

---

# ESSAYS IN EMPIRICAL PUBLIC FINANCE

Inauguraldissertation  
zur Erlangung des akademischen Grades  
eines Doktors der Wirtschaftswissenschaften  
der Universität Mannheim

vorgelegt von

Majed DODIN

im Herbst-/Wintersemester 2021

Abteilungssprecher	Prof. VOLKER NOCKE, Ph.D.
Referent	Prof. Dr. CHRISTOPH ROTHE
Koreferent	Prof. Dr. SEBASTIAN FINDEISEN
Vorsitzender	Prof. Dr. CARSTEN TRENKLER
Datum der Verteidigung	22.12.2021, Mannheim

---

# Acknowledgements

I owe a lot to a lot of people who helped me get through graduate school. First, I would like to express my gratitude to my supervisor Christoph Rothe for his patient support and guidance. I have greatly benefitted from the depth of his knowledge and his ability to convey complex concepts in an intuitive manner. He showed great kindness throughout the last four years and helped me become a lot less confused about some things. His feedback was instrumental in developing and improving the first chapter of this dissertation.

I am also indebted to my teachers, Carsten Trenkler and Markus Frölich, who got me interested in econometrics by demonstrating the scope of the field. Both have patiently answered my questions and provided me with invaluable opportunities to learn, which led me to pursue doctoral studies in the first place.

Further academic thanks go to my coauthors, Sebastian Findeisen, Lukas Henkel, Dominik Sachs and Paul Schüle, for many hours of constructive discussions on the conceptual framework and results presented in the second chapter of this dissertation.

I gratefully acknowledge all the administrative support and funding received during my graduate studies. I am particularly thankful to the Economics Department and the Center for Doctoral Studies in Economics at the University of Mannheim for facilitating my stays abroad. Moreover, this dissertation was supported by the University of Mannheim's Graduate School of Economic and Social Sciences, the Fontana Foundation and the state of Baden-Württemberg through the Landesgraduiertenförderung as well as bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG. I further thank the German Academic Exchange Service, for supporting my stay at Berkeley, and the Joachim Herz Foundation for facilitating access to the German census data.

I have had the great fortune of meeting many wonderful people in Mannheim, Berkeley and New Haven, and made friends there who got me through both good and bad times. I cannot list them all, but give my thanks to Celin, Christoph, Fabian, Julian, Justin, Robert, Sebastian, Tim, Tomasz, Vahe and Xin for being great companions during my years in Mannheim, to

Emily, Lauri, Daniel and Yussuf for making my year in Berkeley an unforgettable one, as well as to Ian, Ipsitaa, Lucas and Sakshi, who made me feel at home in New Haven, despite the weather.

Finally, I would like to thank my family, whom I have spent too much time away from, for their unconditional support and Michaela, who has always cheered me up, especially when I needed it most, for inspiring me with her resolutely positive outlook on life.





# Contents

Acknowledgements	i
List of Tables	vii
List of Figures	viii
Preface	xi
<b>1 Optimized Inference in Regression Kink Designs</b>	<b>1</b>
1.1 Introduction . . . . .	2
1.2 Framework and Notation . . . . .	6
1.2.1 Setup. . . . .	6
1.2.2 Parameter of Interest. . . . .	7
1.3 Optimized Honest Confidence Intervals . . . . .	8
1.3.1 Problem Statement. . . . .	8
1.3.2 General Approach. . . . .	9
1.3.3 Bias Characterization. . . . .	9
1.3.4 Estimation via Dual Optimization. . . . .	11
1.3.5 Interval Construction. . . . .	13
1.3.6 Theoretical Properties. . . . .	14
1.3.7 Practical Implementation. . . . .	18
1.4 Comparison with other Methods . . . . .	20
1.4.1 Local Linear Methods. . . . .	20
1.4.2 Monte Carlo Setup. . . . .	22
1.4.3 Monte Carlo Results. . . . .	23
1.4.4 Gains from Optimization. . . . .	26
1.5 Empirical Illustration . . . . .	28
1.6 Conclusion . . . . .	32

<b>Appendices to Chapter 1</b>	<b>33</b>
1.A Bias Derivation . . . . .	33
1.B Constraint Qualification . . . . .	35
1.C Dual Optimization . . . . .	38
1.D Implementation . . . . .	40
1.E Proof of Proposition 1 . . . . .	42
1.F Extensions . . . . .	43
1.F.1 Fuzzy Discontinuity Design. . . . .	43
1.F.2 Shape Constraints. . . . .	46
1.G Runtime Estimates . . . . .	47
1.H Replication Files and Programs . . . . .	48
1.I Additional Simulation Results . . . . .	49
<b>2 Social Mobility in Germany</b>	<b>53</b>
2.1 Introduction . . . . .	54
2.2 Related Literature and Institutional Background . . . . .	56
2.2.1 Related Literature . . . . .	56
2.2.2 Institutional Background . . . . .	58
2.3 Data and Measurement Strategy . . . . .	60
2.3.1 Variable Definitions . . . . .	61
2.3.2 Mobility Statistics . . . . .	64
2.3.3 Sample Definition and Limitations . . . . .	66
2.4 National Estimates . . . . .	69
2.4.1 Subgroup Estimates . . . . .	71
2.4.2 Time Trends . . . . .	74
2.5 Regional Estimates . . . . .	77
2.5.1 States . . . . .	78
2.5.2 Cities . . . . .	80
2.5.3 Local Labor Markets . . . . .	82
2.6 Conclusion . . . . .	89
<b>Appendices to Chapter 2</b>	<b>91</b>
2.A Additional Information on the Microcensus . . . . .	91
2.B Additional Material: National Estimates . . . . .	92

2.B.1	Sample Income Distribution and Ranks . . . . .	92
2.B.2	Alternative Equalization Schemes . . . . .	92
2.B.3	A-Level Wage Premium 1997-2016 . . . . .	94
2.B.4	Subgroup CEFs and Trends . . . . .	94
2.C	Additional Material: Regional Estimates . . . . .	98
2.C.1	State Level CEFs . . . . .	98
2.C.2	City Level CEFs . . . . .	99
2.C.3	Q5/Q1 Ratio by Local Labor Market . . . . .	100
2.C.4	Mobility Estimates by Planning Region . . . . .	101
2.C.5	Alternative Reference Distributions . . . . .	105
2.C.6	Correlation between Mobility Estimates . . . . .	106
2.C.7	Regional Indicators . . . . .	106
2.C.8	Split Sample Predictors . . . . .	109
	<b>References</b>	<b>111</b>
	<b>Curriculum Vitae</b>	<b>121</b>
	<b>Affidavit</b>	<b>123</b>

# List of Tables

1.1	Monte Carlo Results: Coverage and Relative Length I . . . . .	24
1.2	Monte Carlo Results: Coverage and Relative Length II . . . . .	25
1.3	RKD Estimates: Effect of the Benefit Level on Unemployment Duration . . .	30
1.4	Optimized Intervals under Shape Constraints: Concavity . . . . .	31
1.F.1	Shape of Anderson-Rubin Confidence Sets for $\tau_{RD}$ . . . . .	44
1.G.1	Runtime Estimates . . . . .	47
1.H.1	Dependencies . . . . .	48
1.H.2	Replication Files . . . . .	49
1.I.1	Monte Carlo Results: Coverage and Relative Length I . . . . .	50
1.I.2	Monte Carlo Results: Coverage and Relative Length II . . . . .	51
2.1	Co-Residence Rate by Child Age . . . . .	62
2.2	Monthly Child-related Expenditures of Single Child Households . . . . .	63
2.3	Average Characteristics of Children by Age at Observation . . . . .	67
2.4	National Mobility Estimates . . . . .	70
2.5	Mobility Statistics for Subgroups . . . . .	72
2.6	Social Mobility at the State Level . . . . .	79
2.7	Social Mobility in the 15 Largest Urban Labor Markets . . . . .	81
2.8	The 15 Most Informative Predictors of Relative Mobility . . . . .	88
2.C.1	Correlation between Mobility Statistics . . . . .	106
2.C.2	List of Regional Indicators . . . . .	107
2.C.3	Optimal Predictors of Relative Mobility by Size Category . . . . .	109

# List of Figures

1.1	Monte Carlo Setup: Shape of Conditional Expectation Functions . . . . .	22
1.2	Gains from Optimization: Fixed-Length Excess Length and UMSE . . . . .	27
1.3	Gains from Optimization: Shape of Optimized and Fixed-Length Weights . . .	27
1.4	Empirical Illustration: Weekly UI Benefit Schedule, Louisiana 1979-1984 . . .	29
1.5	Empirical Illustration: Average Duration and Density . . . . .	30
1.D.1	Implementation: Illustration of the Discretization Strategy . . . . .	40
2.1	Move-out Frequency by Parental Income Rank . . . . .	67
2.2	Social Mobility at the National Level . . . . .	69
2.3	Empirical Distribution by Parental Education . . . . .	73
2.4	A-Level Share by Cohort 1980-1996 . . . . .	75
2.5	Parental Income Gradient by Cohort 1980-1996 . . . . .	76
2.6	Quintile Measures by Cohort 1980-1996 . . . . .	76
2.7	A-Level Share by Cohort Quintile 1980-1996 . . . . .	77
2.8	Mobility in Hamburg and Leipzig . . . . .	82
2.9	Histogram of the Number of Observations by Local Labor Market . . . . .	83
2.10	A-Level Share and Q1 Measure by Local Labor Market . . . . .	84
2.11	Parental Income Gradient by Local Labor Market . . . . .	85
2.12	Sorting: Conditional and Unconditional Rank Gradients . . . . .	87
2.A.1	Illustration of the Microcensus Survey Design . . . . .	91
2.B.1	Equalized Household Income by Percentile Rank . . . . .	92
2.B.2	National Estimates under Different Equalization Schemes . . . . .	93
2.B.3	A-Level Wage Premium, Years 1997-2016 . . . . .	94
2.B.4	Social Mobility for Subgroups . . . . .	95
2.B.5	Trends in Parental Income Gradients by Subgroups . . . . .	96
2.B.6	Trends in A-Level Shares by Subgroups . . . . .	97
2.C.1	Social Mobility at the State Level . . . . .	98
2.C.2	Social Mobility in the 15 Largest Local Labor Markets . . . . .	99

