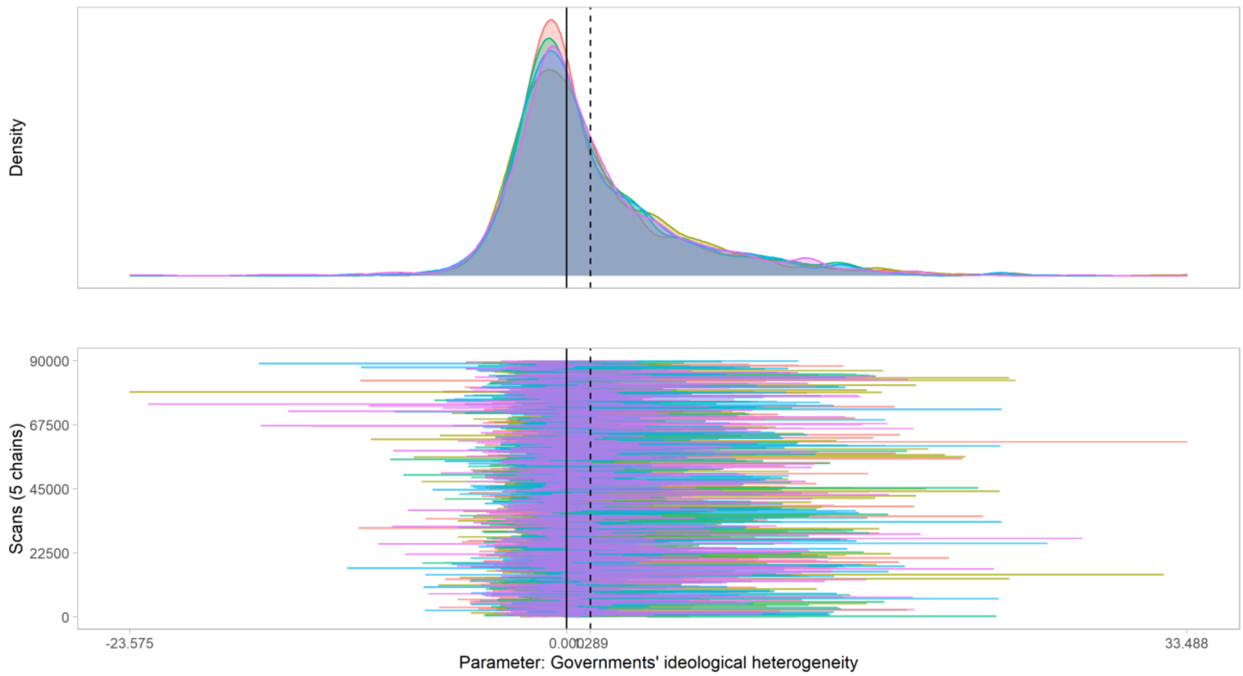
The maximum Gelman & Rubin convergence statistics across all model parameters and chains is less than 1.01, which suggests adequate convergence. The absolute value of the related Geweke z-score is smaller than 2 across all model parameters, which suggests that the mean of the first part and the last part of the chains are the same. The p-values of the Heidelberger and Welch diagnostic are all larger than 0.05 so that the null hypothesis that the Markov Chain is in the stationary distribution cannot be rejected. Finally, the autocorrelation plots across all model parameters and chains likewise do not indicate significant autocorrelation.

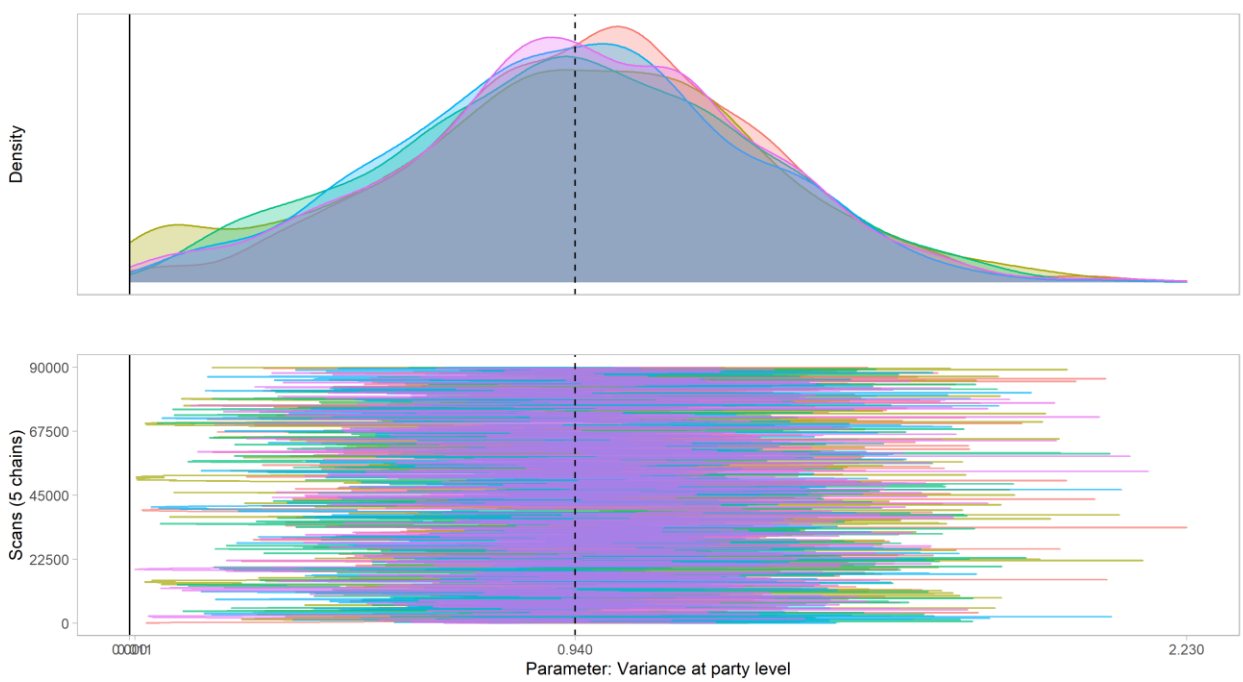
The Monet plots below show the posterior distributions and trace plot of the most important parameters. The solid line indicates zero and the dashed line shows the *arithmetic mean* over all 5 chains.

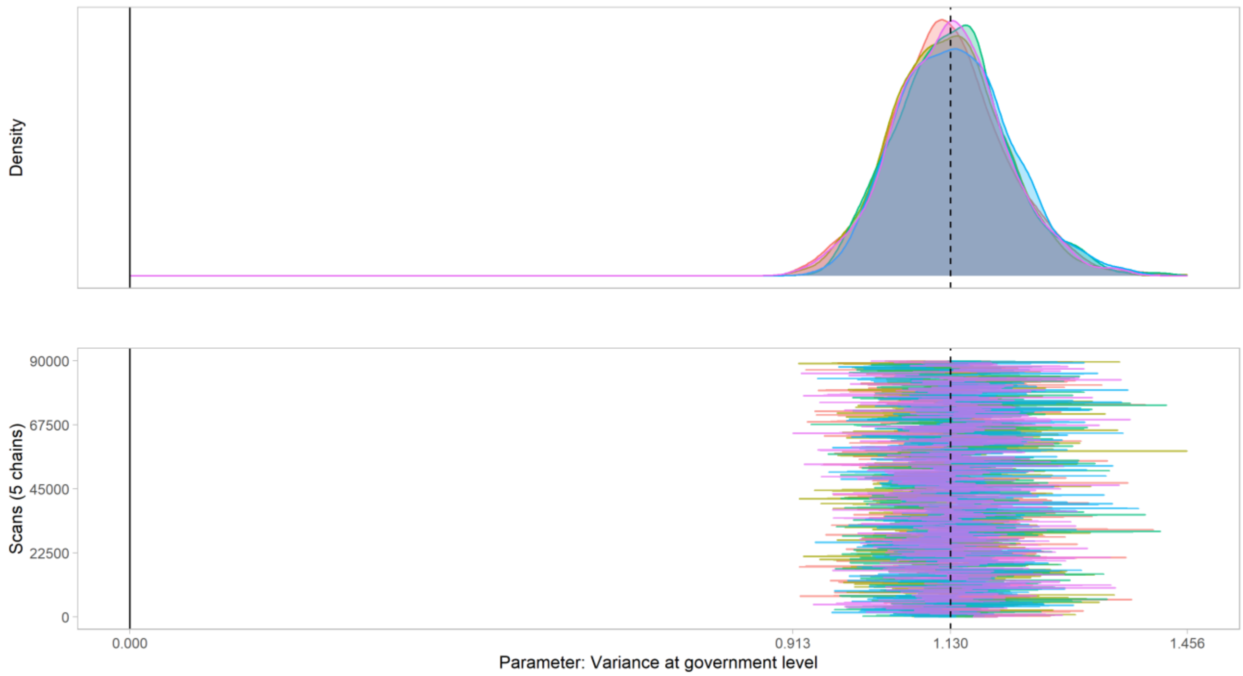
Monetplots

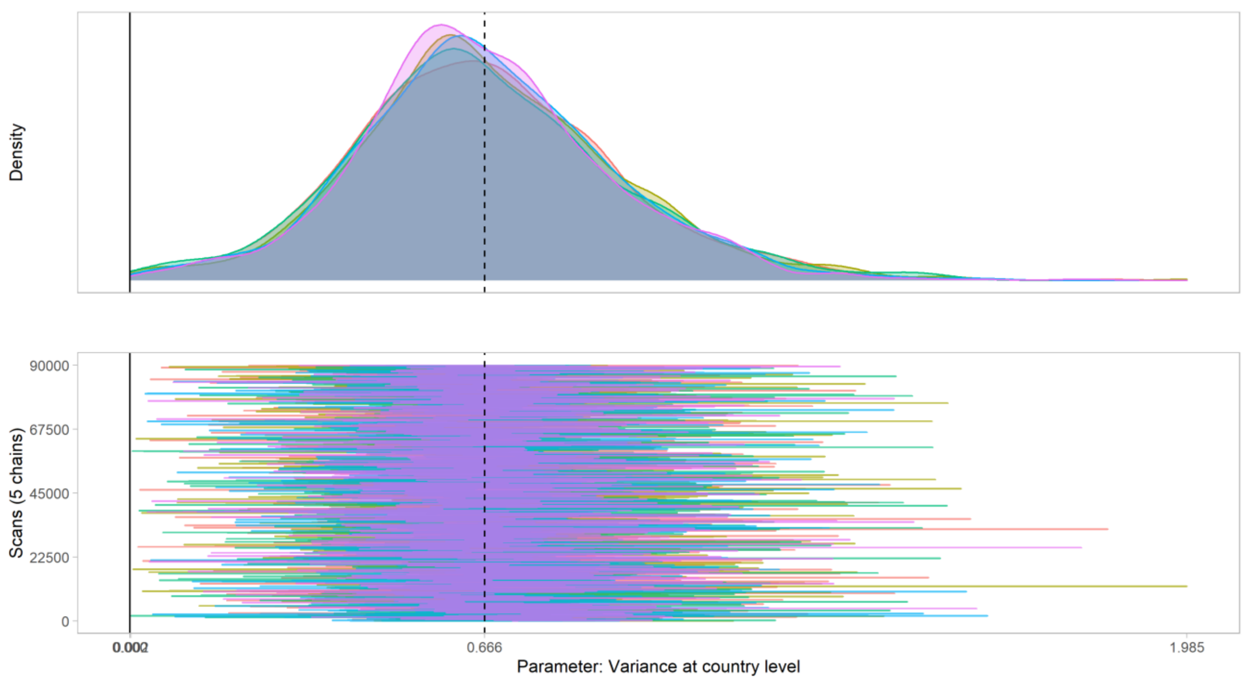

Density
Scans (5 chains)
90000
67500
45000
22500
0
-0.730
0.000
0.687
2.705
Parameter: Parties' financial dependency


Density
Scans (5 chains)
90000
67500
45000
22500
0
-25.610
0.000
8.698
36.118
Parameter: Parties' relative legislative seat share


Density
Scans (5 chains)
90000
67500
45000
22500
0
-23.575
33.488
Parameter: Governments' ideological heterogeneity


Density
Scans (5 chains)
90000
67500
45000
22500
0
0.940
2.230
Parameter: Variance at party level


Density
90000
67500
45000
22500
0
Scans (5 chains)
0.000
0.913
1.130
1.456
Parameter: Variance at government level


Density
90000
67500
45000
22500
0
Scans (5 chains)
0.002
0.666
1.985
Parameter: Variance at country level

APPENDIX A5: EXTENSIONS

Endogenizing the aggregation process makes it possible to test theoretical expectations on aggregation processes, such as the aggregation of player strategies in game-theoretic models (Lupia and Strøm 1995; Tsebelis 2002) and spatial theory (Laver and Shepsle 1996), or indices that aggregate party features to a higher level, such as power indices (Banzhaf 1964). Theoretical expectations on such processes can be tested with the MMMM by translating them into weight functions. The proposed approach also makes it possible to convert other theoretical ideas on the role of parties in coalition governments into testable models. Here I advance three ideas.

Party random effects as a random walk

We may also assume that the effects of unobserved variables change across a party's government participations by including a random walk centered at the party's random effect of its previous government participation: $u_{ij}^{P} \sim N\left(u_{G(i,j),j}^{P}, \sigma_{u^P}^{2}\right)$. The indexing function G(i,j) returns government i^* that party j participated before participating in i.

A social network perspective on government survival

We might theorize that the features of *party dyads* affect government survival. It has been argued, for instance, that the history of cooperation among the government parties is related to government survival as increasing amount of mutual trust, information about each other, and some form of asset specificity reduce bargaining complexity (Saalfeld, 2008: 358-9). Such arguments introduce a social capital perspective, which has not yet been explored empirically in depth. This perspective would recognize yet another level at which dependencies possibly arise – the nesting of party dyads in governments and, conversely, the nesting of governments in party dyads. The nesting, again, is a multiple membership structure. Explicitly modeling the effects of party dyad features, such as years of previous cooperation, would then allow to examine whether social capital in the form of the history of cooperation has a uniform effect on government survival or whether it decays over time. A suitable model would look like this:

$$\lambda_i = \lambda_0 \, exp\left(\dots + \sum_{\substack{i,j \\ j \neq l}} w_{i(j,l)} d_{i(j,l)} \right)$$

$$\text{with } d_{i(j,l)} = \boldsymbol{x}_{i(j,l)}^{D'} \boldsymbol{\beta}^{D} + u_{i(j,l)}^{D} \text{ and } w_{i(j,l)} = \delta^{T_{i(j,l)}^{D}}$$

where $i = 1, \dots, I$ indexes governments, $j = 1, \dots, J$ indexes party 1, and $l = 1, \dots, L$ indexes party 2. The hazard rate of termination of government i at time point t now also includes the aggregated impact of observed ($x_{i(jl)}^{D}$) and unobserved ($u_{i(jl)}^{D}$) *dyad-level features*. A government with, say, four parties exhibits

$\binom{4}{2} = 6$ dyads and its cooperative past would consequently be the weighted sum over the years of cooperation of these 6 dyads up until taking office together in government i at time point t. The weight is modeled in terms of a power law decay function, where $T^{D}_{i(j,l)}$ equals the time the cooperation dates back. Preliminary analyses (available on request) show a significant effect and an estimated discount parameter δ of 0.9, indicating a moderate decay of social capital over time. Tranmer and colleagues (2014) show in more detail how the MMML model can be used to incorporate network effects into our analyses.

The impact of opposition parties on government survival

So far, the focus has been on the parties in government although we might be interested in a total party effect, which would also include opposition parties. The influence of opposition parties can likewise be statistically represented in a multilevel framework. The total party effect can be split up into cross-classification of government and opposition party effects resulting in a bivariate random effect distribution (Goldstein 2011, 243–54). This approach can shed light on the nonelectoral replacement termination hazard. I expect it to be primarily structured by the existence of opportunities with opposition parties. Previous studies regress this type of termination hazard mainly on government properties although opposition attributes probably offer more explanatory power in this regard. A suitable MMML model looks like this:

$$\lambda_i = \lambda_0 \, exp\left(\dots + \sum_{j \in P(i)} w^{G}_{ij} p^{G}_{j} + \sum_{j \notin P(i)} w^{O}_{ij} p^{O}_{j} \right)$$

$$\text{with } w^{G}_{ij} = w^{G*}_{ij} \text{ and } w^{O}_{ij} = w^{O*}_{ij}$$

The effect of opposition parties p^{O}_{j} may only consist of a random effect as current coalition government datasets do not offer much information on parties who are not or never were in government. The weights w^{G}_{ij}, w^{O}_{ij} may or may not be endogenized.

REFERENCES

Andersson, Staffan, Torbjörn Bergman, and Svante Ersson. 2014. "The European representative democracy data archive, release 3."

Aronow, Peter, Cyrus Samii, and Valentina Assenova. 2015. "Cluster–Robust Variance Estimation for Dyadic Data." *Political Analysis* 23 (04): 564–77.

Bäck, Hanna, Marc Debus, and Patrick Dumont. 2011. "Who gets what in coalition governments?: Predictors of portfolio allocation in parliamentary democracies." *European Journal of Political Research* 50 (4): 441–78.

Banzhaf, John F. 1964. "Weighted voting doesn't work: A mathematical analysis." *Rutgers Law Review* 19 (2): 317–43.

Benoit, Kenneth, and Michael Laver. 2007. "Estimating party policy positions: Comparing expert surveys and hand-coded content analysis." *Electoral Studies* 26 (1): 90–107.

Berger, James. 2006. "The case for objective Bayesian analysis." *Bayesian Analysis* 1 (3).

Croon, Marcel A., and Marc J. P. M. van Veldhoven. 2007. "Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model." *Psychological methods* 12 (1): 45–57.

Gamson, William A. 1961. "A Theory of Coalition Formation." *American Sociological Review* 26 (3): 373–82.

Gelman, Andrew. 2006. "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis* 1 (3): 515–34.

Goldstein, Harvey. 2011. *Multilevel statistical models*. 4th ed. Oxford: Wiley.

Hougaard, Philip. 2000. *Analysis of Multivariate Survival Data*. *Statistics for Biology and Health*. New York, NY: Springer.

King, Gary, James Alt, Nancy Burns, and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American journal of political science* 34: 846–71.

Laver, Michael, and Kenneth A. Shepsle. 1996. *Making and breaking governments: Cabinets and legislatures in parliamentary democracies*. Cambridge: Cambridge University Press.

Leenders, Roger T. 2002. "Modeling social influence through network autocorrelation: Constructing the weight matrix." *Social Networks* 24 (1): 21–47.

Lin, D. Y., and L. J. Wei. 1989. "The Robust Inference for the Cox Proportional Hazards Model." *Journal of the American Statistical Association* 84 (408): 1074–78.

Lupia, Arthur, and Kaare Strøm. 1995. "Coalition Termination and the Strategic Timing of Parliamentary Elections." *American Political Science Review* 89 (3): 648–65.

Müller, Wolfgang C., and Kaare Strøm, eds. 1999. *Policy, office, or votes?: How Political Parties in Western Europe Make Hard Decisions*. Cambridge: Cambridge University Press.

Neuhaus, J. M., J. D. Kalbfleisch, and W. W. Hauck. 1991. "A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data." *International Statistical Review* 59 (1): 25.

Newman, Daniel A., and Hock-Peng Sin. 2009. "How Do Missing Data Bias Estimates of Within-Group Agreement?" *Organizational Research Methods* 12 (1): 113–47.

Plummer, Martyn. 2015. *JAGS Version 4.0.0 user manual*.

Scarrow, Susan, Thomas Poguntke, and Paul Webb, eds. 2017. *Organizing Representation. Political Parties, Participation, and Power*. Oxford: Oxford University Press.

Seki, Katsunori, and Laron K. Williams. 2014. “Updating the Party Government data set.” *Electoral Studies* 34: 270–79.

Snijders, Tom. 2016. “The Multiple Flavours of Multilevel Issues for Networks.” In *Multilevel network analysis for the social sciences*, eds. Emmanuel Lazega and Tom Snijders. Springer, 15–46.

Strøm, Kaare. 1990. “A Behavioral Theory of Competitive Political Parties.” *American journal of political science:* 565–98.

Tranmer, Mark, David Steel, and William J. Browne. 2014. “Multiple-membership multiple-classification models for social network and group dependences.” *Journal of the Royal Statistical Society: Series A* 177 (2): 439–55.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton: Princeton University Press.

Volkens, Andrea, Pola Lehmann, Theres Matthieß, Nicolas Merz, and Sven Regel. 2016. “The Manifesto Data Collection. Version 2016b.” Berlin: WZB.

Warwick, Paul. 1994. *Government Survival in Parliamentary Democracies*. Cambridge: Cambridge University Press.

Woldendorp, Jaap, Hans Keman, and Ian Budge. 2000. *Party Government in 48 Democracies (1945 -1998): Composition-Duration-Personnel*. Dordrecht: Springer.

SUPPLEMENTARY MATERIAL CHAPTER 3

A DATA

Appendix A summarizes the operationalization of the employed variables (A1), describes the amount of missing values and the imputation procedure (A2), reports descriptive statistics for both samples (A3), and discusses the use of survey weights (A4).

A1 Operationalization of all employed variables

***Table A1**. Operationalization of all employed variables*

Variable	Groups
Socioeconomic status (SES)	Add Health's pre-constructed SES score is on a z-scale and based on a principal component analysis using as input parental education, parental occupation, parental income, and whether the student receives free lunches. I differentiate students globally (across schools) into low-, medium-, high-SES at the tercile cutoff points. This operationalization is discussed in Appendix B1.
Parental education	I categorize parental education up to high school as low, some college as medium, and at least a 4-year college degree as high.
Parental occupation	Based on the Goldthorpe class typology (Goldthorpe, Llewellyn, and Payne 1980), I categorize parental occupation into low (e.g., unemployed or unskilled), medium (e.g., skilled manual and supervisory), and high (e.g., managers, professionals).
Parental income	I categorize total household income in 1994 dollars into low, medium, high at the tercile cutoff points.
Race	Add health's pre-constructed variable categorizes students into White, Black, Native American, Asian, and Hispanic.
Gender	Add Health's pre-constructed variable categorizes students as girls and boys.
Language spoken at home	Add Health categorizes language spoken at home into English, Spanish, or another language. This categorization is coarse and will, therefore, label some friendship ties as same-language while they are in fact cross-language. As a consequence, this differentiation will overestimate the amount of language-based segregation.
GPA	GPA is averaged across English, Math, Social Science, and Science. To differentiate between same- and cross-GPA friendship, I categorize GPA into low, medium, high at the tercile cutoff points. For the most part, grades are reported by students themselves. For 11% of students, I could take this information directly from their transcripts.
Residential proximity	Add Health includes students' residential locations and neighborhoods, which are placed on an artificial map with the school as a reference point for privacy protection. I calculate the log-Euclidean distance between students.
Overlap in extracurricular activities	Students indicated their participation in a list of 20 possible school clubs (e.g., high school band) and 11 sports teams (e.g., basketball). I measure the degree of overlap of pairs of students as the number of extracurricular activities in which they participate together.
Course overlap	AHAA measures course overlap of pairs of students as the number of courses taken together, weighted by course size and contact time. This score operationalizes the degree to which students share a social and academic space in school by virtue of their course taking patterns (Barber and Wasson 2014; Frank et al. 2008).

A2 Missing values and imputation

Exponential random graph modeling requires that the features of all nodes are observed. As listwise deletion would remove any nodes with incomplete information from the network, I employ multiple imputation using predictive mean matching to eliminate missing values in the final sample. This is an important step because removed students could systematically differ from other students in the network and because removing students from the network would distort the analyzed network structure. That is, the frequency of dyads, triads, and other higher-order network structures would be altered if students were removed from the networks. Distorted network structures would be particularly problematic for the purposes of this paper in which relational mechanisms are examined.

Fortunately, the response rate of the inschool survey in which students identified their friends is high (80 percent on average in the considered schools) and item nonresponse with regard to most considered features is low (see Table A2). The exception is students' course overlap for which about 40 percent of all student pairs lack data. The reason is that the AHAA expansion was carried out during wave III and transcripts (of the wave I school year) are thus only available for the subset of students that also participated in wave III. The data are therefore not missing completely at random but at random conditional on the usual suspects of panel attrition (e.g., race, SES background, school type, and region).[24]

In this context, multiple imputation has shown to be effective at reducing bias and increasing power and may (currently) be the most reliable approach to handle missing data in network analyses (Krause et al. 2020; Smith, Morgan, and Moody 2022). I draw on the multiple imputation by chained equations as implemented in the R package 'mice' (van Buuren and Groothuis-Oudshoorn 2011). I use predictive mean matching because it tends to outperform other methods and can accommodate both categorical (e.g., race) and continuous (e.g., SES score) variables. I create M=10 imputations and estimate separate ERGMs on each imputation sample. I then generate 1,000,000 synthetic networks from the fitted models (100,000 networks per model). Table A2 details the amount of observed and missing values. Note that the proportion of missing data provides limited information about bias reduction and efficiency gains that can be made from multiple imputation (Madley-Dowd et al. 2019). Crucial instead is how informative auxiliary variables are of the patterns of missingness in the data. In the following, I discuss the specification of the imputation model in detail. Both sample I and II contain missing data but I focus on sample II in the following discussion because the variables used in sample I contain few missing values.

[24] Since Add Health oversampled minorities in the inschool but not in the inhome survey and because minorities are more likely to drop out from panel studies, students for which residential locations and course selections are unobserved are more likely nonwhite and from lower SES background. Accordingly, missingness in the final sample is not completely at random and studies employing these data using listwise deletion (e.g., Frank et al. 2008; Mouw and Entwisle 2006) examined networks that, to some degree, lacked majority students' minority friends.

Imputation model for nodal features

I use the total sample (N=91,018) to predict missingness in the final sample (N=5,942) to maximize the information in the imputation model. The imputation model for nodal features includes the following variables: students' age, grade, gender, race, SES, GPA, personality (big five), well-being in school (averaged across 10 items), extracurricular activities, household size, language spoken at home, indicator whether mother is working, rating how much mother cares about them, neighborhood and school mean and variance in SES, and rating how safe they feel in their neighborhood. The auxiliary variables are highly predictive of key variables of interest (e.g., $R^2 = 0.71$ when regressing SES on all auxiliary variables). Trace plots suggest that the estimation algorithm has converged.

Imputation model for edge features

I use the total sample (N=42,574,827) to predict missingness in the final sample (N=3,133,824) to maximize the information in the imputation model. The samples are based on all possible combinations of pairs of students who attend the same school, which results in N(N-1)/2 cases per school as the unit of analysis is the undirected dyad. The imputation model for edge features includes the following variables: school area (rural, suburban or urban), school size, students' absolute difference in age, grade, GPA, SES, well-being in school, household size, neighborhood safety, whether students are of the same gender, race, and speak the same language at home, as well as their course overlap, overlap in extracurricular activities, and neighborhood proximity. For all quantitative variables, I also include their quadratic effect and an interaction with the lower value in the dyad. The auxiliary variables are highly predictive of key variables of interest and trace plots suggest that the estimation algorithm has converged.

Table A2*. Number of observed and missing data.*

Variable	N observed	N missing/imputed	% missing
Socioeconomic status	19,324	1,953	9.2
Parental education	66,651	561	0.8
Parental occupation	60,164	833	1.4
Parental income	15,351	3,040	16.5
Race	88,143	88	0.1
Subracial categories	82,211	1,213	1.5
Gender	89,812	0	0
Language spoken at home	20,737	1,634	7.3
GPA	76,920	1,083	1.4
Big 5	15,675	2,588	14.2
Extracurricular activities	85,627	1,071	1.2
Distance between residential locations	3,271,341	602,942	15.6
Overlap extracurricular activities	42,574,827	0	0
Overlap course selections	1,730,490	1,078,752	38.4

A3 Descriptive statistics

Table A3 provides descriptive statistics of sample I (core sample, weighted statistics) and Table A4 provides descriptive statistics of sample II (saturated sample, unweighted statistics). A comparison shows that there are few differences in the demographic makeup of the two samples.