2.C.3	Q5/Q1 Ratio by Local Labor Market . . . . .	100
2.C.4	Parental Income Gradient by Spatial Planning Region . . . . .	101
2.C.5	Q5/Q1 Ratio by Spatial Planning Region . . . . .	102
2.C.6	Q1 Measure by Spatial Planning Region . . . . .	103
2.C.7	A-Level Share by Spatial Planning Region . . . . .	104
2.C.8	State and Region Specific Parental Income Ranks . . . . .	105





# Preface

This dissertation consists of two self-contained chapters, which are broadly related to questions in empirical public finance. The first chapter contains a methodological contribution that is particularly suited for the study of tax systems and social insurance programs. The second chapter is empirical in nature and concerned with social mobility in Germany.

The question how governments should design tax systems and social insurance programs is a constant source of academic and public debate. Should the income tax schedule be more progressive? Should the government increase unemployment benefits? While different on the surface, both questions share the same structure, in the sense that the design of optimal policies requires knowledge of the behavioural responses of individuals to policy changes. One approach to inform policy is to specify a complete model of economic behaviour and to estimate the primitives of the model, allowing for the simulation of counterfactual policies. A second, less restrictive but empirically more demanding, approach is to derive optimal policy formulas as a function of estimable behavioural elasticities. The method proposed in Chapter 1 of this dissertation aims to provide a practical tool for researchers following the latter approach.

The second chapter is concerned with intergenerational social mobility in Germany. While the public political debate on transfer and education policies in Germany often centers around suspected consequences of proposed reforms for social mobility, and political parties across the spectrum refer to and justify their educational policy platforms by equality of opportunity principles, empirical evidence on the level and evolution of social mobility in Germany is scarce. How socially mobile is the German society? Has Germany become more or less mobile over the last few decades? Are the opportunities of disadvantaged children the same across regions in Germany? The second chapter of this dissertation proposes and implements a measurement framework for social mobility in Germany that allows for the estimation of robust and easy to interpret mobility statistics to document time trends and regional differences at a higher level of detail than previously possible.

In the following, I will briefly describe the content of each of the two chapters in more detail.

## **Chapter 1: Optimized Inference in Regression Kink Designs**

Tax and benefit schedules are often defined as a piecewise linear function of an assignment variable, resulting in kinks at the thresholds between adjacent parts of a schedule. Regression Kink Designs use such kinks to estimate the local average behavioral response of individuals to changes in the policy variable by comparing how the outcome of interest changes with the assignment variable in a small neighborhood around the threshold. In practise, this requires estimates of the derivative of the conditional expectation function of the outcome variable given the assignment variable to the left and the right of the threshold, which can be obtained non-parametrically, that is without imposing functional form assumptions. However, the statistical properties of nonparametric estimators depend on the smoothness of the unknown function, which complicates estimation and inference. This chapter proposes a method to remedy finite sample coverage problems and improve upon the efficiency of commonly employed procedures for the construction of nonparametric confidence intervals in regression kink designs. The proposed interval is centered at the linear minimax estimator over distributions with Lipschitz constrained conditional mean functions and obtained via numerical convex optimization. This approach offers several theoretical and practical advantages over traditional methods, leading to substantial improvements in the length and coverage properties of the resulting intervals.

## **Chapter 2: Social Mobility in Germany**

This chapter is based on joint work with Sebastian Findeisen, Lukas Henkel, Dominik Sachs and Paul Schüle. It proposes a measurement strategy for social mobility in Germany that is based on the association between parental income and the educational opportunities of children, which allows for the use of census data. Our measure of educational opportunities captures whether a child obtains an A-Level degree, the highest secondary schooling degree in Germany which grants direct access to the tuition-free national university system and marks an important sign of social distinction in the German society. We find that, at the national level, a 10 percentile increase in the parental income rank is associated with a 5.2 percentage point increase in the probability of obtaining an A-Level degree. This parental income gradient has not changed for the birth cohorts of 1980-1996, despite a large-scale policy of expanding upper secondary education. At the regional level, there exists substantial variation in mobility estimates, which cannot be explained by sorting of different households into different regions.

*To my family.*



# Chapter 1

## Optimized Inference in Regression Kink Designs

### Abstract

In this chapter, I propose a method to remedy finite sample coverage problems and improve upon the efficiency of commonly employed procedures for the construction of nonparametric confidence intervals in regression kink designs. The proposed interval is centered at the half-length optimal, numerically obtained linear minimax estimator over distributions with Lipschitz constrained conditional mean function. Its construction ensures excellent finite sample coverage and length properties which are demonstrated in a simulation study and an empirical illustration. Given the Lipschitz constant that governs how much curvature one plausibly allows for, the procedure is fully data driven, computationally inexpensive and valid irrespective of the distribution of the assignment variable.

## 1.1 Introduction

Research designs that warrant a causal interpretation of model parameter estimates based on observational data are becoming increasingly popular in the empirical social sciences. The quality of such observational studies is often assessed in terms of how well the assumptions that allow for the estimation of counterfactual quantities can be justified. In this context, methods that exploit discontinuities arising naturally from institutional rules are particularly attractive, as they allow the researcher to be precise about the source of variation utilized in the estimation of the parameter and discuss the assumptions that allow for a causal interpretation.

The Regression Kink Design (RKD) can be employed to estimate causal parameters when the variable of interest is a kinked function of an assignment variable. Analogous to the better-known Regression Discontinuity Design (RDD), which estimates the effect of a variable that changes its level discontinuously at a threshold, kink designs utilize discontinuous changes in the slope of the policy variable, effectively exposing units on each side of the threshold to different incentives. Such kinks arise naturally whenever marginal rates change discontinuously or benefit formulas involve maxima or minima and can be exploited to address important questions that are often difficult to be addressed experimentally, e.g. questions regarding the optimal design of unemployment insurance policies (see [Card et al., 2015a](#); [Landais, 2015](#); [Kolsrud et al., 2018](#)).

For example, a common feature of unemployment insurance systems is that unemployment benefits, typically a function of income in some base period, are capped at some maximum level. Consequently, the incentives for reemployment differ at each side of the threshold defined by the cap, and the (local) causal effect of the benefit level on unemployment duration can be estimated by comparing how the duration outcome changes with prior income at each side of the cutoff relative to the size of the kink. In practice, this amounts to estimating the jump in the first derivative of the conditional expectation function (CEF) of the duration outcome given the assignment variable at the kink point, and dividing the estimate by the kink size. If unobserved confounders vary smoothly at the discontinuity point, this corresponds up to scale to the (local) elasticity of unemployment duration with respect to the benefit level, a key parameter in dynamic labor supply models and corresponding welfare optimal benefit formulas.

Since the welfare effects of policy changes can often be expressed in terms of elasticities, kink designs can be used in a sufficient statistics approach to policy evaluation, avoiding the need for parametric assumptions and the estimation of structural primitives of the model, combining credible identification with the ability to make welfare predictions ([Chetty, 2009](#)).

A practical challenge for researchers utilizing discontinuity designs is to choose estimation and inference procedures that preserve the credibility of their findings stemming from the nonparametric identification of the causal parameter. The standard approach to inference in regression kink designs relies on local polynomial regression, that is fitting a polynomial model of order  $p \geq 1$  using only observations within a prespecified window of length  $2h$  around the threshold  $c$  by weighted least squares. The theoretical properties of local polynomial estimators have been studied extensively (e.g. [Ruppert and Wand, 1994](#); [Fan and Gijbels, 1996](#); [Fan et al., 1997](#)) and it is well understood that if the true CEF differs from a polynomial of order  $p$  on  $[c - h, c + h]$ , the resulting estimator for the kink parameter is generally biased. The two standard approaches to inference in regression kink designs recognize this by combining an asymptotic normality result with an argument that addresses this smoothing bias.

One approach is to choose the bandwidth  $h$  sufficiently small such that for a given sample size  $n$  the resulting bias is (hopefully) negligible relative to the estimator’s standard deviation. This strategy is referred to as undersmoothing in the nonparametric regression literature and uses plug-in estimates of bandwidth sequences that shrink faster than the asymptotic mean squared error (MSE) optimal sequence, eliminating the bias from the asymptotic approximation invoked for inference. In practice, the bandwidth is chosen by multiplying a regularized estimate of the pointwise, that is evaluated at the unknown true CEF, asymptotic MSE optimal bandwidth by  $n^{-\delta}$  for some small  $\delta > 0$  (cf. [Imbens and Lemieux, 2008](#)).