Table A3. *Descriptive statistics sample I (Add Health core sample, weighted statis).*

Variable	M	SD	Min, Max	N
SES score	-0.03	1.2	-5.59; 3.51	83,444
Parental education	Low: 0.411 Med: 0.322 High: 0.267			83,444
Parental occupation	Low: 0.336 Med: 0.193 High: 0.471			83,444
Parental income (in 1994 dollars)	44,535	43,014	0; 999,000	83,444
Age	14.7	1.71	10; 19	83,444
Race	White: 0.703 Black: 0.160 Native: 0.011 Asian: 0.040 Hispanic: 0.086			83,444
Gender	Boys: 0.498 Girls: 0.502			83,444
Language spoken at home	English: 0.923 Spanish: 0.054 Other: 0.023			83,444
GPA	2.79	0.833	0; 4	83,444
Outdegree	2.74	2.57	0; 10	83,444
Indegree	2.68	2.84	0; 27	83,444

Table A4. *Descriptive statistics for sample 2 (Add Health schools 2, 3, 6, 7, 8, 12, 28, 58, 77, 81, 88, 90, 106, 126, 194, 369).*

Variable	M	SD	Min, Max	N
SES score	0.05	1.15	-4.33; 3.51	5,942
Parental education	Low: 0.404 Med: 0.306 High: 0.290			5,942
Parental occupation	Low: 0.360 Med: 0.207 High: 0.433			5,942
Parental income (in 1994 dollars)	47,199	57,115	0; 999,000	5,942
Age	15.1	1.52	10; 19	5,942
Race	White: 0.727 Black: 0.150 Native: 0.011 Asian: 0.030 Hispanic: 0.082			5,942
Gender	Boys: 0.501 Girls: 0.499			5,942
Language spoken at home	English: 0.952 Spanish: 0.035 Other: 0.013			5,942
GPA	2.75	0.83	0; 4	5,942
Outdegree	3.00	2.53	0; 10	5,942
Indegree	3.10	2.90	0; 21	5,942
Overlap: Courses (# of courses, centered and weighted by course size and contact time)	0.081	0.310	-0.10; 4.57	3,133,824
Overlap: Extracurriculars (# of extracurricular activities, centered)	0.076	0.340	-0.10; 3.37	3,133,824
Overlap: Neighborhoods (log-Euclidean distance in meters, centered)	0.172	1.18	-8.14; 3.92	3,133,824

A4 Survey weights for dyadic data

Our knowledge about segregation patterns in friendship networks is based, on the one hand, on samples that are confined single locations or regions (e.g., Wimmer and Lewis 2009; McMillan 2022) and, on the other hand, on samples that are nationally representative (e.g., Add Health, CILS4EU). To generalize results to the national population, however, the latter data require design weights because the sample is stratified and select groups are oversampled. Using race as an example, Table 2 shows that the racial distribution in Add Health deviates substantially from the proportions reported in NCES (1998a, 1998b) without using survey weights. When using survey weights, in contrast, the racial distribution is approximated much better.

Table A5: *The racial distribution in Add Health with and without survey weights.*

	NCES 1995 estimate[25]	Unweighted sample (sample 1)	Weighted sample (sample 1)
White	67.2	56.9	70.3
Black	16.0	17.0	16.0
Native American	1.0	3.5	1.1
Asian	3.8	5.9	4.0
Hispanic	12.0	16.7	8.6

Most network studies do not make use of survey weights, mainly because popular network formation models (e.g., ERGM) have not been extended to accommodate survey weights. As the analyses in this paper are simulation-based, I can weigh each dyad by its inclusion probability in the sample. Regardless of whether or why a dyad exists, it's inclusion probability in the sample is the product of the inclusion probability of the involved students (Chantala 2001). Accordingly, the weight of a dyad equals $w_{\text{dyad}} = w_{\text{ego}} \cdot w_{\text{alter}}$ when total sample weights are used and $w_{\text{dyad}} = \frac{w_{\text{ego}} \cdot w_{\text{alter}}}{w_{\text{school}}}$ when poststratification weights are employed. I validated this weighting approach with a small simulation study (available upon request) in which I sample units unequally from a population and create networks among the sampled units. Weighing dyads by the product of the unit weights yields the same proportions as when calculated from the population networks.

For all analyses involving sample 1, I make use of the grand sample weight in Add Health (GSWGT1), which corrects both for selection and nonresponse bias. This weight is available for Add Health's core sample and based on student and school attributes (race, SES background, school type, and region). Since not all students in the friendship network are part of the core sample, I predict their grand sample weight based on the same school and student attributes and then rescale the weights such that the distribution of weights in the full sample (sample 1) matches the distribution of weights in the core sample. Table A6

[25] Private elementary and secondary school enrollment by race in 1995: White: 77.9, Black: 9.2, Native: 0.4, Asian: 4.6, Hispanic: 8.0. Public elementary and secondary school enrollment by race in 1995: White: 64.8, Black: 16.8, Native: 1.1, Asian: 3.7, Hispanic: 13.5. Total private-school enrollment: 5,032,200. Total public-school enrollment: 44,840,000.

shows that this procedure works quite well. The proportion of each dyad type in the socioeconomic mixing matrix is close to its mathematical expectation (11.1%) when weighing dyads by the product of students' weights in the completely random network. Without weighing dyads, in contrast, the proportions deviate from the mathematical expectation because medium-SES students are somewhat overrepresented in the sample.

Table A6: *Proportion of each dyad type in the socioeconomic mixing matrix with and without weighing dyads in the completely random network.*

Tie combination	low→ low	low→ med	low→ high	med→ low	med→ med	med→ high	high→ low	high→ med	high→ high
unweighted	6.9	11.1	8.2	11.1	18	13.3	8.2	13.3	9.8
weighted	11.1	11.1	11.1	11.1	11.1	11.1	11.1	11.1	11.1

I cannot use survey weights for any of the analyses involving sample 2 since the saturated sample is a nonrandom subset of Add Health's core sample. Effect estimates are therefore not population-averaged associations but average associations in the sample. However, comparisons show that students in both samples exhibit similar sociodemographic profiles and their friendship networks are similarly segregated along sociodemographic lines (see Appendix A3).

B DESCRIPTION OF SOCIOECONOMIC SEGREGATION IN HIGH SCHOOL FRIENDSHIP NETWORKS

Appendix B examines the robustness of the descriptive results. The section analyzes in particular the robustness of the results with respect to different operationalizations of SES background and definitions of friendship.

Figure B1 depicts the density of Add Health's pre-constructed SES score. Figures B2 to B5 replicate Figure 2 with the following cutoff points: 50/50, 33/66, 25/75, and 25/50/75 percentiles. The figures demonstrate that the key pattern – the unilateral exclusion of students at the bottom and closure among students in the upper part of the SES distribution – is evident regardless of the exact cutoff points.

Figure B6 shows that the segregation pattern is also similar when only reciprocated friendship ties are analyzed with some nuances: segregation is greater for low→low than for high→high ties. This is because low-SES students are excluded both by medium- and high-SES students and thus reciprocated ties are likely only among themselves. In contrast, for high-SES students, reciprocated ties are likely not only amongst themselves but anyone else they nominate. Therefore, the reason why segregation is greater for low→low than for high→high ties is that reciprocated ties do not differentiate exclusion and closure.

Finally, Figure B7 shows that this key pattern is more difficult to appreciate when the continuous SES score is used to examine socioeconomic segregation as linear distance between friends' socioeconomic background. The red line in the figure depicts the observed linear distances. The black line depicts the linear distances expected if students selected friends at random. Comparing them shows that small distances between friends' SES background are more frequent than expected and large distances are less frequent than expected. However, the linear analysis underestimates the true segregation. The reason is that the linear distance measure combines low↔med and med↔high ties and ties with the same distance but different directions (e.g., low→high and high→low ties) even though their tie probabilities differ systematically. Therefore, the operationalization of socioeconomic segregation in prior studies masks the unilateral and nonlinear segregation pattern to some extent.

Figure B1: *Density of Add Health's SES score.*

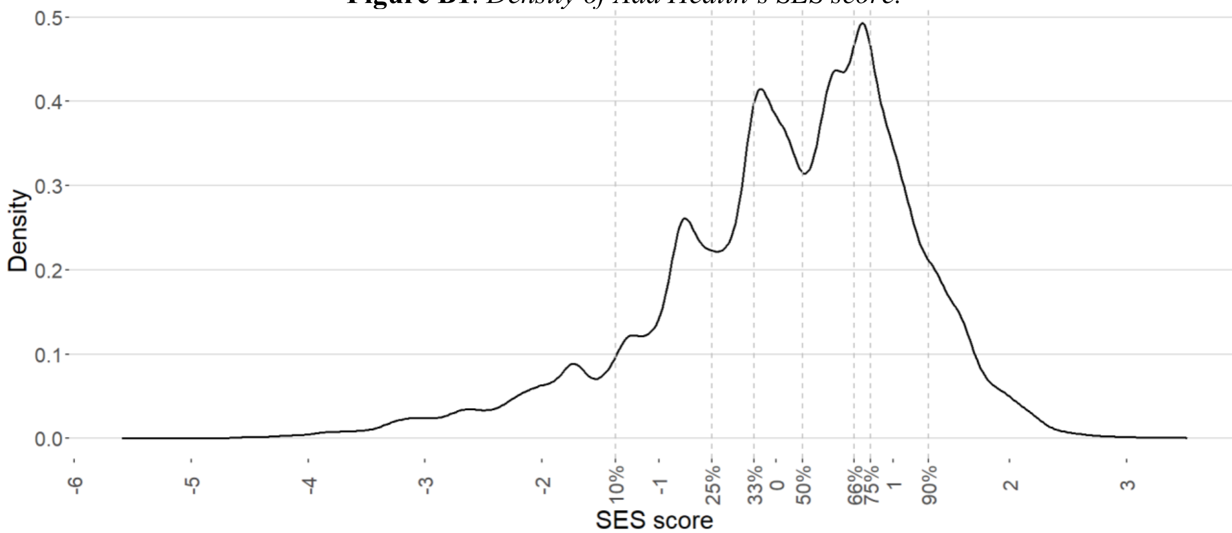
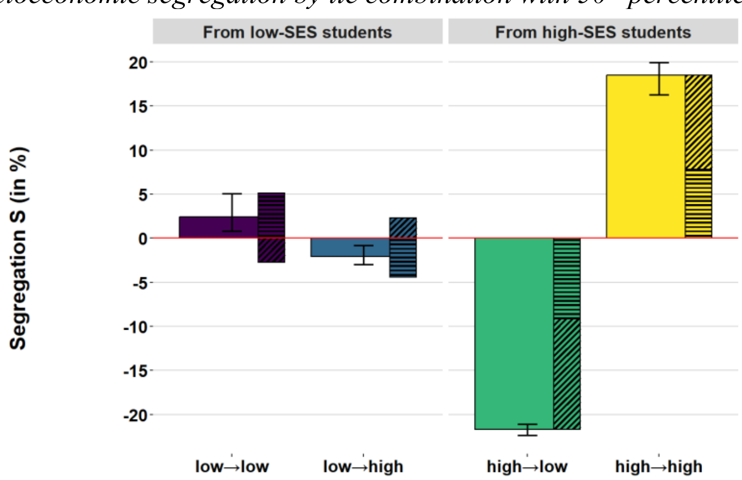
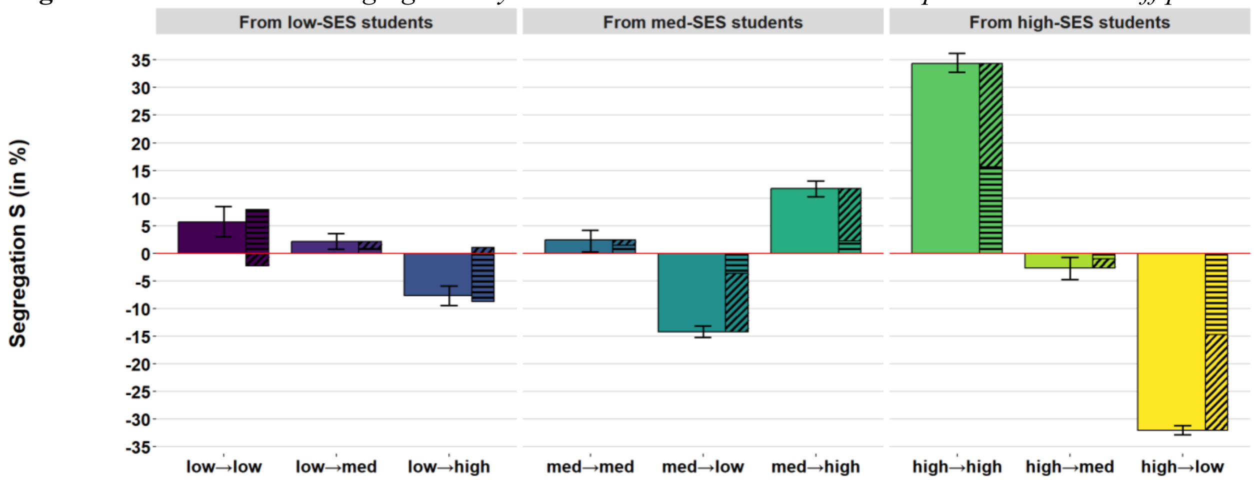


Figure B2: *Socioeconomic segregation by tie combination with 50th percentile as cutoff point.*