A second approach relies on estimating the leading bias term using a higher order polynomial and constructing the interval around a bias corrected point estimate, taking into account the additional variance introduced by bias estimation in the asymptotic approximation. This approach is referred to as robust bias correction ([Calonico et al., 2014](#)) and is implemented using a pilot bandwidth tuned for bias estimation and a regularized estimate of the pointwise asymptotic MSE-optimal bandwidth for confidence interval (CI) construction.

In a recent paper, [Armstrong and Kolesár \(2020\)](#) demonstrate that both default approaches to nonparametric confidence interval construction can lead to severe undercoverage in finite samples when implemented using bandwidth selectors justified by pointwise asymptotics, as is common. They attribute this finding to the pointwise statistical guarantees underlying these procedures in general, and in particular to the fact that, irrespective of the global curvature of the CEF, the pointwise asymptotic MSE-optimal bandwidth selector can be arbitrarily large if the  $(p + 1)$ -th derivative of the CEF is close to zero at the threshold, resulting in large bandwidth choices for potentially highly nonlinear functions. While this problem was previously recognized, the regularization terms that are added to prevent the empirical bandwidth selectors

from selecting too large bandwidths are sensitive to ad-hoc choices of tuning parameters that drive the finite sample coverage rates of the resulting intervals. As a solution, the authors propose to explicitly restrict the parameter space by placing a bound on the  $(p+1)$ -th derivative of the unknown CEF and select bandwidths according to a minimax criterion, avoiding the need for regularization and allowing for the construction of intervals that are honest, in the sense that they are valid uniformly over the considered parameter space. This approach is feasible, as the bound allows for the computation of the magnitude of the exact worst-case smoothing bias of the local polynomial estimator over the class of functions restricted by the bound, which is taken into account by adjusting critical values accordingly. It is shown that explicitly accounting for possible smoothing bias in this fashion can substantially sharpen inference relative to traditional methods.

This paper proposes confidence intervals for regression kink designs that leverage this potential. Our proposed method is optimal in the sense that it minimizes the interval length amongst all honest procedures that utilize linear estimators and thus improves upon the performance of intervals based on local polynomial regression in a minimax sense. The efficiency gains in terms of the length of two-sided 95% intervals relative to the uniform MSE-optimal local linear<sup>1</sup> intervals for continuous designs are approximately 6 percent, and increase when the assignment variable is discrete. It follows from the arguments in [Armstrong and Kolesár \(2020\)](#) that the efficiency gains relative to undersmoothing and robust bias-correction under valid bandwidth choices are even larger, which we demonstrate in a simulation study and an empirical illustration. Our results suggest that the efficiency gains relative to traditional approaches to nonparametric inference in regression kink designs are substantial. These improvements are particularly valuable for applied work that utilizes RKDs, as power concerns are common and existing procedures that take into account the smoothing bias often yield uninformative confidence intervals despite graphical evidence suggesting the existence of kink effects ([Card et al., 2017](#)). While optimally tuned honest intervals based on local polynomial regression attain on par efficiency in continuous designs, the optimization approach to inference allows us to flexibly incorporate shape constraints that can further shrink the confidence set. We extend the optimization approach to fuzzy designs by employing the inversion strategy proposed in [Noack and Rothe \(2019\)](#) and state conditions that ensure that the optimized linear confidence intervals are honest in the sense of [Li \(1989\)](#). Finally, we illustrate the utility of the

---

<sup>1</sup>We focus on local linear estimators as the relevant benchmark as they are a popular choice in applied work and have leading bias proportional to the second derivative of the unknown function.



procedure in the context of inference on the elasticity of unemployment duration with respect to the benefit level and provide software that implements the procedure.

The key tuning parameter of our proposed method is the bound on the second derivative of the CEF, which governs how much curvature one plausibly allows for. In contrast to undersmoothing and robust bias correction, our method requires that this bound is specified explicitly. However, as pointed out in [Armstrong and Kolesár \(2020\)](#), these methods can not factually avoid this choice, as they must also implicitly restrict a derivative of order two or higher in order to maintain coverage over a given nonparametric class of functions. Once this bound is specified, the proposed method is fully data driven and avoids the choices associated with local polynomial regression regarding the polynomial order, the kernel or the bandwidth. These advantages come at the cost of a closed-form expression for the estimator that we center our interval around, which is defined as the solution to a linear minimax problem over the space of distributions defined by the bound and obtained by numerical convex optimization.

The study of linear minimax estimators for nonparametric regression problems (among other problems) goes back to [Donoho \(1994\)](#), who showed that, under broad conditions, the ratio of the linear minimax risk to the general minimax risk is bounded by 1.25, and that the minimax linear estimator for linear functionals over convex parameter spaces can be obtained via convex optimization. In the context of inference in discontinuity designs, [Armstrong and Kolesár \(2018\)](#) first applied this result to RDDs over a class of functions proposed by [Sacks and Ylvisaker \(1978\)](#) that restricts the approximation error of a Taylor expansion about the discontinuity point<sup>2</sup>. In the same context, [Imbens and Wager \(2019\)](#) impose a bound on the second derivative and propose a numerical optimization strategy to construct honest RDD intervals based on the linear minimax estimator under homoskedasticity, emphasizing the advantages of direct optimization in settings with discrete assignment variables and multivariate assignment rules.

The method proposed in this paper contributes to the methodological literature on inference in regression kink designs ([Calonico et al., 2014](#); [Card et al., 2015b](#); [Ganong and Jäger, 2018](#); [Noack and Rothe, 2019](#)). In implementing our method, we rely on the discretization strategy of [Imbens and Wager \(2019\)](#), which is employed to the dual of the linear minimax problem for kink estimation. Conceptually, this paper builds on the "bias-aware" approach to nonparametric inference in particular ([Armstrong and Kolesár, 2018](#); [Kolesár and Rothe, 2018](#); [Imbens and](#)

---

<sup>2</sup>In this class, the minimax linear estimator can be obtained in closed form. However, since discontinuity designs are predicated on the assumption of continuity of the CEF away from the threshold it is conceptually less appealing than the Hölder class considered in [Imbens and Wager \(2019\)](#), [Armstrong and Kolesár \(2020\)](#) and this paper. Moreover, permitting functions that are discontinuous away from the threshold also leads to inference that is too conservative at smooth functions, as the worst-case bias is larger in this class.

Wager, 2019; Ignatiadis and Wager, 2019; Noack and Rothe, 2019; Rambachan and Roth, 2019; Schennach, 2020; Armstrong and Kolesár, 2020) and a large methodological literature on inference in discontinuity designs in general (cf. Lee and Lemieux, 2010; Cattaneo et al., 2019).

The remainder of this paper is structured as follows: Section 1.2 introduces our notation, sketches the identification result underlying regression kink designs and defines the parameter of interest. Section 1.3 formally states the objective of this paper, explains how the bound permits a bias characterization that facilitates the implementation of the procedure, states assumptions under which optimized linear confidence intervals attain honesty and discusses practical questions related to their implementation. In Section 1.4, we conduct a simulation study to investigate the performance of optimized linear intervals relative to a variety of methods based on local linear regression. In Section 1.5, we demonstrate the utility of our method in a sensitivity analysis to Landais (2015). Section 1.6 concludes. The Appendix contains proofs and additional results.

## 1.2 Framework and Notation

**1.2.1 Setup.** We observe a random sample of  $n$  independent pairs  $(X_i, Y_i)$ , where  $Y_i \in \mathbb{R}$  is the outcome of interest and  $X_i \in \mathbb{R}$  is the assignment variable for unit  $i$ . We write the nonparametric regression model as

$$Y_i = \mu(X_i) + u_i \quad \mathbb{E}[u_i|X_i] = 0 \quad \mathbb{V}[u_i|X_i] = \sigma_i^2 \quad i = 1, \dots, n,$$

so that  $\mu(X_i) = \mathbb{E}[Y_i|X_i]$  denotes the conditional expectation function of the outcome given the assignment variable. The subscript  $N$  indicates a column vector of length  $n$  where the  $i$ -th element corresponds to individual  $i$ , such that e.g.  $X_N = (X_1, X_2, \dots, X_n)^T$ . We normalize the threshold  $c$  to zero and define an indicator function  $D(x) = \mathbb{1}_{[x \geq 0]}$ . For a general function  $f(x)$  we write  $f_+(x) = f(x)D(x)$  and  $f_-(x) = f(x)(1 - D(x))$  and denote the  $j$ -th derivative of  $f(x)$  with respect to  $x$  by  $f^j(x)$ , where  $f^0(x)$  is understood to mean the function itself. Moreover, let  $f_{\pm}^j(0) = \lim_{\pm x \downarrow 0} f_{\pm}^j(x)$  so that in particular  $\mu_+^1(0) = \lim_{x \downarrow 0} \mu_+^1(x)$  and  $\mu_-^1(0) = \lim_{x \uparrow 0} \mu_-^1(x)$ . Let  $R_X$  denote the support of  $X_i$ , i.e. the smallest closed set in  $\mathbb{R}$  such that  $\Pr\{X_i \in R_X\} = 1$ . We assume that  $R_X$  covers zero, is bounded and denote by  $\underline{x}$  and  $\bar{x}$  its minimal and maximal element. Let  $\mathcal{X}_- = [\underline{x}, 0)$  and  $\mathcal{X}_+ = (0, \bar{x}]$ . We assume that the CEF  $\mu$  is a member of the class

$$\mathcal{F}(L) = \{f : |f_{\pm}^1(x) - f_{\pm}^1(x')| \leq L|x - x'|, (x, x') \in \mathcal{X}_{\pm}\},$$

which formalizes the notion that  $\mu$  is two times differentiable on either side of the threshold, potentially discontinuous at the threshold, and has second derivative bounded by  $L$  uniformly over  $\mathcal{X} = \mathcal{X}_+ \cup \mathcal{X}_-$ . The bound  $L$  effectively governs how much curvature one allows for, as values of  $L$  close to zero imply that the members of  $\mathcal{F}(L)$  are close to linear, while larger values of  $L$  allow for increasing amounts of curvature. Given  $L$  and  $\sigma_N$ , the method that is proposed in this paper is fully data-driven and we will assume throughout the derivation that both are known. The choice of  $L$  and estimation of  $\sigma_N$  are discussed separately in Section 1.3.7.