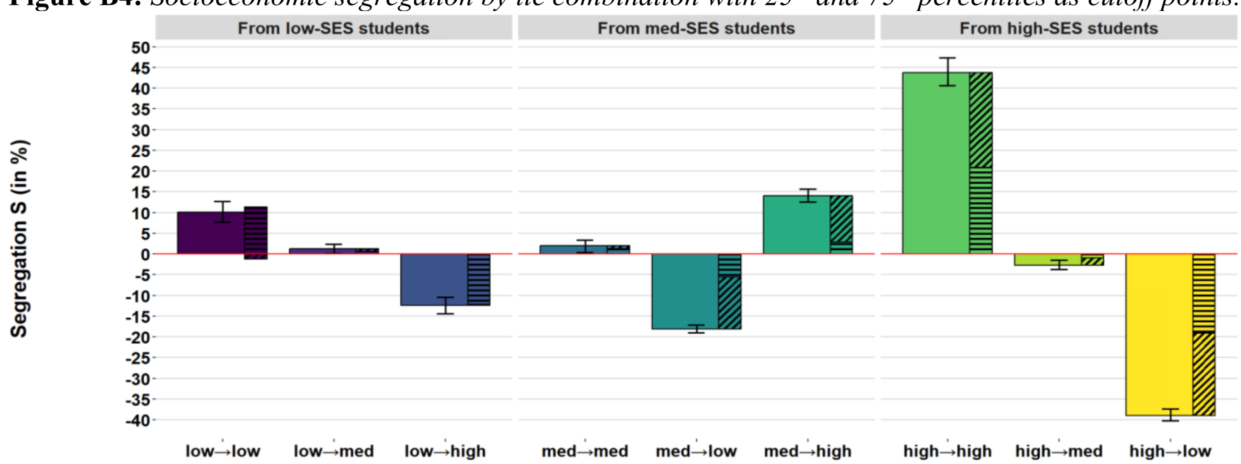
	low→low	low→high	high→low	high→high
Expected	19.4	22.8	26.6	31.2
95% PI	[18.9, 19.7]	[22.5, 23]	[26.4, 26.8]	[30.8, 31.8]
Expected within	20.4	21.8	24.2	33.7
95% PI	[19.8, 20.8]	[21.4, 22.3]	[21.6, 24.9]	[32.7, 36.3]
Observed	19.9	22.3	20.8	37
S (within)	-2.7	2.3	-12.5	10.7
S (between)	5.2	-4.4	-9.2	7.8

Figure B3: *Socioeconomic segregation by tie combination with 33rd and 66th percentiles as cutoff points.*



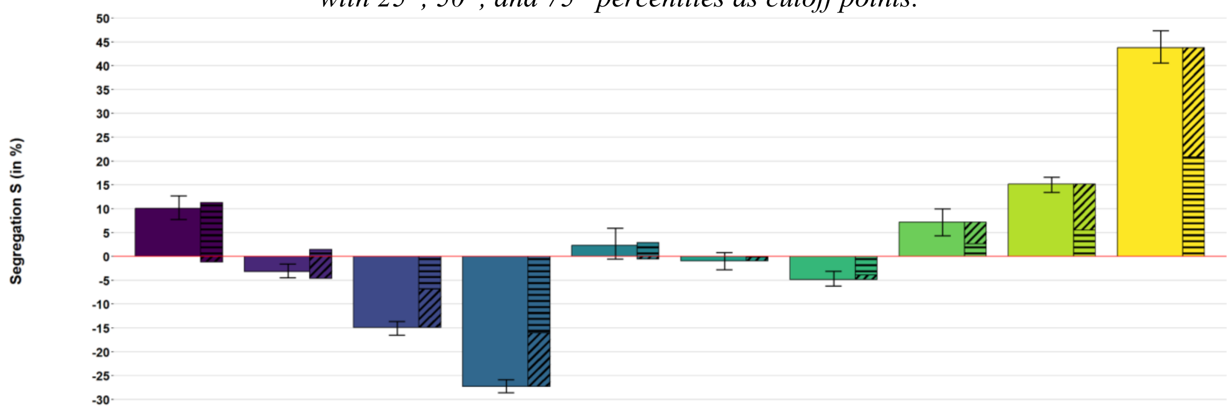
	low→low	low→med	low→high	med→med	med→low	med→high	high→high	high→med	high→low
Expected	9.8	9.9	9.9	11.2	11.1	11.2	12.4	12.4	12.2
95% PI	[9.5, 10]	[9.7, 10]	[9.7, 10.1]	[11, 11.5]	[11, 11.2]	[11.1, 11.4]	[12.2, 12.5]	[12.1, 12.6]	[12.1, 12.4]
Expected within	10.5	10	9	11.4	10.7	11.5	14.3	12.2	10.4
95% PI	[10.2, 10.8]	[9.8, 10.1]	[8.8, 9.2]	[11.2, 11.6]	[9.8, 11]	[11.2, 12.3]	[13.8, 16.1]	[12, 12.5]	[8.8, 10.9]
Observed	10.3	10.1	9.1	11.5	9.5	12.5	16.6	12	8.3
S (within)	-2.3	1.3	1	1	-10.6	9.5	18.7	-1.6	-17.3
S (between)	7.9	0.9	-8.7	1.4	-3.7	2.2	15.7	-1.1	-14.8

Figure B4: *Socioeconomic segregation by tie combination with 25th and 75th percentiles as cutoff points.*



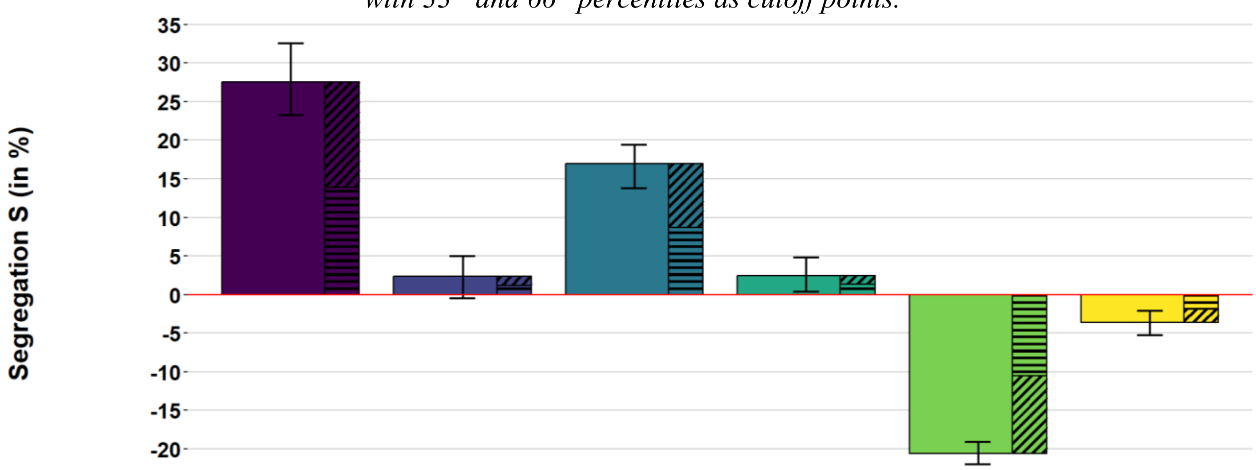
	low→low	low→med	low→high	med→med	med→low	med→high	high→high	high→med	high→low
Expected	5.4	10.8	5.4	25.1	12.5	12.6	7.1	14.1	7
95% PI	[5.3, 5.5]	[10.7, 10.9]	[5.3, 5.6]	[24.8, 25.5]	[12.3, 12.6]	[12.4, 12.8]	[6.9, 7.2]	[13.9, 14.2]	[6.8, 7.1]
Expected within	6	10.9	4.8	25.4	11.8	13	8.6	13.9	5.7
95% PI	[5.9, 6.1]	[10.7, 11]	[4.7, 4.9]	[25.1, 25.9]	[10.5, 12.2]	[12.6, 14.1]	[8.1, 9.8]	[13.7, 14.1]	[4.6, 6]
Observed	5.9	11	4.8	25.6	10.2	14.4	10.2	13.7	4.3
S (within)	-1.2	0.7	-0.3	0.6	-12.6	11.1	22.9	-1.7	-19.8
S (between)	11.3	0.5	-12.1	1.3	-5.5	2.9	20.8	-1	-19.1

Figure B5: *Socioeconomic segregation by tie combination with 25th, 50th, and 75th percentiles as cutoff points.*



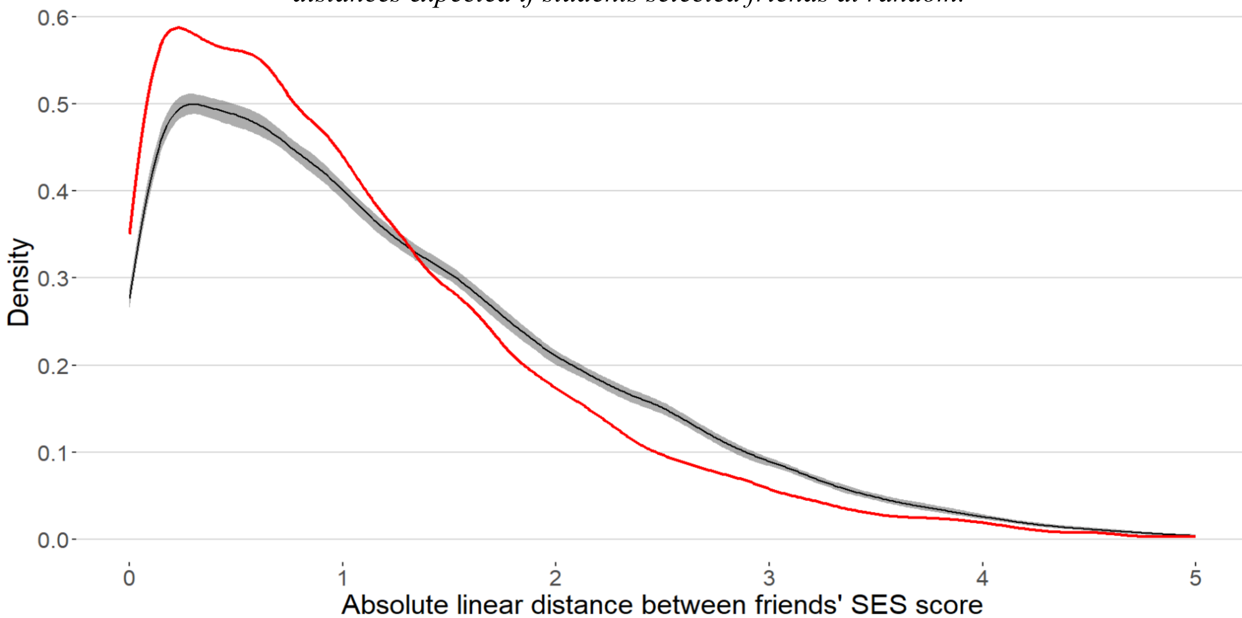
	low↔low	low↔medlow	low↔medhigh	low↔high	medlow↔medlow	medlow↔medhigh	medlow↔high	medhigh↔medhigh	medhigh↔high	high↔high
Expected	5.4	11.5	11.8	12.4	6.1	12.6	13.2	6.4	13.5	7.1
95% PI	[5.3, 5.5]	[11.3, 11.7]	[11.6, 12]	[12.2, 12.6]	[5.9, 6.3]	[12.3, 12.8]	[13, 13.4]	[6.2, 6.6]	[13.3, 13.7]	[6.9, 7.2]
Expected within	6	11.7	11	10.4	6.3	12.5	12.7	6.6	14.2	8.6
95% PI	[5.9, 6.1]	[11.3, 11.9]	[10.2, 11.3]	[9.3, 10.8]	[6.1, 6.5]	[12.3, 12.8]	[12.4, 12.9]	[6.4, 6.8]	[13.8, 15.2]	[8.1, 9.8]
Observed	5.9	11.2	10	9	6.3	12.4	12.6	6.9	15.5	10.2
S (within)	-1.2	-4.6	-8.1	-11.3	-0.6	-0.8	-0.9	4.6	9.6	22.9
S (between)	11.3	1.5	-6.9	-16.1	2.9	-0.2	-3.9	2.6	5.6	20.8

Figure B6: *Socioeconomic segregation in friendship networks of reciprocated ties by tie combination with 33rd and 66th percentiles as cutoff points.*



	low↔low	med↔med	high↔high	low↔med	low↔high	med↔high
Expected	7.2	11.2	15.8	17.9	21.3	26.6
95% PI	[6.9, 7.4]	[10.9, 11.5]	[15.5, 16.2]	[17.5, 18.3]	[20.9, 21.7]	[26.2, 27.1]
Expected within	8.2	11.3	17.1	18.2	19.1	26.1
95% PI	[7.9, 8.7]	[11.1, 11.7]	[16.8, 17.8]	[17.8, 18.6]	[18, 19.5]	[25.7, 26.6]
Observed	9.2	11.5	18.5	18.4	16.9	25.6
S (within)	13.6	1.2	8.3	1.1	-10.1	-1.7
S (between)	13.9	1.2	8.7	1.3	-10.5	-1.9

Figure B7: *Socioeconomic segregation as linear distance between friends' SES background as a continuous SES score. The red line depicts the observed linear distances. The black line depicts the linear distances expected if students selected friends at random.*



C ANALYSIS OF DETERMINANTS OF SOCIOECONOMIC SEGREGATION

Appendix C provides details for the analysis of determinants of socioeconomic segregation. Section 1 offers a comprehensive review of prior studies. Section 2 presents measures of correlation between SES and third variables to better understand their impact on socioeconomic segregation. Sections 3-6 offer more details on the ERGM results (specification details, full results, robustness checks, and convergence and goodness-of-fit statistics).

C1 Review of prior work

Table C1: *Review of prior work on the determinants of racial and socioeconomic segregation in high school friendship networks.*

Paper	Summary	Stratified settings	Homophily, Popularity, Sociality	Relational mechanisms	Generalizability
Current study	Determinants of socioeconomic segregation in high school friendship networks. Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: Directed dyad $N_{schools} = 128, N_{nodes} = 83{,}000$ Method: ERGM simulations to directly measure the impact of tie-formation mechanisms on socioeconomic segregation	Courses, Extracurriculars, Neighborhoods Not included: Parental networks, Primary / middle school networks	*Homophily*: SES, GPA, race, gender, grade *Popularity*: SES, GPA, race *Sociality*: Students are constrained to have the same outdegree as observed.	Reciprocity, Triadic closure, Preferential attachment, Homophily * reciprocity, Homophily * triadic closure	Partly nationally representative of students in US high schools in 1994/5 → Descriptive analyses are weighted (AH grand sample weight) → Decomp. of determinants within schools is unweighted
Moody 2001	School-level moderators of racial segregation Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: School $N_{schools} = 130$ Method: Linear regression	Extracurriculars, Neighborhoods Not included: Courses	None	None	Nationally representative of US high schools in 1994/5 (AH school weights).
Joyner and Kao 2001	School-level moderators of racial segregation Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: Student $N_{schools} = 134, N_{nodes} = 78{,}007$ Method: Logistic regression	None	*Homophily:* race *Popularity*: none *Sociality*: none	None	Sample of students in US high schools in 1994/5 (no survey weights)
Mouw & Entwisle 2006	Neighborhood effects on racial segregation in in high school. Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: Directed dyad $N_{schools} = 134, N_{nodes} = 14{,}500$ Method: Quasi ERGM	Neighborhoods Not included: Courses, Extracurriculars	*Homophily*: SES, race Not included: GPA *Popularity*: none *Sociality*: none	Reciprocity, Triadic closure, Preferential attachment	Partly nationally representative of students in US high schools in 1994/5 → Decomposition of within- and between-school components is weighted based on NCES data → Within-school analysis is unweighted
Goodreau et al. 2009	Effects of relational mechanisms on racial segregation in high school. Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: Undirected dyad $N_{schools} = 59, N_{nodes} = 63{,}408$ Method: ERGM	None	*Homophily:* race, gender, grade *Degree:* same as for homophily Not included: SES, GPA	Triadic closure Not included: Preferential attachment	Sample of students in US high schools in 1994/5 (no survey weights)

Wimmer & Lewis 2009	Determinants of racial segregation at uni. Data: Facebook friendships of one cohort at Texas A&M (year: 2009) Type: Cross-sectional Unit of analysis: Directed dyad $N_{nodes} = 1{,}640$ Method: ERGM	Same major Same neighborhood, residence, room Not included: Courses, Extracurriculars	*Homophily:* SES, race and sub-racial groups, cultural preferences Not included: GPA *Popularity:* none *Sociality:* race Not included: SES, GPA	Reciprocity, Triadic closure Not included: Preferential attachment	Facebook friendships at Texas A&M in 2009 (full population, one cohort)
Mc Farland et al. 2014	Contextual moderators of segregation in HS. Outcomes: Race, gender, and age segregation Data: Add Health (year: 1994/5) Type: Cross-sectional and longitudinal Unit of analysis: Directed dyad $N_{schools} = 129, N_{nodes} = 75{,}122$ Method: ERGM meta-analysis	Extracurriculars Not included: Courses, Neighborhoods	*Homophily*: race, SES, GPA, age, gender, grade *Popularity:* students are constrained to have a similar indegree as observed. *Sociality:* same as with popularity	Reciprocity, Triadic closure Not included: Preferential attachment	Sample of students in US high schools in 1994/5 → no survey weights (meta-analysis weighs by parameter precision, not by population weight)
Hofstra et al. 2017	Outcomes: Ethnic and gender segregation in online and offline networks of Dutch youth. Data: Facebook and CILS4EU (year: 2011) Type: Cross-sectional Unit of analysis: Reciprocated dyad (?) $N_{schools} = 112, N_{nodes} = 2{,}549$ Method: Multilevel linear regression	School and classroom composition (on segregation in online networks)	*Homophily:* ethnicity, gender Not included: SES, GPA *Popularity:* none *Sociality:* none	Preferential attachment Not included: Reciprocity, Triadic closure	Sample of students in Dutch secondary schools in 2011 (no survey weights)
Kruse 2017	Neighborhood effects on ethnic segregation in German secondary school classrooms. $N_{classrooms} = 144, N_{nodes} = 2{,}393$ Data: CILS4EU (year: 2011) Unit of analysis: Directed dyad Type: Cross-sectional Method: Linear regression	Neighborhoods	None	None	Sample of students in German secondary schools in 2011 (no survey weights)
Malacarne 2017	Individual and school-level determinants of cross-SES and cross-race ties in high school. Data: Add Health (year: 1994/5) Type: Cross-sectional Unit of analysis: Directed dyad $N_{schools} = 132, N_{nodes} = 11{,}033$ Method: Logistic regression	Extracurriculars, Neighborhoods Not included: Courses	*Homophily:* race, SES Not included*: GPA* *Sociality:* race, SES, GPA *Popularity:* none	None	Nationally representative of students in US high schools in 1994/5 (AH grand sample weights)
Kruse & Kroneberg 2019	Ethnic identity and ethnic segregation in German secondary school classrooms. Data: CILS4EU (year: 2011) Unit of analysis: Directed dyad Type: Cross-sectional $N_{classrooms} = 150, N_{nodes} =$ not stated Method: ERGM meta-analysis	None	*Homophily:* ethnicity, gender Not included: SES, GPA *Popularity*: none *Sociality*: none	Reciprocity Triadic closure Not included: Preferential attachment	Sample of students in German secondary schools in 2011 (no survey weights)
An 2022	Segregation patterns in Chinese middle school classrooms. Outcomes: Gender, height, age, SES, personality, academic performance, smoking Data: ANSR (year: 2010/1) Unit of analysis: Directed dyad	Same grade, Same classroom Not included: Courses, Extracurriculars, Neighborhoods	*Homophily:* SES, GPA (rank), gender, age, height, smoking, personality (optimism) *Popularity*: same as for homophily *Sociality:* same as for homophily	Reciprocity Triadic closure Preferential attachment	Non-random sample of 6 middle schools in central China in 2010/1

	Type: Cross-sectional $N_{schools} = 6, N_{nodes} = 4{,}094$ Method: ERGM meta-analysis		Not included: ethnicity		
Chetty et al 2022	Determinants of socioeconomic segregation in online friendship networks in the US. Data: Facebook friendships (year: 2022) Unit of analysis: Directed dyad Type: cross-sectional $N_{nodes} = 70.3m$ Method: Dyadic model	Not studied within high schools	*Homophily*: SES *Popularity*: none *Sociality*: SES Not included: race	Sociality Not included: Reciprocity, Triadic closure, Preferential attachment	Facebook friendships of US population aged 25-44 in 2022 → quasi population data
McMillan 2022	Racial and gender homophily by friendship strength in middle and high schools. Data: PROSPER (year: 2002/3) Unit of analysis: Directed dyad Type: Cross-sectional (networks are observed across multiple time points but analyzed cross-sectionally) $N_{schools} = 51, N_{nodes} = 8{,}530$ Method: Valued ERGM meta-analysis	None	*Homophily:* SES, race, gender *Popularity*: same as for homophily *Sociality*: same as for homophily Not included: GPA	Reciprocity, Triadic closure Not included: Preferential attachment	Non-random sample of 28 school districts in rural PA and IA in 2002/3 in which a RCT was carried out at middle and high schools
Zhao 2022	Consolidation and ethnic segregation in European secondary school classrooms. Data: CILS4EU (year: 2011) Unit of analysis: Directed dyad Type: Cross-sectional $N_{classrooms} = 503, N_{nodes} = 10{,}904$ Method: ERGM meta-analysis	None	*Homophily:* SES, ethnicity *Sociality:* same as for homophily *Popularity*: same as for homophily Not included: GPA	Triadic closure, Reciprocity, Tension (dgwnsp) Not included: Preferential attachment	Sample of students in European secondary schools in 2011 → no survey weights (meta-analysis weighs by parameter precision, not by population weight)
Zhao 2023	Ethnic homophily and ethnic segregation in Swedish secondary school classrooms. Data: CILS4EU (year: 2011) Unit of analysis: Directed dyad Type: Cross-sectional $N_{classrooms} = 160, N_{nodes} = 3{,}255$ Method: ERGM meta-analysis	None	*Homophily:* ethnicity, gender *Sociality:* same as for homophily *Popularity*: same as for homophily Not included: SES, GPA	Triadic closure, Reciprocity, Preferential attachment	Sample of students in Swedish secondary schools in 2011 → no survey weights (meta-analysis weighs by parameter precision, not by population weight)
Hoffman & Chabot 2023	Determinants of socioeconomic segregation at a French summer camp. Data: Summer camp in France in 2019 Unit of analysis: Directed dyad Type: longitudinal $N_{nodes} = 60$ Method: Exp. random partition model	Shared rooms Joint activities	*Homophily:* SES, gender, age *Popularity*: same as for homophily *Sociality*: same as for homophily Not included: ethnicity, GPA	Reciprocity, Triadic closure Not included: Preferential attachment	Nonrandom sample of French teenagers aged 10-14
Zwier & Geven 2023	Study socioeconomic segregation in Dutch primary school classrooms. Data: PRIMS (year: 2020) Unit of analysis: Directed dyad Type: Cross-sectional (analysis) $N_{schools} = 55, N_{nodes} = 1{,}416$ Method: ERGM meta-analysis	Parental networks Not included: Courses, Extracurriculars, Neighborhoods	*Homophily:* SES, ethnicity, grade, gender *Sociality*: same as for homophily *Popularity*: same as for homophily Not included: GPA	Reciprocity, Triadic closure, Not included: Preferential attachment	Representative of Dutch primary school students aged 11-12 → no survey weights used but sample is self-weighting

Chabot 2024	Socioeconomic segregation in French middle schools. Data: 4 French middle schools (year: 2014) Unit of analysis: Directed dyad Type: Longitudinal $N_{schools} = 4, N_{nodes} = 820$ Method: SAOM simulations	Courses, Neighborhoods, Primary school network Not included: Extracurriculars	*Homophily:* SES, ethnicity, GPA, gender *Sociality*: same as for homophily *Popularity*: same as for homophily	Reciprocity, Triadic closure, Preferential attachment, Reciprocity*Triadic closure	Nonrandom sample of students in four middle schools in Savoie and Paris
Raabe et al 2024	Study of friendship nominations by parental income in Swedish secondary school classrooms Data: CILS4EU (Sweden, year: 2011) Unit of analysis: Directed dyad (?) Type: Longitudinal $N_{schools} = 129, N_{nodes} = 4787$ Method: Multilevel SAOM simulations	None	*Homophily:* SES, ethnicity, gender, family structure, region *Popularity*: same as for homophily *Sociality*: same as for homophily Not included: GPA	Reciprocity (?), Triadic closure, Preferential attachment	Sample of students in secondary schools Sweden in 2011 (no survey weights)

The review in Table C1 is confined to observational studies of determinants of racial and socioeconomic segregation using sociometric network data. I focus on racial and socioeconomic segregation because their determinants overlap to a great extent. Note that many (if not most) studies never intended to disentangle the main drivers of segregation. I examine prior studies through this lens to assess the extent to which observational studies have disentangled the relative contributions of determinants.

The review shows that much prior work omits important stratified settings (e.g., Goodreau et al. 2009; Mouw and Entwisle 2006), correlated homophilies (e.g., Chetty et al. 2022b; Goodreau et al. 2009), or relational mechanisms by employing dyadic network models (e.g., Chetty et al. 2022b; Malacarne 2017). Recent studies have made great progress towards parsing the main determinants but examine socioeconomic segregation in other countries (An 2022; Chabot 2024; Hoffman and Chabot 2023; Zwier and Geven 2023). Differences in degree and pattern of socioeconomic segregation across countries indicate that socioeconomic segregation in friendship networks is specific to its institutional and cultural context.

Abbreviations

- degree = sociality and popularity were not differentiated because the unit of analysis is the undirected dyad
- preferential attachment = shorthand to refer to the “gwidegree” and “gwodegree” statistics (propensity for in-/outgoing ties to form where there are already ties)
- Add Health = National Longitudinal Study of Adolescent Health
- PROSPER = Promoting School-Community Partnerships to Enhance Resilience study
- ANSR = Adolescent Smoking and Network Research study
- CILS4EU = Children of Immigrants Longitudinal Survey in Four European Countries Study
- PRIMS = Transition from Primary to Secondary education study
- SAOM = Stochastic actor-oriented modeling
- ERGM = Exponential random graph modeling

C2 SES disparities by third variables

This sections provides information on SES disparities by third variables to substantiate some of the interpretations in the results and conclusion section.

SES disparities by schools

Students differ in SES background more within schools than between schools. To show this, I decompose the variance in the continuous SES score into within- and between-school components across the 128 schools in the core sample. The variance within schools equals $\mathrm{V_W} = \frac{\sum_s w\sigma_s^2}{128}$ and the variance between schools equals $\mathrm{V_W} = \frac{\sum_s\left(w\bar{\mu}_s - \frac{\sum w\bar{\mu}_s}{128}\right)^2}{128}$, where $\bar{\mu}_s$ and σ_s^2 are the school-specific means and variances and w is Add Health's school weight. The school-specific means (μ_s) and variances (σ_s^2) are calculated using Add Health's inschool weights. The variance in the SES score equals $\mathrm{V_W} = 1.35$ on average within schools. In contrast, the variance in the average SES score between schools equals $\mathrm{V_B} = 0.25$.

SES disparities by race

In the Add Health schools, the race-SES correlation between schools is $\rho_B = 0.24$. Within schools, in contrast, it is only $\rho_W = 0.11$. To calculate the race-SES correlation within and between schools, I fit a linear model using Add Health's grand sample weight that regresses students' SES background on schools' racial distributions (fully interacted proportions). The adjusted $\mathrm{R}^2 = 0.059$. Subsequently, I add students' race to the model. The adjusted R^2 increases by 0.012. The race-SES correlations reported on p.8 are the square root of these figures. Note that these correlations provide an intuition for the race-SES correlation within and between schools. Technically, they are conditional correlations between fitted and observed SES scores.

In the Add Health schools, SES disparities are greater within racial groups than between them. To show this, I decompose the variance in the continuous SES score by racial group into within- and between-group components. In a given school, the variance within racial groups equals $\mathrm{V_W}(s) = \frac{1}{5}\sum_i \sigma_{is}^2$ and the variance between racial groups equals $\mathrm{V_B}(s) = \frac{1}{5}\sum_i(\mu_{is} - \bar{\mu}_s)^2$, where μ_{js} and σ_{is}^2 are the race-specific means and variances in school s (calculated using Add Health's inschool weight). I then average the within- and between-group components across schools using Add Health's school weights. On average across schools, the variance in the SES score within racial groups equals $\mathrm{V_W} = 0.97$. In contrast, the variance in the average SES score between racial groups equals $\mathrm{V_B} = 0.27$.

SES disparities by GPA

Using the same calculations as above, the GPA-SES correlation between schools is $\rho_B = 0.19$. Within schools, it is $\rho_W = 0.18$. Therefore, between schools, the GPA-SES correlation is lower than the race-SES correlation, but within schools, the GPA-SES correlation is greater than the race-SES correlation.

SES disparities by neighborhoods

Using the same variance decomposition as for race, I find that, in the Add Health schools, SES disparities are greater between neighborhoods than within them. On average across schools, the variance in the SES score within neighborhoods equals $\mathrm{V_W} = 0.56$. In contrast, the variance in the average SES score between neighborhoods equals $\mathrm{V_B} = 0.84$. Neighborhoods may thus contribute to socioeconomic segregation. However, their specific impact on socioeconomic segregation cannot be determined from these variances alone as it depends, among other things, on the distribution of students across neighborhoods and the effect of residential proximity on friendship formation.