Since conditional expectation functions are unique only over the support of the conditioning variable, we follow [Kolesár and Rothe \(2018\)](#) in that, assuming  $\mu \in \mathcal{F}(L)$  is understood to mean that there exist  $\mu$  in  $\mathcal{F}(L)$  such that  $\Pr\{\mu(X_i) = \mathbb{E}[Y_i|X_i]\} = 1$ . While the canonical regression kink design is predicated on continuous assignment variables, our setup therefore does not assume the assignment variable to be of any specific type. This is advantageous in settings in which we only have coarse measurements at our disposal, in particular if the support of the assignment variable does not contain an open neighborhood around the threshold. In such situations, the bound  $L$  allows for extrapolation that ensures meaningful partial identification of  $\mu_+^1(0)$  and  $\mu_-^1(0)$ , as discussed in Section 1.3.6.

**1.2.2 Parameter of Interest.** Regression Kink Designs consider structural models of the form

$$Y = Y(T, X, E) \quad T = T(X),$$

where the function  $Y : \mathbb{R}^3 \mapsto \mathbb{R}$  describes how the outcome is produced, the random variable  $E \in \mathbb{R}$  aggregates unobserved influences potentially correlated with the assignment variable, and the policy function  $T : \mathbb{R} \mapsto \mathbb{R}$  is differentiable away from a kink location normalized to zero. The parameter of interest is the average partial effect of the policy variable at the kink

$$\tau_{RKD} = \mathbb{E} \left[ \frac{\partial Y(T, X, E)}{\partial T} \Big| X = 0 \right],$$

which inherits its causal interpretation from the definition of  $Y$ . Let  $f_{E|X}(e|x)$  denote the conditional probability density function of  $E$  given  $X = x$ . Under regularity conditions, we can write the first derivative of the CEF in this framework as

$$\mu^1(x) = \mathbb{E} \left[ \frac{\partial Y(T, X, E)}{\partial T} \Big| X = x \right] T^1(x) + \mathbb{E} \left[ \frac{\partial Y(T, X, E)}{\partial X} \Big| X = x \right] + \mathbb{E} \left[ Y \frac{\partial \ln f_{E|X}(e|X)}{\partial X} \Big| X = x \right].$$

Since  $T$  is kinked at zero, this decomposition implies that, if the average partial effect of the assignment variable and the distribution of unobservables are continuous at zero,  $\tau_{RKD}$  is identified by

$$\tau_{RKD} = \frac{\mu_+^1(0) - \mu_-^1(0)}{T_+^1(0) - T_-^1(0)}.$$

Card et al. (2015b) characterize models for which this is the case and discuss conditions under which  $\tau_{RKD}$  is equivalent to the "treatment on the treated" parameter in Florens et al. (2008) or the "local average response" parameter in Altonji and Matzkin (2005), respectively. In the next section, we focus on the sharp RKD, that is we assume that the denominator of  $\tau_{RKD}$  is known, such that inference on  $\tau_{RKD}$  is solely concerned with the jump in the first derivative of  $\mu(x)$ ,  $\theta = \mu_+^1(0) - \mu_-^1(0)$ , which we refer to as the kink parameter. In Appendix 1.F, we extend our method to fuzzy designs by using a strategy recently proposed by Noack and Rothe (2019).

## 1.3 Optimized Honest Confidence Intervals

**1.3.1 Problem Statement.** We seek to construct efficient confidence intervals of the form  $\mathcal{I}_\alpha = [\underline{\theta}, \bar{\theta}]$  which cover the kink parameter  $\theta$  with at least probability  $1 - \alpha$  for some prespecified level  $\alpha > 0$  in large samples. Furthermore, we strengthen this requirement by demanding our confidence intervals to be honest in the sense of Li (1989) with respect to the class  $\mathcal{F}(L)$

$$\liminf_{n \rightarrow \infty} \inf_{\mu \in \mathcal{F}(L)} \Pr \left[ \mu_+^1(0) - \mu_-^1(0) \in \mathcal{I}_\alpha \right] \geq 1 - \alpha.$$

The uniform requirement imposed by honesty disciplines our inference in the sense that it requires us to specify and take into account plausible adversarial distributions in our asymptotic approximation. In the present setting, this means to specify  $L$  and guarantee coverage for the worst-case function in  $\mathcal{F}(L)$ . Honesty ensures that, for any tolerance level  $\eta$ , we can find a sample size  $n_\eta$  such that for  $n > n_\eta$  coverage of  $\mathcal{I}_\alpha$  is above  $1 - \alpha - \eta$  for all  $f \in \mathcal{F}(L)$ . As discussed in Armstrong and Kolesár (2020), the requirement to explicitly specify  $L$  is not a disadvantage of uniform procedures, as methods that rely on pointwise guarantees of the type

$$\text{for every } f \in \mathcal{F}, \liminf_{n \rightarrow \infty} \Pr \left[ \mu_+^1(0) - \mu_-^1(0) \in \mathcal{I}_\alpha \right] \geq 1 - \alpha,$$

must implicitly restrict  $L$  to justify that coverage is controlled over a given function class  $\mathcal{F}$ . This holds true, irrespective of the regularization problem that such procedures need to solve.

**1.3.2 General Approach.** We center the confidence interval around a linear estimator of the form  $\hat{\theta} = \sum_{i=1}^n w_i Y_i$ , with weights  $w_N$  optimized for a uniform criterion. The estimator is linear in the sense that the weights depend only on  $X_N$  and a non-random tuning parameter  $\kappa > 0$ , which we keep implicit in our notation. This choice is motivated by the relative minimax-efficiency result in [Donoho \(1994\)](#) and the fact that the mean squared error of linear estimators, and related quantities governing the length of our confidence intervals, depend on the unknown function only through the bias. In order to ensure honesty of the enclosing interval, we compute the magnitude of the exact worst-case conditional bias  $\bar{B}(w_N) = \sup_{\mu \in \mathcal{F}(L)} \mathbb{E}[\hat{\theta} - \theta | X_N]$  of the estimator over  $\mathcal{F}(L)$  during the optimization, and inflate the interval width accordingly. Following [Imbens and Wager \(2019\)](#), we obtain the estimate for the kink parameter by numerical convex optimization, solving a version of the linear minimax problem under known variance  $\sigma_N^2$

$$\min_{w_N \in \mathbb{R}^n} \sum_{i=1}^n w_i^2 \sigma_i^2 + \kappa \bar{B}(w_N)^2 \quad \bar{B}(w_N) = \sup_{\mu \in \mathcal{F}(L)} \left\{ \sum_{i=1}^n w_i \mu(X_i) - (\mu_+^1(0) - \mu_-^1(0)) \right\}. \quad (1.1)$$

Note that, since the class  $\mathcal{F}(L)$  is symmetric with respect to zero, we do not need an absolute value inside the supremum, as the worst-case negative and positive biases over  $\mathcal{F}(L)$  have the same magnitude. For  $\kappa = 1$ , the objective function thus corresponds to the uniform conditional MSE of  $\hat{\theta}$  over  $\mathcal{F}(L)$ . In general,  $\kappa$  governs the worst-case conditional bias-variance tradeoff and is either chosen to minimize the uniform MSE or the interval length. Since the solution of (1.1) depends on the data only through  $X_N$ , the confidence interval obtained by optimizing the interval length over  $\kappa$  provides the same statistical guarantee as those obtained for a fixed  $\kappa$ .