SES disparities by course selections

Using the same variance decomposition as for race, I find that, in the Add Health schools, SES disparities are greater between courses than within them. On average across schools, the variance in the SES score within courses equals $\mathrm{V_W} = 0.64$. In contrast, the variance in the average SES score between courses equals $\mathrm{V_B} = 0.72$. Courses may thus contribute to socioeconomic segregation. However, their specific impact on socioeconomic segregation cannot be determined from these variances alone as it depends, among other things, on the distribution of students across courses and the course effect on friendship formation.

SES disparities by extracurricular participation

Using the same variance decomposition as for race, I find that, in the Add Health schools, SES disparities are greater within extracurricular activities than between them. On average across schools, the variance in the SES score within extracurricular activities equals $\mathrm{V_W} = 0.93$. In contrast, the variance in the average SES score between extracurricular activities equals $\mathrm{V_B} = 0.38$. Extracurricular activities may thus attenuate socioeconomic segregation. However, their specific impact on socioeconomic segregation cannot be determined from these variances alone as it depends, among other things, on the distribution of students across extracurricular activities and the effect of extracurricular activities on friendship formation.

C3 ERGM specifications (M3)

Here I provide more details on the specification of models 1-3. Table C2 details the specific ERGM terms I used to build the model. As can be seen from the table, M3 is relatively complex. Substantial performance improvements since version 4 of the R package "ergm" make it possible to fit such models on larger networks in reasonable time (Krivitsky et al. 2022). The joint estimation also improves model convergence and predictive performance (Stewart et al. 2019; Tolochko and Boomgaarden 2024). As a robustness check, I estimated the same models using the multilevel model proposed by Krivitsky et al. (2023), which produces very similar results. I decided against this approach because the R package "ergm.multi" is slower during simulation and does not allow me to specify the desired constraint combination.

To parse the relative impact of the modeled mechanisms, I turn on the estimated conditional effects one by one (in silico experiments). I start by turning on the effects of stratified settings to examine how neighborhoods, courses, and extracurricular activities contribute to socioeconomic segregation in the absence of homophilous tendencies, popularity differences, and relational mechanisms. Subsequently, I switch on homophilous tendencies to analyze how homophily on the basis of GPA, race, and SES background within these settings contribute to socioeconomic segregation. Then, I turn on popularity differences by GPA, race, and SES background to examine the extent to which segregation is the result of an unequal rejection of outgroup members rather than an equal one (i.e., homophily). Finally, I switch on relational mechanisms to observe how they amplify the existing socioeconomic segregation. While I chose this order carefully, I show in Appendix C5 that the impact of each mechanism on socioeconomic segregation does not depend much on the ordering.

To further clarify the approach, I discuss in the following the impact of race homophily, SES homophily, and popularity differences by SES background. I model the differential effect of race homophily on friendship formation using the main diagonal of the racial mixing matrix (i.e., $\hat{\theta}_{\text{White}\rightarrow\text{White}}$, $\hat{\theta}_{\text{Black}\rightarrow\text{Black}}$, $\hat{\theta}_{\text{Asian}\rightarrow\text{Asian}}$, $\hat{\theta}_{\text{Hispanic}\rightarrow\text{Hispanic}}$)[26]. Like in any other multi-variable regression model, these are conditional effects that represent the effect of racial homophily on friendship formation net of other modeled mechanisms. This effects on friendship formation then turns into an impact on socioeconomic segregation, $\hat{\tau}_{\text{race}}$, through the (unconditional) correlation between students' race and SES background. In other words,

[26] For categorical statistics, I chose the modal category as baseline. Note that this choice does not affect the results very much. To see this, consider the impact of racial homophily on socioeconomic segregation. The impact is calculated by comparing the segregation in synthetic networks in which the effect of racial homophily is turned on to synthetic networks in which it is turned off. By choosing ties among white students as baseline, in networks in which racial homophily is turned off, *all* ties have the same friendship propensity as ties among whites. The choice of baseline does not affect $\hat{\tau}_{\theta_k}$ because the impact of racial homophily on segregation arises from differences in tie propensity across racial combinations and these differences do not change depending on which baseline category is chosen.

the impact of racial homophily on socioeconomic segregation arises from the structural intersectionality between these attributes. Through interaction effects, I also consider behavioral intersectionalities between race and SES background.

I model the differential effect of SES homophily on friendship formation using the main diagonal of the socioeconomic mixing matrix (i.e., $\hat{\theta}_{\text{low}\to\text{low}}$, $\hat{\theta}_{\text{med}\to\text{med}}$, $\hat{\theta}_{\text{high}\to\text{high}}$). The impact of SES homophily on socioeconomic segregation, $\hat{\tau}_{\text{SES}}$, should be understood as a residual effect that captures the remaining unobserved drivers of socioeconomic segregation net of other modeled mechanisms[27].

Finally, I model the differential effect of popularity differences by SES background on friendship formation using the off-diagonal elements of the socioeconomic mixing matrix (i.e., $\hat{\theta}_{\text{low}\to\text{med}}$, $\hat{\theta}_{\text{low}\to\text{high}}$, $\hat{\theta}_{\text{med}\to\text{low}}$, et cetera). This operationalization builds and improves upon prior work that separates homophily and popularity mechanisms using mixing matrices (An 2022; An and McConnell 2015). The separation of homophily and popularity in this prior work is incomplete because outdegree differences are not taken out. In addition to group-specific homophily, the incidences in mixing matrices also reflect group differences in- and outdegree. Homophily and popularity effects will thus be confounded by differences in sociality without removing outdegree differences. I take them out by holding constant the outdegree distribution during estimation and simulation, which allows for a more precise separation of these two mechanisms.

[27] Including the socioeconomic mixing matrix in the model does not affect the results or pose estimation issues because M3 is a model of friendship formation, not of socioeconomic segregation. The model is thus not saturated. I simply include socioeconomic mixing matrix to be able to turn on and off its effects when simulating networks.

Table C2: *ERGM statistics of models 1-3.*

ERGM statistics	Explanation
Random selection between and within schools (M1)	
formula = g ~ edges constraints = ~ odegrees	Random selection between and within schools. Students in simulated networks have the same outdegree as in the observed networks.
Random selection within schools (M2)	
formula = g ~ edges + offset(F(~edges, ~!nodematch(“schoolid”))) constraints = ~ odegrees	Random selection within schools. Students in simulated networks have the same outdegree as in the observed networks.
Full model (M3)	
formula = g ~ edges + offset(F(~edges, ~!nodematch(“schoolid”))) + constraints = ~ odegrees	Students in simulated networks have the same outdegree as in the observed networks.
+ Stratified settings	
edgecov(x) + edgecov(x^2) + edgecov(x^3) + x = course overlap, extracurricular overlap, residential proximity	Course overlap, extracurricular activity overlap, and residential proximity
+ Homophily and Popularity differences	
nodemix(SES, levels2=mm_SES) + nodemix(GPA, levels2=mm_GPA) + nodemix(race, levels2=mm_race) + nodematch(“race_sub”) + mm_SES = matrix(c(“11”, “12”, “13”, “21”, NA, “23”, “31”, “32”, “33”), 3, 3) mm_GPA and mm_race are constructed analogously also included (using F() operator): – Intersectional homophily: same_SES * same_race, same_SES * race, same_race * SES – Intersectional popularity: race * SES	Homophily and popularity on the basis of SES, GPA, as well as race and subracial categories. SES: Homophilous and aspirational tendencies are separated using the diagonal statistics as homophily (11, NA, 33) and the off-diagonal terms as aspiration (12, 13, 21, 23, 31, 32). The baseline is homophily among medium-SES students (22). Analogous separations are performed for GPA and race. I include several interactions to measure potential behavioral peculiarities at the intersection of race and SES background.
+ Relational mechanisms	
mutual F(~mutual, ~nodemix(SES, mm_xses) + F(~mutual, ~nodemix(SES, mm_gpa) + F(~mutual, ~nodemix(SES, mm_xrace) +	Reciprocity I also include differential reciprocity to estimate whether the tendency to reciprocate ties depends on whether they cross group boundaries.
gwesp(0, fixed=T) + gwdsp(0, fixed=T), F(~gwesp(0, fixed=T), ~nodemix(SES, mm_xses) + F(~gwesp(0, fixed=T), ~nodemix(SES, mm_gpa) + F(~gwesp(0, fixed=T), ~nodemix(SES, mm_xrace) +	Triadic closure – I include “gwdsp” to estimate “gwesp” more accurately as some of the transitive triangles may in fact be created by open two paths. – I also include differential triadic closure to estimate whether the tendency to close open triads depends on whether the open edge crosses group boundaries.

ERGM statistics	Explanation
gwidegree(0, fixed=T)	Preferential attachment I include preferential attachment to estimate the effect of the other modeled mechanisms more accurately. I leave it out of the paper because its impact on segregation is negligible and readers find the difference between preferential attachment and popularity differences by student attributes hard to comprehend. A negative coefficient indicates preferential attachment (i.e., ties are more likely to be directed toward a few subjects) (Hunter 2007).
absdiff(grade) + absdiff(age) + nodematch(gender)	I include these statistics to improve model-fit. They do not affect the results as they are uncorrelated with SES background.

C4 ERGM regression results (M3)

This section provides the regression output for M3. Table C3 presents the same results as in Figure 5: the average marginal effect of each variable (Duxbury and Wertsching 2023). Table C4 presents the normal ERGM regression output: the effect of each mechanism on the conditional log-odds of a tie.

Table C3: **Average marginal effects (AME)**. *The table provides the AME of each variable, measuring the average change in tie probability as a variable increases by one unit. AMEs are on a probability scale, which allows for effect size comparisons. Effects should be interpreted relative to the baseline tie probability, which is 0.01% on average across networks. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, + p<0.01.*
Interpretation example: The popularity of high-SES students increases the baseline tie probability of ties to them by 0.1 percentage points (from 0.01% to 0.02%).

Variable	Beta	SE	P
	Stratified settings		
(Overlap courses)^1	1.3857	0.01509	***
(Overlap courses)^2	-0.75137	0.014005	***
(Overlap courses)^3	0.12304	0.003593	***
(Overlap extras)^1	0.16764	0.028651	***
(Overlap extras)^2	0.071723	0.055016	
(Overlap extras)^3	-0.05136	0.025036	*
(Residential proximity)^1	-0.21934	0.015068	***
(Residential proximity)^2	0.12401	0.017061	***
(Residential proximity)^3	-0.021468	0.004617	***
	Homophily and popularity		
Popularity: low-GPA	-0.01288	0.008048	
Popularity: med-GPA	0.023038	0.009489	*
Popularity: high-GPA	-0.01519	0.007588	*
Homophily: low-GPA	0.055254	0.010134	***
Homophily: med-GPA		reference	
Homophily: high-GPA	0.052023	0.009897	***
Popularity: low-SES	0.018778	0.036628	
Popularity: med-SES	0.04936	0.036275	

Popularity: high-SES	0.10029	0.036909	**
Homophily: low-SES	0.006558	0.010736	
Homophily: med-SES		reference	
Homophily: high-SES	0.022362	0.010123	*
Popularity: White	0.013734	0.022035	
Popularity: Black	-0.06361	0.022489	**
Popularity: Asian	-0.05906	0.023969	*
Popularity: Hispanic	0.013015	0.020746	
Homophily: White		reference	
Homophily: Black	0.22148	0.016283	***
Homophily: Asian	0.32663	0.016254	***
Homophily: Hispanic	0.22304	0.016648	***
Homophily: subracial match	0.15067	0.008417	***
Homophily: grade level	-0.02912	0.003152	***
Homophily: age	-0.03465	0.002836	***
Homophily: gender	0.067425	0.003268	***
Popularity: White x low-SES	-0.00553	0.028734	
Popularity: Black x low-SES	-0.02511	0.033872	
Popularity: Asian x low-SES	0.011678	0.041551	
Popularity: Hispanic x low-SES	-0.00297	0.02259	
Popularity: White x high-SES	-0.01875	0.028306	
Popularity: Black x high-SES	-0.07022	0.029448	*
Popularity: Asian x high-SES	-0.04548	0.027831	
Popularity: Hispanic x high-SES	-0.03454	0.028777	
Homophily: White x same-SES	0.008101	0.019942	
Homophily: Black x same-SES	0.002659	0.02082	
Homophily: Asian x same-SES	0.008189	0.021591	
Homophily: Hispanic x same-SES	0.058642	0.02064	**
Homophily: low-SES x same-race	-0.00333	0.013234	
Homophily: med-SES x same-race	-0.0205	0.012498	
Homophily: high-SES x same-race	-0.06265	0.013353	***
Homophily: same-SES x same-race	0.012609	0.018276	
Relational mechanisms			
Reciprocity (mutual)	0.5294	0.013975	***
Reciprocity x cross-SES	0.002851	0.014419	***
Reciprocity x cross-GPA	0.029358	0.013753	
Reciprocity x cross-race	0.15214	0.018799	*
Triadic closure (gwesp)	0.29171	0.004279	***
Triadic closure x cross-SES	0.009359	0.004572	*
Triadic closure x cross-GPA	0.016467	0.004549	***
Triadic closure x cross-race	0.008171	0.011608	
Open two paths (gwdsp)	-0.02694	0.000903	***
Preferential attachment (gwideg)	-0.19168	0.010006	***

Table C4: **ERGM results.** *The table provides the regular ERGM output: the effect of each variable on the conditional log-odds of a tie. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, + p<0.01.*