**1.3.3 Bias Characterization.** In order to translate (1.1) into a tractable optimization problem we rely on the restrictions defining  $\mathcal{F}(L)$ . For any  $\mu \in \mathcal{F}(L)$  and  $X_i \in R_X$  we can write

$$\mu_+(X_i) = \mu_+(0) + \mu_+^1(0)X_i + R_+(X_i) \quad \mu_-(X_i) = \mu_-(0) + \mu_-^1(0)X_i + R_-(X_i),$$

where  $R_+$  and  $R_-$  denote the remainders of the expansions. Note that by definition  $R(0) = 0$ ,  $R^1(0) = 0$ , and  $|R^2(x)| \leq L$  for all  $\mu \in \mathcal{F}(L)$  and  $x \in \mathcal{X}$ . The worst-case conditional bias  $\bar{B}(w_N)$  of a general linear estimator of  $\theta$  over  $\mathcal{F}(L)$  is

$$\sup_{\mu \in \mathcal{F}(L)} \left\{ \sum_{i=1}^n w_{i,+} \left[ \mu_+(0) + \mu_+^1(0)X_i + R_+(X_i) \right] + \sum_{i=1}^n w_{i,-} \left[ \mu_-(0) + \mu_-^1(0)X_i + R_-(X_i) \right] - \theta \right\}.$$

Since the assumption  $\mu \in \mathcal{F}(L)$  does not impose any constraints on  $(\mu^0, \mu^1)$  at zero, this expression is infinite unless the following discrete moment conditions are satisfied

$$\sum_{i=1}^n w_{i,+} = 0 \quad \sum_{i=1}^n w_{i,+} X_i = 1 \quad \sum_{i=1}^n w_{i,-} = 0 \quad \sum_{i=1}^n w_{i,-} X_i = -1.$$

As a consequence, any solution to (1.1) must satisfy these constraints, a fact that we utilize in the implementation of our estimator. We refer to a linear estimator and the corresponding set of weights with associated  $X_N$  as "of the correct order" if these constraints are met, in which case the conditional bias of  $\hat{\theta}$  is given by the weighted sum of approximation errors of the Taylor approximation to  $\mu_{\pm}$  near zero. Since  $\mu \in \mathcal{F}(L)$  implies that  $\mu^1$  is absolutely continuous, an integration by parts argument allows further characterization of the conditional bias under the constraints, yielding

$$\mathbb{E}[\hat{\theta} - \theta | X_N] = \sum_{i=1}^n w_i R(X_i) = \sum_{i=1}^n w_{i,+} \int_0^{X_i} \mu^2(t)(X_i - t) dt - \sum_{i=1}^n w_{i,-} (X_i) \int_{X_i}^0 \mu^2(t)(X_i - t) dt.$$

Applying an argument based on Fubini's Theorem then obtains

$$\mathbb{E}[\hat{\theta} - \theta | X_N] = \int_0^{\infty} \mu^2(t) \sum_{i: X_i \in [t, \infty)} w_{i,+} (X_i - t) dt - \int_{-\infty}^0 \mu^2(t) \sum_{i: X_i \in (-\infty, t]} w_{i,-} (X_i - t) dt.$$

This representation of the conditional bias characterizes the choice of an adversarial nature that needs to pick  $\mu$  out of  $\mathcal{F}(L)$  in response to  $(w_N, X_N)$  for weights that satisfy the above constraints. Let  $\bar{w}(t) = D(t) \sum_{i: X_i \geq t} w_{i,+} (X_i - t) - (1 - D(t)) \sum_{i: X_i < t} w_{i,-} (X_i - t)$ . In this notation, the conditional bias is  $\int_{\mathbb{R}} \mu^2(t) \bar{w}(t) dt$ , which an adversarial nature maximizes by setting  $\mu^2(t) = \text{sign}(\bar{w}(t)) L$ , yielding

$$\bar{B}(w_N) = L \int_{\mathbb{R}} |\bar{w}(t)| dt. \tag{1.2}$$

It follows that the worst-case conditional bias of a linear estimator, if it is finite, is proportional to  $L$  and that the worst-case function in  $\mathcal{F}(L)$  at which it is attained is a quadratic spline with piecewise constant second derivative of magnitude  $L$ . Another consequence of (1.2) is that, for any given set of weights of the correct order and bound  $L$ , the computation of  $\bar{B}(w_N)$  amounts to finding the roots of  $\bar{w}(t)$ . This has practical value, as the sign of  $\bar{w}(t)$  is known for local polynomial estimators under regularity conditions, a fact we utilize in our estimation strategy.

**1.3.4 Estimation via Dual Optimization.** Our approach to implementing the optimized linear confidence intervals relies on a renormalization justified by the result in (1.2) and the convexity of (1.1). From equation (1.2) it follows that for linear estimators with finite worst-case bias, it holds that  $\sup_{\mu \in \mathcal{F}(L)} \mathbb{E}[\hat{\theta} - \theta | X_N] = L \sup_{R \in \bar{\mathcal{F}}(1)} \sum_{i=1}^n w_i R(X_i)$  with  $\bar{\mathcal{F}}(1)$  defined as

$$\bar{\mathcal{F}}(1) = \{f : f(0) = f^1(0) = 0 \wedge |f_{\pm}^1(x) - f_{\pm}^1(x')| \leq |x - x'|, (x, x') \in \mathcal{X}_{\pm}\}.$$

The class  $\bar{\mathcal{F}}(1)$  can be understood as the class of remainder functions corresponding to the conditional expectation functions in  $\mathcal{F}(1)$ , which is reflected by the additional constraints that correspond to the aforementioned properties of remainders. The renormalization allows us to equivalently state the optimization problem defining  $\hat{\theta}$  as follows. For a fixed  $\kappa$ , we write the primal problem as

$$\begin{aligned} & \underset{w_N, r}{\text{minimize}} && \sum_{i=1}^n w_i^2 \sigma_i^2 + \kappa L^2 r^2 \\ & \text{subject to} && \sup_{R \in \bar{\mathcal{F}}(1)} \left[ \sum_{i=1}^n w_i R(X_i) \right] \leq r \\ & && \sum_{i=1}^n w_{i,+} = 0 \quad \sum_{i=1}^n w_{i,-} = 0 \\ & && \sum_{i=1}^n w_{i,+} X_i = 1 \quad \sum_{i=1}^n w_{i,-} X_i = -1. \end{aligned} \tag{1.3}$$

The weights solving problems (1.1) and (1.3) are equivalent since, at the optimal value of  $r$ , the objective functions are identical and any candidate solution to (1.1) lies in the feasible set for  $w_N$  of (1.3). This reformulation is helpful, as we require the optimal weights as well as a sharp uniform upper bound on  $\mathbb{E}[\hat{\theta} - \theta | X_N]$  to construct our confidence interval and, more importantly, rely on the additional constraints of  $\bar{\mathcal{F}}(1)$  in our implementation. The Lagrangian of (1.3) is given by

$$\begin{aligned} L(w_N, r, \nu, \lambda) = & \sup_{R \in \bar{\mathcal{F}}(1)} \sum_{i=1}^n w_i^2 \sigma_i^2 + \kappa L^2 r^2 + \nu \left( \sum_{i=1}^n w_i R(X_i) - r \right) + \lambda_1 \left( \sum_{i=1}^n w_{i,+} \right) \\ & + \lambda_2 \left( \sum_{i=1}^n w_{i,-} \right) + \lambda_3 \left( \sum_{i=1}^n w_{i,+} X_i - 1 \right) + \lambda_4 \left( \sum_{i=1}^n w_{i,-} X_i + 1 \right). \end{aligned}$$

In order to solve (1.3) we rely on a second equivalence result. In Appendix 1.B, we show that the local polynomial weights with corresponding worst-case conditional bias lie in the feasible set of (1.3), implying that a refined Slater's condition applies to (1.3). As a consequence, strong

duality holds and any primal optimal point is also a minimizer of  $L(w_N, r, \nu^*, \lambda^*)$  where  $(\nu^*, \lambda^*)$  is the solution to the dual problem

$$\begin{aligned} \underset{\nu, \lambda}{\text{maximize}} \quad & q(\nu, \lambda) = \inf_{w_N, r} L(w_N, r, \nu, \lambda) \\ \text{subject to} \quad & \nu \in \mathbb{R}_+^1, \lambda \in \mathbb{R}^4. \end{aligned}$$

In Appendix 1.C, we show that we can interchange the order of the infimum and the supremum in the dual objective by applying a minimax theorem, yielding an inner convex quadratic minimization problem that is solved analytically. This results in closed-form expressions for the primal parameters as functions of the dual parameters and a remainder function  $R \in \bar{\mathcal{F}}(1)$ ,

$$w_i = \frac{\lambda_1 D(X_i) + \lambda_2(1 - D(X_i)) + \lambda_3 D(X_i)X_i + \lambda_4(1 - D(X_i))X_i + \nu R(X_i)}{-2\sigma_i^2} \quad r = \frac{\nu}{2\kappa L^2} \quad (1.4)$$

as well as a simplified expression for the dual objective

$$\begin{aligned} q(\nu, \lambda) = \sup_{R \in \bar{\mathcal{F}}(1)} & -\frac{1}{4} \sum_{i=1}^n \frac{[\lambda_1 D(X_i) + \lambda_2(1 - D(X_i)) + \lambda_3 D(X_i)X_i + \lambda_4(1 - D(X_i))X_i + \nu R(X_i)]^2}{\sigma_i^2} \\ & -\frac{1}{4} \frac{\nu^2}{\kappa L^2} - \lambda_3 + \lambda_4. \end{aligned}$$

Let  $(w^*, r^*) = (w(\nu^*, \lambda^*), r(\nu^*, \lambda^*))$  for the element of  $\bar{\mathcal{F}}(1)$  that attains the supremum in  $q(\nu, \lambda)$ . It follows from strong duality and strict convexity of  $L(w, r, \nu^*, \lambda^*)$  that  $(w^*, r^*)$  is the solution of (1.3). As a consequence, we can recover the weights solving (1.1) as well as the associated worst-case conditional bias over  $\mathcal{F}(L)$  by the solution of the simplified dual problem

$$\begin{aligned} \underset{\nu, \lambda, R}{\text{maximize}} \quad & -\frac{1}{4} \sum_{i=1}^n \frac{[\lambda_1 D(X_i) + \lambda_2(1 - D(X_i)) + \lambda_3 D(X_i)X_i + \lambda_4(1 - D(X_i))X_i + \nu R(X_i)]^2}{\sigma_i^2} \\ & -\frac{1}{4} \frac{\nu^2}{\kappa^2 L^2} - \lambda_3 + \lambda_4. \\ \text{subject to} \quad & \nu \in \mathbb{R}_+^1, \lambda \in \mathbb{R}^4, R \in \bar{\mathcal{F}}(1), \end{aligned} \quad (1.5)$$

in conjunction with the mapping (1.4) between primal and dual parameters and the result (1.2). This translates the primal problem of  $n + 1$  parameters into a problem over the space  $\bar{\mathcal{F}}(1)$  and five dual parameters, which we solve numerically by discretization as described in Appendix 1.D.

*Remark 1.* Abstracting from the constraint  $R \in \bar{\mathcal{F}}(1)$ , the dual problem (1.5) is a standard quadratic program and the remaining challenge is to find a suitable approximation strategy to the functional constraint. In our implementation, we approximate the function on



an equidistant grid to permit approximation of the second order constraint via finite central differences, e.g.  $R^2(x) = [R(x+h) - 2R(x) + R(x-h)]/h^2 + O(h^2)$  (see Appendix 1.D for details). This approach is formally justified by Proposition 2 of Imbens and Wager (2019), which states that, for assignment variables with compact and convex support, the optimal weights can be recovered with arbitrary small  $L_2$ - error under the proposed discretization strategy.

*Remark 2.* Under the proposed strategy, the optimization approach to bias-aware inference allows us to incorporate shape constraints in a simple fashion. As explained in detail in Appendices 1.D and 1.F, any additional constraint on the CEF that can be approximated in terms of finite differences of Taylor remainders can be utilized by modifying the feasible set of (1.5).

**1.3.5 Interval Construction.** Given a solution  $(\nu^*, \lambda^*, R^*)$  to (1.5) for a fixed value of  $\kappa$ , we recover the optimal weights  $w_N \leftarrow w^*$  and the corresponding worst-case bias magnitude  $\bar{B}(w_N) \leftarrow r^*L$  via the mapping (1.4) and the result (1.2) to construct the optimized linear interval. Intuitively, the construction relies on the following decomposition of  $\hat{\theta} - \theta$

$$\hat{\theta} - \theta = \underbrace{\sum_{i=1}^n w_i R(X_i)}_{=\mathbb{E}[\hat{\theta} - \theta | X_N]} + \underbrace{\sum_{i=1}^n w_i u_i}_{=\hat{\theta} - \mathbb{E}[\hat{\theta} | X_N]} .$$

By definition,  $\mathbb{E}[\hat{\theta} - \theta | X_N]$  is bounded in absolute value by  $\bar{B}(w_N) = \sup_{\mu \in \mathcal{F}(L)} E[\hat{\theta} - \theta | X_N]$  uniformly over  $\mathcal{F}(L)$ . Let  $s_n^2 = \sum_{i=1}^n w_i^2 \sigma_i^2$  denote the conditional variance of  $\hat{\theta}$  given  $X_N$  and define the conditional bias to standard deviation ratio  $t_n = s_n^{-1} \mathbb{E}[\hat{\theta} - \theta | X_N]$ . The uniform bound implies that the t-statistic

$$\frac{\hat{\theta} - \theta}{s_n} = \frac{\mathbb{E}[\hat{\theta} - \theta | X_N]}{s_n} + \frac{\hat{\theta} - \mathbb{E}[\hat{\theta} | X_N]}{s_n}$$

is the sum of a term that is bounded in absolute value by  $\bar{t}_n = \bar{B}(w_N) s_n^{-1}$  uniformly over  $\mathcal{F}(L)$  and a term that, under conditions stated in the next section, converges to a standard normal distribution uniformly over  $\mathcal{F}(L)$  by a suitable central limit theorem. Provided that this is the case, it follows that an honest  $(1 - \alpha)$  confidence interval for  $\theta$  is given by

$$\mathcal{I}_\alpha = \left[ \hat{\theta} \pm s_n \text{cv}_{1-\alpha}(\bar{t}_n) \right], \quad (1.6)$$

where  $\text{cv}_{1-\alpha}$  denotes the  $(1 - \alpha)$  quantile of the folded normal distribution  $|N(\bar{t}_n, 1)|$  with mean  $\bar{t}_n$  and variance one, that is the distribution of the absolute value of a normal distribution with

mean  $\bar{t}_n$  and variance one. Intuitively, this construction works because the bias can not be negative and positive at the same time, which implies that a hypothetical interval that adds and subtracts  $\bar{B}(w_N) + z_{1-\alpha/2}s_n$  from  $\hat{\theta}$  would be too conservative. For the sharp<sup>3</sup> regression kink design, an honest  $(1 - \alpha)$  interval for  $\tau_{RKD}$  is then immediately obtained by rescaling the upper and lower ends of  $\mathcal{I}_\alpha$  by the inverse of the magnitude of the kink.

We construct two types of optimized linear confidence intervals according to (1.6): Uniform MSE-optimal ( $\kappa_{UMSE} = 1$ ) and length-optimal ( $\kappa_{LE} = \arg \min_{\kappa > 0} s_n \text{cv}_{1-\alpha}(\bar{t}_n)$ ) intervals. However, the same construction principle can be applied to any uniform performance criterion that specifies a worst-case bias-variance trade-off, as the guarantees of (1.6) hold for any fixed  $\kappa > 0$ .

*Remark 3.* In order to obtain the length-optimal interval, we search for the optimal value  $\kappa_{LE}$  using a combination of golden section search and successive parabolic interpolation as implemented in standard derivative free optimization libraries. This is feasible at high accuracy in practice as the runtime of a single optimization iteration as implemented is low (approximately 0.129 seconds for a sample size of 6000), leading to an average total runtime of 3.59 seconds for the same sample size on a standard desktop computer (see Appendix 1.G for more details). This could in principle be further improved by restricting attention to values of  $\kappa$  smaller than the uniform MSE-optimal choice  $\kappa = 1$ . This is because the length-optimal weights will "over-smooth" relative to the uniform MSE-optimal weights, which [Armstrong and Kolesár \(2020\)](#) show for estimators in their regularity class (cf. Figure 1 therein).

**1.3.6 Theoretical Properties.** In order to discuss the statistical properties of confidence intervals constructed according to (1.6) we impose the following assumptions.

**Assumption 1** Let  $(C, \delta, \sigma_{\min}, \sigma_{\max}) \in \mathbb{R}_+^4$  be some fixed vector.

- (i)  $\{Y_i, X_i\}_{i=1}^n$  is an i.i.d. random sample of size  $n$  from a fixed population.
- (ii)  $\mu \in \mathcal{F}(L)$  for some  $L > 0$ .
- (iii)  $0 < \sigma_{\min}^2 \leq E[(Y_i - \mu(X_i))^2 | X_i = x] \leq \sigma_{\max}^2$  for all  $x \in R_X$  and  $\mu \in \mathcal{F}(L)$ .
- (iv)  $E[|Y_i - \mu(X_i)|^{2+\delta} | X_i = x] \leq C$  for all  $x \in R_X$  and  $\mu \in \mathcal{F}(L)$ .
- (v) The solution  $w_N^*$  satisfies
 
$$\frac{\max_i w_i^2}{\sum_{i=1}^n w_i^2} \xrightarrow{P} 0.$$

---

<sup>3</sup>In fuzzy discontinuity designs, the strategy underlying the construction of the interval (1.6) can also be employed in principle. However, in this case the smoothing bias of the first stage estimator must be dealt with and additional problems arise. In Appendix 1.F, we discuss how our implementation utilizes the strategy proposed in [Noack and Rothe \(2019\)](#) to extend the optimization approach to fuzzy discontinuity designs.

Assumption 1 is sufficient for a central limit theorem to apply to  $s_n^{-1}w_N^T u_N = s_n^{-1}(\hat{\theta} - \mathbb{E}[\hat{\theta}|X_N])$  uniformly over  $\mathcal{F}(L)$  and ensures the consistency of  $\hat{\theta}$  in the identified setting. Part (i) is the standard model for survey data and provides that the kink parameter is a well defined quantity. Part (ii) and (iii) ensure that the quadratic program defining  $\hat{\theta}$  is strictly convex, which guarantees that the optimal weights are uniquely recovered by the dual optimal parameters, provided that the data contains at least two distinct points on either side of the threshold.

Assumptions (iii) and (iv) guarantee the existence of and establish bounds on the second and  $(2 + \delta)$ -th absolute conditional moment functions of the CEF error uniformly over the support of the assignment variable and the class of permitted CEFs. The two assumptions restrict the class of permitted distributions beyond the CEF constraint (ii) in that they require uniformly bounded and non-zero conditional variances as well as the existence of a strictly finite higher order moment function. Assumptions 1 (i)-(v) are sufficient to establish that Lyapunov's condition applies to each element of  $\mathcal{F}(L)$ , implying convergence of  $s_n^{-1}w_N^T u_N$  to a standard normal variable uniformly over  $\mathcal{F}(L)$ .