Variable	Beta	SE	P
Stratified settings			
(Overlap courses)^1	6.230888	0.067856	***
(Overlap courses)^2	-3.37867	0.062978	***
(Overlap courses)^3	0.553254	0.016156	***
(Overlap extras)^1	0.753829	0.128835	***
(Overlap extras)^2	0.322515	0.24739	
(Overlap extras)^3	-0.23094	0.112578	*
(Residential proximity)^1	0.986308	0.067757	***
(Residential proximity)^2	-0.557618	0.076716	***
(Residential proximity)^3	0.096534	0.020762	***
Homophily and popularity			
Popularity: low-GPA	-0.057925	0.036189	
Popularity: med-GPA	0.103593	0.042669	*
Popularity: high-GPA	-0.068299	0.034119	*
Homophily: low-GPA	0.248461	0.04557	***
Homophily: med-GPA	reference		
Homophily: high-GPA	0.233932	0.044503	***
Popularity: low-SES	0.08444	0.164704	
Popularity: med-SES	0.221958	0.163116	
Popularity: high-SES	0.450987	0.16597	**
Homophily: low-SES	0.029487	0.048277	
Homophily: med-SES	reference		
Homophily: high-SES	0.100555	0.04552	*
Popularity: White	0.061757	0.099083	
Popularity: Black	-0.265553	0.107783	**
Popularity: Asian	-0.286051	0.101125	*
Popularity: Hispanic	0.058525	0.093287	
Homophily: White	reference		
Homophily: Black	1.468739	0.07309	***
Homophily: Asian	0.995916	0.073218	***
Homophily: Hispanic	1.002955	0.074862	***
Homophily: subracial match	0.677521	0.037848	***
Homophily: grade level	-0.130959	0.014173	***
Homophily: age	-0.155812	0.012754	***
Homophily: gender	0.303191	0.014697	***
Popularity: White x low-SES	-0.024851	0.129207	
Popularity: Black x low-SES	-0.112927	0.152313	
Popularity: Asian x low-SES	0.052512	0.186842	
Popularity: Hispanic x low-SES	-0.013361	0.101582	
Popularity: White x high-SES	-0.084301	0.127282	
Popularity: Black x high-SES	-0.315761	0.132419	*
Popularity: Asian x high-SES	-0.204514	0.125147	
Popularity: Hispanic x high-SES	-0.155301	0.129401	
Homophily: White x same-SES	0.036429	0.089672	
Homophily: Black x same-SES	0.011954	0.093622	
Homophily: Asian x same-SES	0.036825	0.097089	
Homophily: Hispanic x same-SES	0.263695	0.092812	**
Homophily: low-SES x same-race	-0.014967	0.059511	
Homophily: med-SES x same-race	-0.092179	0.056198	
Homophily: high-SES x same-race	-0.281701	0.060045	***
Homophily: same-SES x same-race	0.0567	0.082181	

Relational mechanisms			
Reciprocity (mutual)	2.380564	0.062841	***
Reciprocity x cross-SES	0.012821	0.06484	***
Reciprocity x cross-GPA	0.132014	0.061841	
Reciprocity x cross-race	0.684125	0.084531	*
Triadic closure (gwesp)	1.311728	0.019243	***
Triadic closure x cross-SES	0.042083	0.020557	*
Triadic closure x cross-GPA	0.074047	0.020455	***
Triadic closure x cross-race	0.036744	0.052197	
Open two paths (gwdsp)	-0.121156	0.004061	***
Preferential attachment (gwideg)	-0.861941	0.044993	***

C5 Robustness checks: sequence of effects (M3)

This sections demonstrates that the results presented in Figure 7 are not artifacts of the sequence in which mechanisms are added. In Figure 7, I first turn on the effects of stratified settings to examine how neighborhoods, courses, and extracurricular activities contribute to socioeconomic segregation absent friending biases. Subsequently, I switch on homophilous tendencies to analyze how homophily on the basis of GPA, race, and SES background within these settings contribute to socioeconomic segregation. Then, I turn on popularity differences by GPA, race, and SES background to examine the extent to which segregation is the result of an unequal rejection of outgroup members rather than an equal one (i.e., homophily). Finally, I switch on relational mechanisms to examine if they amplify the existing socioeconomic segregation.

Figure C1 shows that the impact of mechanisms does not depend much on the ordering. In the figure, I exemplarily switched the ordering of race and SES for homophilous tendencies and popularity differences. While the proportions are not identical to Figure 7 (partly because I only simulate 100,000 networks), the overall picture is the same. Both racial and socioeconomic homophily contribute to socioeconomic segregation while popularity differences by race barely affect it and popularity differences by SES reduce socioeconomic segregation. The reason why the ordering does not matter much is that I switch on conditional effects estimated in a single model rather than unconditional effects estimated in a sequence of models. An exception is relational mechanisms; their impact on socioeconomic segregation depends on the existing segregation in the network, which is why I add them last.

An alternative approach to measuring relative contributions would be to switch off mechanisms while keeping all other mechanisms switched on. A disadvantage of this approach is that contributions do not sum up (i.e., $\sum \tau_{\theta_k} \neq \hat{\pi}_t^{\text{M3,full}} - \hat{\pi}_t^{\text{M3,empty}}$) and are much smaller because relational mechanisms amplify all but the absent mechanism and because mechanisms compensate for each other due to intercorrelations. This can be seen in Figure C2, where the impact of a mechanism was measured by switching it off while keeping all other mechanisms switched on. The overall picture is again very similar, but effect sizes are much lower.

Figure C1: *As compared to Figure 7, the sequence of race and SES for homophilous tendencies and popularity differences was switched during simulation.*

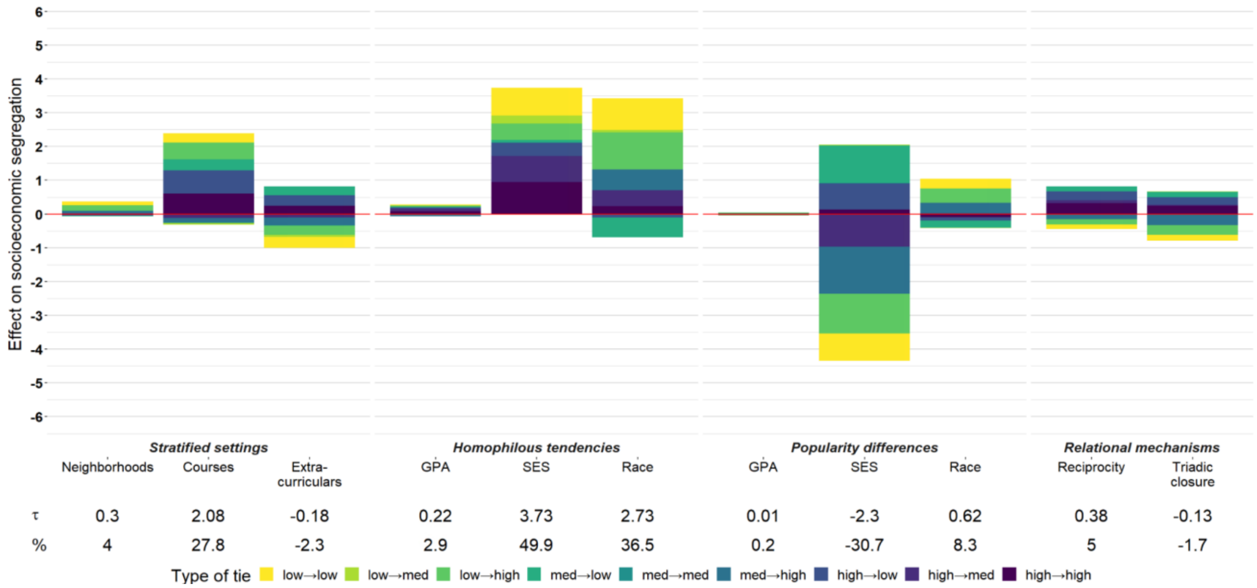
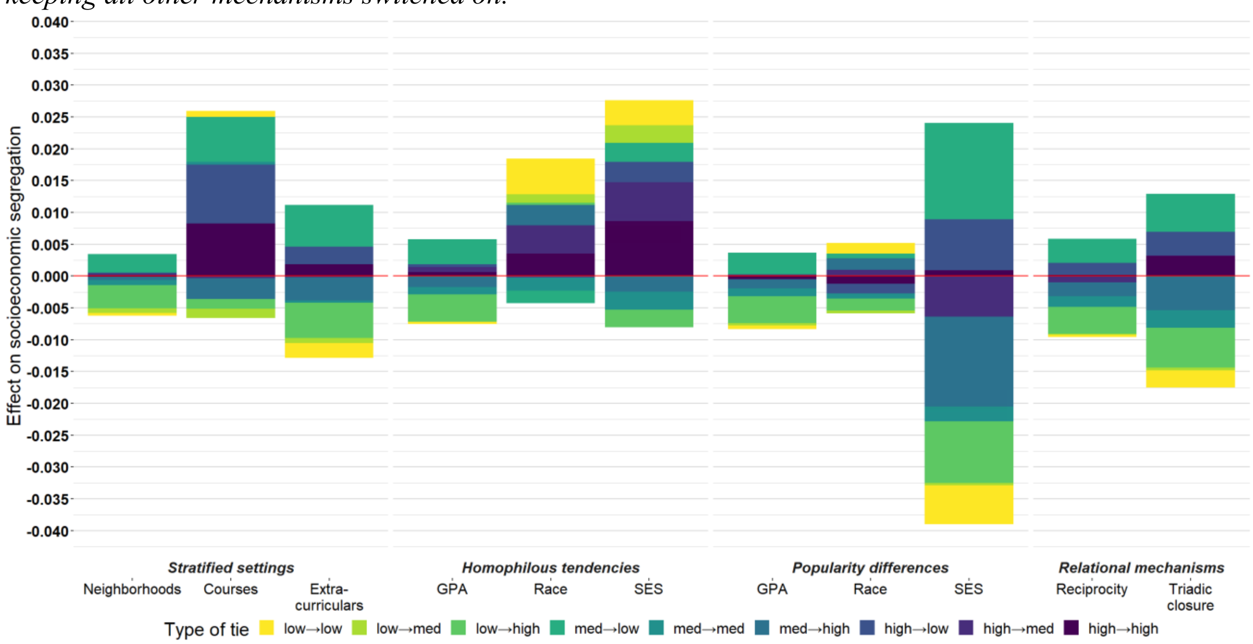


Figure C2: *As compared to Figure 7, the impact of each mechanism was measured by switching it off while keeping all other mechanisms switched on.*



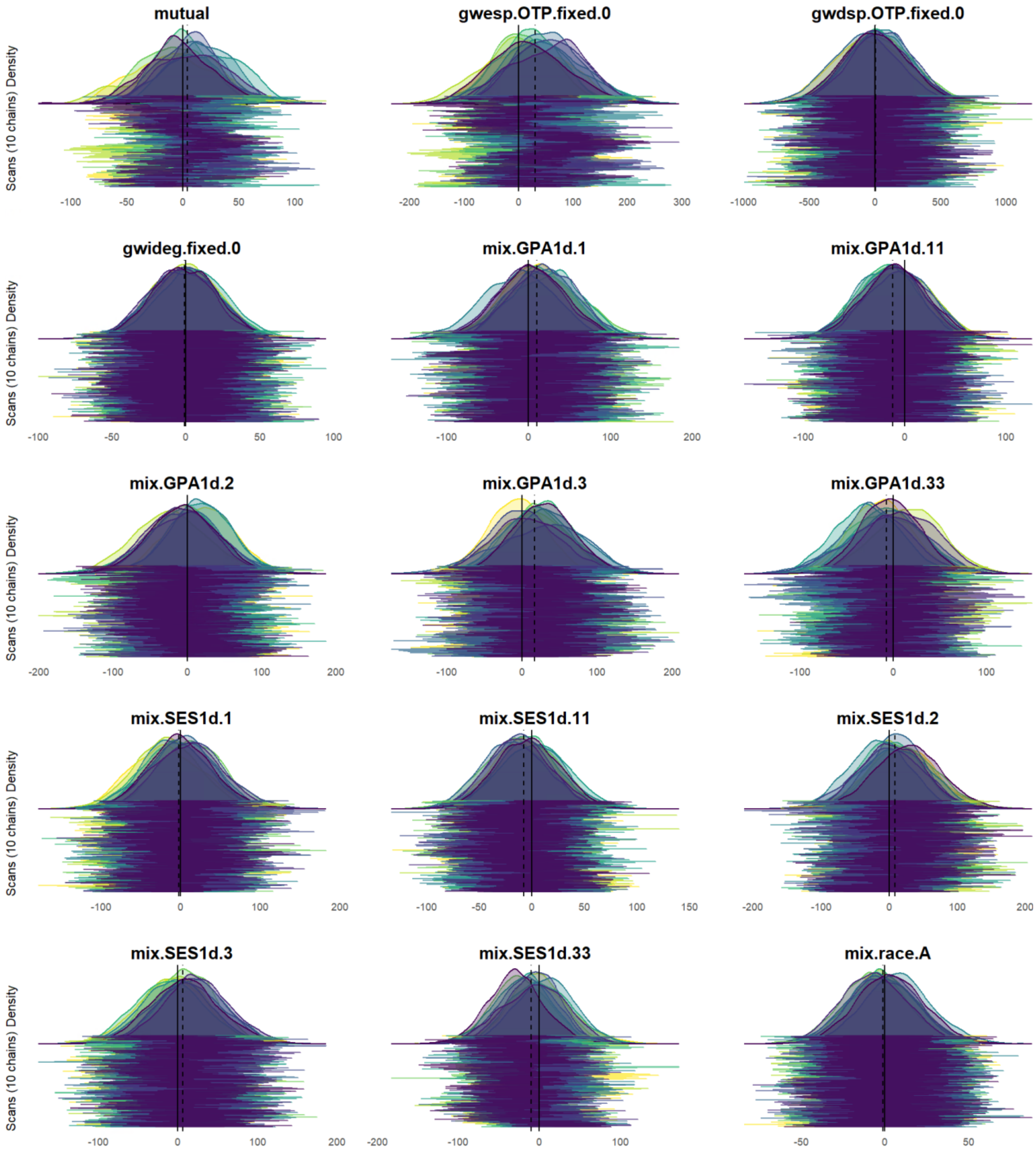
C6 Convergence statistics and goodness-of-fit statistics (M3)

The MCMC diagnostics suggest that the chains successfully converged to the posterior distributions. Figure C3 depicts what I call "monetplots", a combination of density and trace plot, depicting the density of the MCMC sample statistics and the trace that lead to it. For the most part, the monetplots indicate that the 10 specified chains are mixing well since densities are bell-shaped and centered around zero. The reciprocity and triadic closure parameters show trace plots with some patterns. To investigate, I fitted a simpler model without interactions between reciprocity and cross-group ties as well triadic closure and cross-group ties. Without these interactions the patterns disappear. Therefore, the interactions introduce some estimation uncertainty in the model.

Apart from reciprocity and triadic closure, MCMC chains show little serial correlation between sample statistics at different points in the chain, which suggests that they are independent draws. Similarly, the overall Geweke burn-in diagnostic is nonsignificant, indicating that the means at different locations of the chains are equal.

Figure C4 presents goodness-of-fit diagnostics. The plots show that the model recapitulates well the observed in- and outdegree distribution, edge-wise shared partners, and geodesic distances. The plot on the top left depicts the models' observed sufficient statistics as quantiles of the simulated sample. The observed statistics are near the sample median, indicating that the model recapitulates observed structures to a satisfactory degree.

Figure C3: *"Monetplots": a combination of density and trace plot, depicting the density of the MCMC sample statistics and the trace that lead to it.*



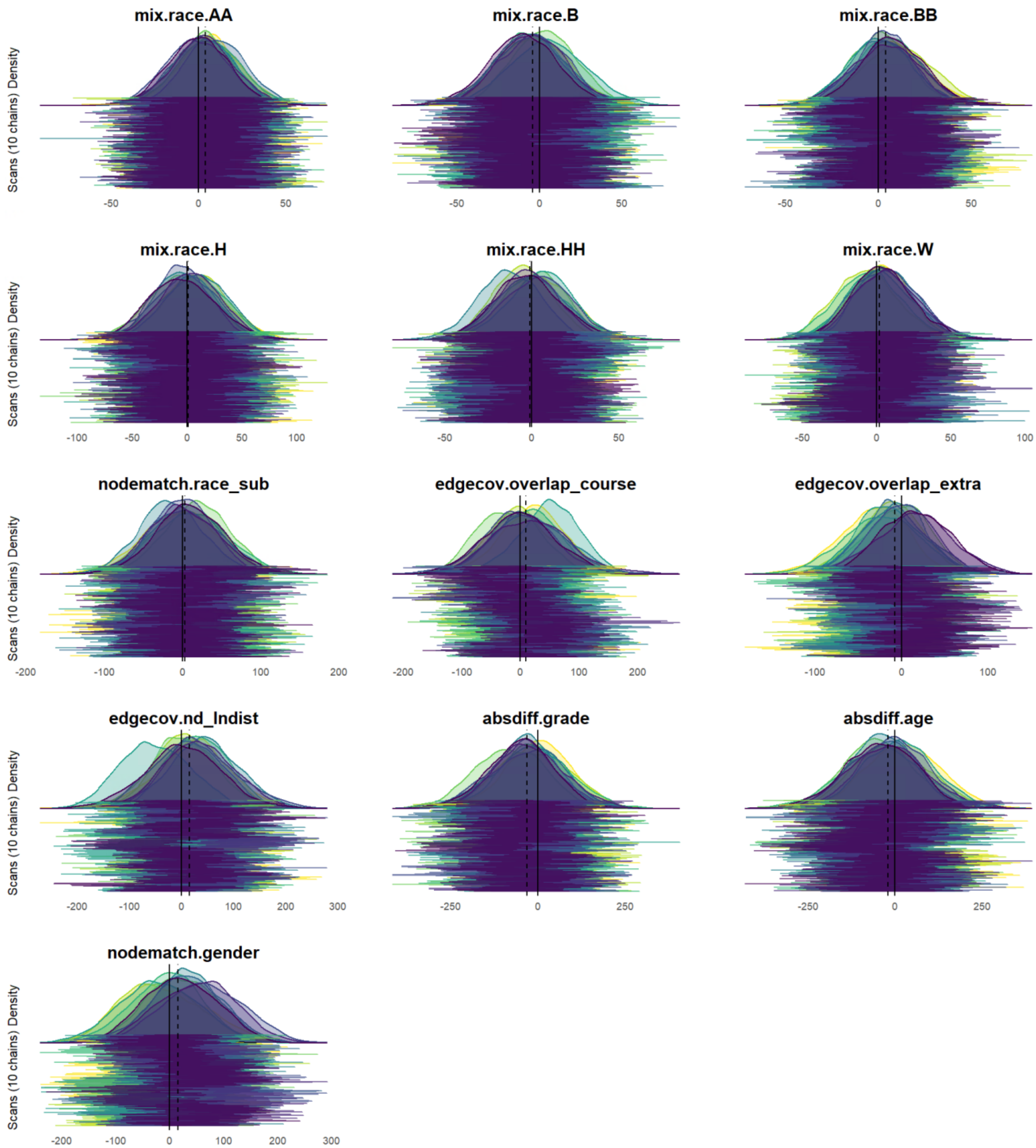
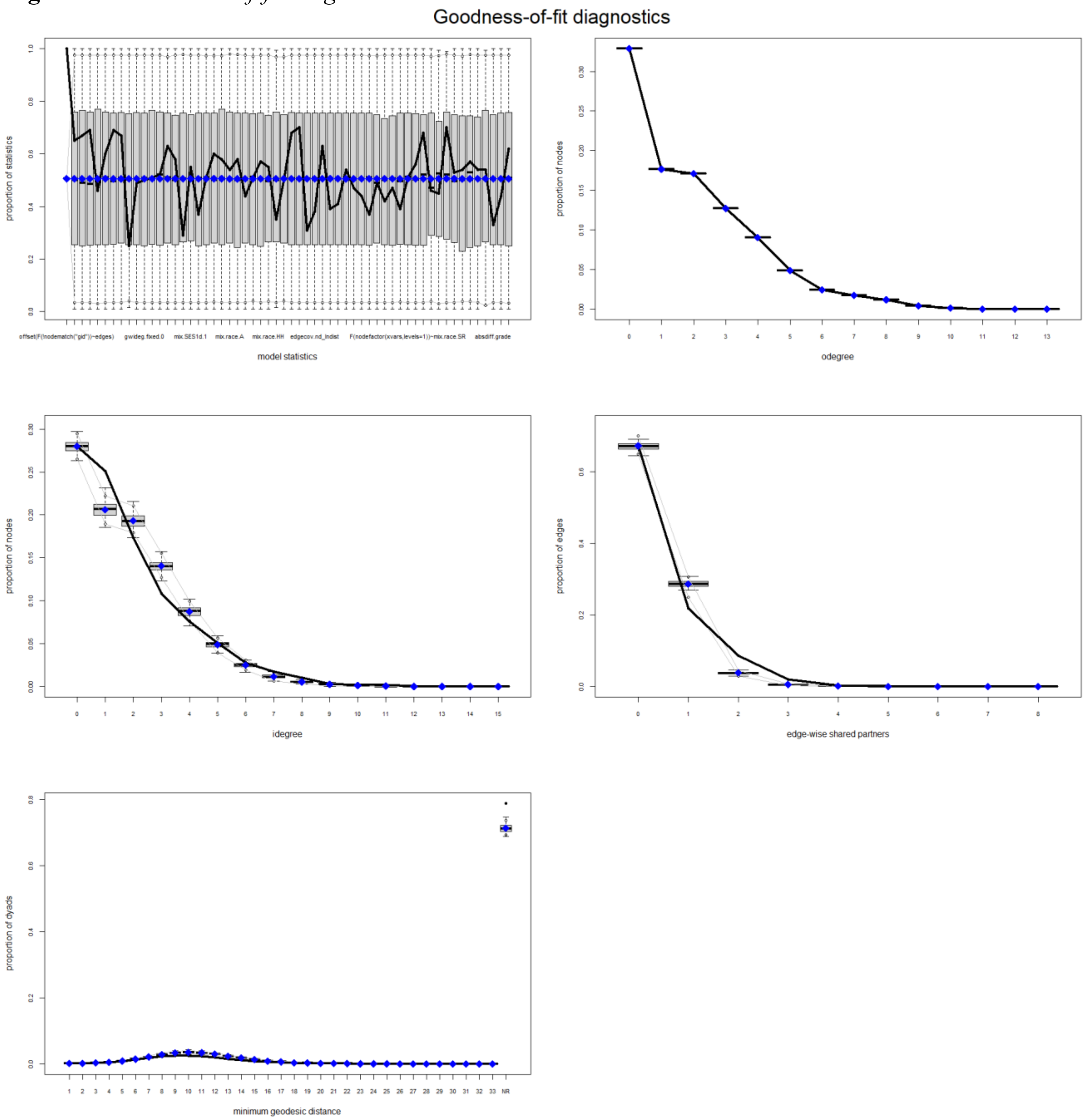

mix.race.AA
mix.race.B
mix.race.BB
mix.race.H
mix.race.HH
mix.race.W
nodematch.race_sub
edgecov.overlap_course
edgecov.overlap_extra
edgecov.nd_lndist
absdiff.grade
absdiff.age
nodematch.gender
Scans (10 chains) Density

Figure C4: *Goodness-of-fit diagnostics.*



REFERENCES

An, Weihua. 2022. 'Friendship Network Formation in Chinese Middle Schools: Patterns of Inequality and Homophily'. *Social Networks* 68:218–28. doi: 10.1016/J.SOCNET.2021.07.003.

An, Weihua, and Will R. McConnell. 2015. 'The Origins of Asymmetric Ties in Friendship Networks: From Status Differential to Self-Perceived Centrality'. *Network Science* 3(2):269–92. doi: 10.1017/NWS.2015.12.

Barber, Carolyn, and Jillian Woodford Wasson. 2014. 'A Comparison of Adolescents' Friendship Networks by Advanced Coursework Participation Status'. *Gifted Child Quarterly* 59(1):23–37. doi: 10.1177/0016986214559639.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. 'Mice: Multivariate Imputation by Chained Equations in R'. *Journal of Statistical Software* 45(3):1–67. doi: 10.18637/JSS.V045.I03.

Chabot, Timothée. 2024. 'How Does Socioeconomic Homophily Emerge? Testing for the Contribution of Different Processes to Socioeconomic Segregation in Adolescent Friendships'. *Social Networks* 76:160–73. doi: 10.1016/J.SOCNET.2023.09.002.

Chantala, K. 2001. *National Longitudinal Study of Adolescent Health: Constructing Weights to Use in Analyzing Pairs of Individuals from Add Health Data*.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022. 'Social Capital II: Determinants of Economic Connectedness'. *Nature 2022 608:7921* 608(7921):122–34. doi: 10.1038/s41586-022-04997-3.

Duxbury, Scott W., and Jenna Wertsching. 2023. 'Scaling Bias in Pooled Exponential Random Graph Models'. *Social Networks* 74:19–30. doi: 10.1016/J.SOCNET.2023.02.003.

Frank, Kenneth A., Chandra Muller, Kathryn S. Schiller, Catherine Riegle-Crumb, Anna Strassmann Mueller, Robert Crosnoe, and Jennifer Pearson. 2008. 'The Social Dynamics of Mathematics Coursetaking in High School'. *American Journal of Sociology* 113(6):1645–96. doi: 10.1086/587153.

Goldthorpe, John H., Catriona. Llewellyn, and Clive. Payne. 1980. *Social Mobility and Class Structure in Modern Britain*. Clarendon Press.

Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. 'Birds of a Feather, Or Friend of a Friend?: Using Exponential Random Graph Models to Investigate Adolescent Social Networks'. *Demography* 46(1):103–25. doi: 10.1353/dem.0.0045.

Hoffman, Marion, and Timothée Chabot. 2023. 'The Role of Selection in Socioeconomic Homophily: Evidence from an Adolescent Summer Camp'. *Social Networks* 74:259–74. doi: 10.1016/J.SOCNET.2023.04.002.

Hunter, David R. 2007. 'Curved Exponential Family Models for Social Networks'. *Social Networks* 29(2):216–30. doi: 10.1016/J.SOCNET.2006.08.005.

Krause, Robert W., Mark Huisman, Christian Steglich, and Tom Snijders. 2020. 'Missing Data in Cross-Sectional Networks – An Extensive Comparison of Missing Data Treatment Methods'. *Social Networks* 62:99–112. doi: 10.1016/J.SOCNET.2020.02.004.

Krivitsky, Pavel N., Pietro Coletti, Niel Hens, and South Wales. 2023. 'A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks'. *Journal of the American Statistical Association* 1–22. doi: 10.1080/01621459.2023.2242627.

Krivitsky, Pavel N., David R. Hunter, Martina Morris, and Chad Klumb. 2022. 'Ergm 4: Computational Improvements'. *ArXiv 2203.08198, ArXiv.Org E-Print Archive.*

Madley-Dowd, Paul, Rachael Hughes, Kate Tilling, and Jon Heron. 2019. ‘The Proportion of Missing Data Should Not Be Used to Guide Decisions on Multiple Imputation’. *Journal of Clinical Epidemiology* 110:63–73. doi: 10.1016/J.JCLINEPI.2019.02.016.

Malacarne, Timothy. 2017. ‘Rich Friends, Poor Friends: Inter–Socioeconomic Status Friendships in Secondary School’. *Socius: Sociological Research for a Dynamic World* 3:237802311773699. doi: 10.1177/2378023117736994.

Mouw, Ted, and Barbara Entwisle. 2006. ‘Residential Segregation and Interracial Friendship in Schools’. *American Journal of Sociology* 112(2):394–441. doi: 10.1086/506415.

NCES. 1998a. *Overview of Public Elementary and Secondary Schools and Districts: School Year 1995-96, Table 6*.

NCES. 1998b. *Private School Universe Survey, 1995-96, Table 15*.

Smith, Jeffrey A., Jonathan H. Morgan, and James Moody. 2022. ‘Network Sampling Coverage III: Imputation of Missing Network Data under Different Network and Missing Data Conditions’. *Social Networks* 68:148–78. doi: 10.1016/J.SOCNET.2021.05.002.

Stewart, Jonathan, Michael Schweinberger, Michal Bojanowski, and Martina Morris. 2019. ‘Multilevel Network Data Facilitate Statistical Inference for Curved ERGMs with Geometrically Weighted Terms’. *Social Networks* 59:98–119. doi: 10.1016/J.SOCNET.2018.11.003.

Tolochko, Petro, and Hajo G. Boomgaarden. 2024. ‘Same but Different: A Comparison of Estimation Approaches for Exponential Random Graph Models for Multiple Networks’. *Social Networks* 76:1–11. doi: 10.1016/J.SOCNET.2023.05.003.

Zwier, Dieuwke, and Sara Geven. 2023. ‘Knowing Me, Knowing You: Socio-Economic Status and (Segregation in) Peer and Parental Networks in Primary School’. *Social Networks* 74:127–38. doi: 10.1016/J.SOCNET.2023.03.003.