Assumption (v) restricts the limit behavior of the set of optimal weights and is difficult to derive from higher-level conditions. This is because, to the best of our knowledge, a closed form solution to (1) is not known and a general characterization of  $w_N^*$  beyond the spline property derived in Section 1.3 is difficult. While this is unattractive from a theoretical point of view, the good news is that we can verify that the condition is approximately met in any given application, and our implementation reports the finite sample counterparts to Assumption (v). Assumption (v) together with (iii) is sufficient for  $s_n^2 = o_p(1)$  uniformly over  $\mathcal{F}(L)$  and implies consistency of  $\hat{\theta}$  in the identified setting.

Roughly speaking, Assumption 1 rules out distributions such that, for some  $X_i \in R_X$  and  $\mu \in \mathcal{F}(L)$ , in the limit  $w_i u_i$  is "too large", in the sense that it dominates the behavior of the sequence  $s_n^{-1}w_N^T u_N$ . This rules out that only a "small" proportion of the data is driving the estimate under this function. Appendix 1.E contains a formal discussion of how the relevant components of Assumption 1 can be used to show that Lyapunov's condition holds conditionally on  $X_N$  uniformly over  $\mathcal{F}(L)$ , which is the key ingredient in ensuring that the optimized interval attains honesty. Once uniform convergence of  $s_n^{-1}[\hat{\theta} - \mathbb{E}[\hat{\theta}|X_N]] \xrightarrow{D} N(0, 1)$  is established, it follows from the definitions of the worst-case bias to standard deviation ratio  $\bar{t}_n$  and the critical

value  $cv_{1-\alpha}$  that, for a standard normal random variable  $Z \sim N(0, 1)$ , it holds uniformly over  $\mathcal{F}(L)$  that

$$\liminf_{n \rightarrow \infty} \Pr \left( |Z + t_n| \leq cv_{1-\alpha}(\bar{t}_n) | X_n \right) \geq 1 - \alpha.$$

Taken together, the two results imply that (1.6) is honest, which we record in Proposition 1.

**Proposition 1** Suppose that Assumption 1 holds. Then uniformly over  $\mathcal{F}(L)$

$$s_n^{-1}[\hat{\theta} - \mathbb{E}[\hat{\theta} | X_N]] \xrightarrow{D} N(0, 1) \quad \text{and} \quad t_n \leq \bar{t}_n = s_n^{-1} \bar{B}(w_N),$$

and the interval  $\mathcal{I}_\alpha = [\hat{\theta} \pm s_n cv_{1-\alpha}(\bar{t}_n)]$  satisfies

$$\liminf_{n \rightarrow \infty} \inf_{\mu \in \mathcal{F}(L)} \Pr \left[ \mu_+^1(0) - \mu_-^1(0) \in \mathcal{I}_\alpha \right] \geq 1 - \alpha$$

for  $t_n$ ,  $\bar{B}(w_N)$  and  $cv_{1-\alpha}$  as defined in Section 1.3.5.

The relevant difference of the statistical guarantee given in Proposition 1 relative to those that pointwise approaches to nonparametric confidence interval construction rely on is as follows: It ensures that, for any tolerance level  $\eta$ , one can find a sample size  $n_\eta$  such that for all  $n > n_\eta$  coverage is at least  $1 - \alpha - \eta$  for *all* functions in  $\mathcal{F}(L)$ . In contrast, pointwise procedures can not generally ensure the existence of such a sample size for any given non-trivial tolerance level without restricting the curvature or a higher order derivative, since the true CEF is unknown. Thus, their coverage properties can theoretically be poor even in large samples. Consequently, the uniform guarantee provided by honesty is required for reliably good finite sample performance. Once such a restriction is imposed, it follows from the definition of optimized intervals that they are the minimax optimal choice (under known variances) amongst all linear intervals.

Another attractive property of (1.6) and "bias-aware" intervals in general is that they remain valid, irrespective of whether the assignment variable has support arbitrarily close to the threshold or not, in the sense that the statistical guarantee of the interval remains the same. In settings in which this is not the case, the interval will have positive length in the limit, but not necessarily cover the whole identification interval, that is the interval of values for  $\theta$  that are consistent with the distribution  $(X, Y)$  and the restriction imposed by  $\mu \in \mathcal{F}(L)$ . Partial identification intervals of this type were proposed in [Imbens and Manski \(2004\)](#) and are also

useful in other non-standard situations, e.g. if one wants to exclude data for reasons such as data entry errors or other institutional characteristics that could justify such a choice.

In order to discuss two potential threats to the quality of the approximation underlying the construction of (1.6), we further introduce the following two assumptions.

**Assumption 2** Let  $C_1 \in \mathbb{R}_+$  be fixed and  $\hat{s}_n$  denote an estimator at our disposal.

- (i) Assumption 1 holds with  $\delta = 1$ .
- (ii) Assumption 1 holds and  $\hat{s}_n - s_n = o_p(1)$  uniformly over  $\mathcal{F}(L)$ .

Assumption 2 (i) allows us to clarify the role of the ratio  $\bar{w}_R = [\max_i |w_i|] [\sum_{i=1}^n |w_i|]^{-1}$  addressed by Assumption 1 (v) with respect to the quality of the normal approximation underlying (1.6). It follows from the Berry-Essen Theorem (cf. Theorem 3 in De Brabanter et al., 2013) that under Assumption 2 (i)

$$\sup_{z \in \mathbb{R}^1} \sup_{\mu \in \mathcal{F}(L)} \left| \Pr \left[ \left( \frac{\hat{\theta} - \mathbb{E}[\hat{\theta} | X_N]}{s_n} \right) \leq z \mid X_N \right] - \Phi(z) \right| \leq D \frac{C}{\sigma_{min}^3} \frac{\max_i |w_i|}{\sum_{i=1}^n |w_i|},$$

where  $\Phi$  denotes the standard normal CDF and the constant  $D$  lies in  $0.4097 < D \leq 0.56$ . This illustrates the role of Assumption 1 (v) in ensuring the quality of the distributional approximation. In particular, under the maintained assumptions, one would expect that the finite sample coverage rate of (1.6) at the worst-case function is close to nominal whenever the ratio  $\bar{w}_R$  is small. We therefore report  $\bar{w}_R^2$  as a diagnostic statistic in our implementation and recommend to verify that this is the case in practice. If the ratio is "large", in the sense that the weights concentrate on a small set of observations, it is recommended to modify  $\kappa$ . In doing so, one effectively trades the quality of the estimator resulting from the initial choice of  $\kappa$  in terms of the respective performance criterion for an improvement in the quality of the distributional approximation.

Assumption 2 (ii) emphasizes that the honesty property of (1.6) was derived under known variances and that, in principle, one needs an appropriate estimator for  $\sigma_N$  in order to preserve honesty under estimated conditional variance. Assumption 2 (ii) provides that such a uniformly consistent estimator of  $s_n$  is available. In this case, the feasible interval that replaces  $s_n$  with  $\hat{s}_n$  remains asymptotically uniformly valid. This is pointed out because commonly used estimators for the conditional variance have a leading bias that is proportional to  $\mu^1$ , which is unrestricted over  $\mathcal{F}(L)$ . Noack and Rothe (2019) propose an estimator for the conditional variance based

on a regression adjusted version of the nearest-neighbor estimator of [Abadie et al. \(2014\)](#) that has leading bias proportional to  $\mu^2$  and can thus preserve honesty under second order bounds. Our implementation contains their proposed estimator as well as standard estimators of  $s_n$ .

Note that the estimator solving (1.1) is the finite-sample minimax linear estimator of  $\theta$  over  $\mathcal{F}(L)$  only under known conditional variances  $\sigma_N^2$ . If the conditional variances need to be estimated, the estimator is no longer guaranteed to achieve the minimax risk in finite samples. However, in the case of homoskedasticity, it suffices to estimate  $\sigma_N^2$  by an efficient estimator for the conditional variance to obtain honest and asymptotically minimax optimal intervals. Under heteroskedasticity, a uniformly consistent estimator for  $\sigma_N^2$  is required to maintain honesty as discussed above, and the minimax properties of the estimator solving (1.1) depend on this choice.

**1.3.7 Practical Implementation.** So far we have assumed that the curvature bound  $L$  and the conditional variances  $\sigma_N$  are known. In practice, it needs to be specified how  $\sigma_N$  should be estimated and  $L$  chosen. While the previous discussion gives some guidance on how to estimate  $\sigma_N$ , the most important choice in implementing our method is the choice of the tuning parameter  $L$ , which, without additional assumption, can not be determined from the data without undermining the honesty of (1.6) (cf. [Armstrong and Kolesár \(2018\)](#) and references therein). This is due to a result in [Low \(1997\)](#), who shows that when  $\mathcal{F}$  is a derivative smoothness class, it is, without further assumptions, not possible to adapt to  $\mathcal{F}$  while maintaining uniform coverage at the same time.

**1.3.7.1. Choice of  $L$ .** As a consequence of Low’s impossibility result, the curvature bound  $L$  has to be chosen a priori and application-specific knowledge on what constitutes plausible amounts of curvature is required to obtain suitable values of  $L$ . We reiterate that this requirement is not unique to bias-aware approaches to inference, as confidence intervals based on pointwise procedures must implicitly restrict  $L$  to be informative at any given tolerance level.

In the absence of reliable information on the magnitude of  $L$ , it is recommended to conduct a sensitivity analysis by considering a range of plausible bounds together with rule of thumb (ROT) estimates of  $L$  based on modelling the CEF over the largest part of its domain that is plausibly informative. In our implementation, we consider three rules of thumb. The first was suggested by [Armstrong and Kolesár \(2020\)](#) and is based on fitting a global quartic polynomial on each side of the threshold. The ROT estimate of  $L$  is then obtained by computing the global maximum of the absolute value of the second derivative of the polynomial implied by

the estimated coefficients. The second rule of thumb we consider was proposed in [Imbens and Wager \(2019\)](#). It involves fitting a quadratic polynomial on each side of the cutoff and computing an estimate of  $L$  by scaling the maximum second derivative magnitude by a factor of 2-4. Finally, we propose a third rule of thumb that is based on fitting a cubic smoothing spline with evenly spaced knots on each side of the threshold. The bound  $\hat{L}_{ROT}$  is then estimated by the maximum magnitude of the implied second derivative. This approach is heuristically motivated by the fact that the worst-case function of our estimator is a quadratic spline. In our simulation study, we report results based on this approach.

While it is not possible to consistently recover  $L$  from the data, we are aware of two methods that were proposed to guard against overly optimistic choices of  $L$  and to gain intuition for what might constitute plausible degrees of curvature. The first method is due to [Kolesár and Rothe \(2018\)](#), who propose a method to estimate a lower bound on  $L$  based on the observations that any function in  $\mathcal{F}(L)$  can, between any two points that are  $\Delta$  units apart, not depart from a straight line by more than  $\frac{L\Delta^2}{8}$ . The second method is due to [Noack and Rothe \(2019\)](#), who propose a graphical procedure based on the solution to a constrained least squares problem to visualize "extreme" elements of  $\mathcal{F}(L)$ . The idea is to plot this element while iteratively increasing the curvature bound until the resulting function becomes implausibly erratic. We generally recommend to combine subject knowledge, ROT estimates and such heuristic devices to gain intuition in any given application.

**1.3.7.2. Estimation of  $\sigma_N$ .** The discussion in the previous section implies that the estimator derived under known variances underlying the optimized linear confidence intervals can be understood as motivated by a homoskedastic model. In order to ensure that the inference based on (1.6) is robust to heteroskedasticity, it is required to construct confidence intervals using an appropriate estimator for the conditional variances  $\sigma_N^2$ , analogous to a regression analysis that uses ordinary least squares estimators but builds confidence intervals using Eicker–Huber–White standard errors. In our implementation, we initialize  $\sigma_N$  by a naive homoskedastic estimate to obtain the weights, before building confidence intervals based on a function that implements different heteroskedasticity-robust estimators of the conditional variance, including estimators based on standard estimates of  $\sigma_i$  relying on the residuals of linear regressions, nearest-neighbor estimates proposed and considered in [Abadie and Imbens \(2006\)](#) and [Abadie et al. \(2014\)](#), as well as the uniformly consistent modification proposed in [Noack and Rothe \(2019\)](#). The results reported in the simulation study are obtained using the nearest-neighbor approach of [Abadie et al. \(2014\)](#) to estimate  $\sigma_N$  based on 10 nearest-neighbor matches.

## 1.4 Comparison with other Methods

In this section, we compare the performance of optimized linear confidence intervals to a variety of procedures based on local linear regression in a simulation study. In order to be precise about the comparison we briefly introduce the considered methods.

**1.4.1 Local Linear Methods.** The local linear estimate  $\hat{\theta}_{LL}$  of  $\theta$  with bandwidth  $h$  is the coefficient on  $X_i D_i$  in a weighted OLS regression of  $Y_i$  on the vector  $(1, X_i, D_i, X_i D_i)$ , using only observations  $i$  such that  $|X_i| \leq h$ , with weights determined by a kernel function. Under regularity conditions, it holds that if the density of the assignment variable  $f_X(x)$  is continuous and bounded away from zero in an open neighborhood around the threshold, the MSE of the local linear estimator with bandwidth sequence  $h_n \rightarrow 0$  evaluated at a function  $\mu$  is

$$\text{MSE}(h_n; \mu) = h_n [B^2 + o_p(1)] + \frac{1}{nh_n^3} [V + o_p(1)],$$

where  $B \propto (\mu_+^2(0) - \mu_-^2(0))$  is the leading asymptotic bias and  $V \propto (\sigma_+(0)^2 + \sigma_-(0)^2) f_X(0)^{-1}$  is the asymptotic variance. If  $B \neq 0$ , an asymptotic MSE-optimal bandwidth sequence is thus

$$h_{PMSE} = n^{-1/5} \left[ \frac{3V}{2B^2} \right]^{1/5}.$$

In our simulation study, we compare the performance of uniform procedures to methods that rely on plug-in estimates of this quantity which add a regularization term to the denominator that shrinks with the sample size (Imbens and Kalyanaraman, 2012). We denote such estimates by  $\hat{h}_{PMSE}$  and compute them using the plug-in estimators proposed by Calonico et al. (2014). The undersmoothing bandwidths are computed relative to the obtained pointwise asymptotic MSE-optimal estimate  $\hat{h}_{US} = n^{-1/20} \hat{h}_{PMSE}$ . The two bandwidth choices required for the RBC intervals are either both set pointwise optimal  $b = \hat{b}_{PMSE}$ ,  $h = \hat{h}_{PMSE}$ , where  $\hat{b}_{PMSE}$  refers to the pointwise asymptotic MSE-optimal plug-in estimate for the local quadratic bias estimator, or both set to the pointwise asymptotic MSE-optimal estimate  $\hat{h}_{PMSE}$  for the local linear estimator. In addition, we consider RBC bandwidth choices  $h_{CE}$ ,  $b_{CE}$  that optimize the pointwise asymptotic coverage error (Calonico et al., 2018), which can be considered an intermediate form of undersmoothing and robust bias correction.

Under regularity conditions and restrictions on the rate of the bandwidth sequence  $h_n \rightarrow 0$

$$\sqrt{nh_n^3} [\hat{\theta}_{LL}(h_n) - \theta - h_n B_n] \xrightarrow{d} N(0, V),$$



where  $B_n \xrightarrow{P} B$ . The methods that we consider differ in whether and how they take into account the incorrect centering induced by the smoothing bias. In our simulation, we consider the following four types of two-sided local linear intervals for the above bandwidth choices:

$$\begin{aligned}
\textit{Conventional} & \quad \mathcal{I}_{Conv.} = \left[ \hat{\theta}_{LL}(\hat{h}_{PMSE}) \pm z_{1-\alpha/2} \sqrt{\hat{V}_{Conv.}/n\hat{h}_{PMSE}^3} \right], \\
\textit{Undersmoothed} & \quad \mathcal{I}_{US} = \left[ \hat{\theta}_{LL}(\hat{h}_{US}) \pm z_{1-\alpha/2} \sqrt{\hat{V}_{US}/n\hat{h}_{US}^3} \right], \\
\textit{Robust Bias Correction} & \quad \mathcal{I}_{RBC} = \left[ \hat{\theta}_{LL}(\hat{h}_{RBC}) - \hat{h}_{RBC} \hat{B}_n(\hat{b}_{RBC}) \pm z_{1-\alpha/2} \sqrt{\hat{V}_{RBC}/n\hat{h}_{RBC}^3} \right], \\
\textit{Fixed Length} & \quad \mathcal{I}_{FL} = \left[ \hat{\theta}_{LL}(\hat{h}_{FL}) \pm cv_{1-\alpha}(\bar{t}_n) \sqrt{\hat{V}_{FL}/n\hat{h}_{FL}^3} \right],
\end{aligned}$$

where  $z_{1-\alpha/2}$  denotes the  $(1 - \alpha/2)$  quantile of the standard normal distribution,  $\hat{V}$  denotes an estimate of the respective asymptotic variance, and  $cv_{1-\alpha}$  as well as  $t_n$  are defined as in (1.5) for the worst-case magnitude of the conditional bias of  $\hat{\theta}_{LL}(\hat{h}_{FL})$ .

The first two types, conventional and undersmoothed confidence intervals, essentially assume the smoothing bias away. While the conventional method directly assumes that  $h_n B_n s_n^{-1} \approx 0$ , undersmoothed intervals rely on the asymptotic promise that  $h_{US}/h_{PMSE} \xrightarrow[n \rightarrow \infty]{} 0$ , implying that  $\sqrt{nh_n^3} [\hat{\theta}(h_n) - \theta - h_n B_n] = \sqrt{nh_n^3} [\hat{\theta}(h_n) - \theta] + o_p(1) \xrightarrow{d} N(0, V)$ , which is uninformative about the smoothing bias in a given sample.

The last two types explicitly address the smoothing bias. RBC intervals are centered around a bias-corrected point estimate, using a higher order local polynomial estimator to estimate the bias. They differ from traditional bias-corrected intervals, which are known to perform poorly in finite samples (Hall, 1992), in that they do not require  $h_{RBC}/b_{RBC} \xrightarrow[n \rightarrow \infty]{} 0$ . As a consequence, the standardized bias-correction term is not negligible asymptotically, leading to a different asymptotic variance  $V_{RBC}$  that captures the additional uncertainty introduced by bias estimation. In our simulation, all RBC bias estimates are based on local quadratic estimators. The fixed length intervals are constructed in the same spirit as the optimized intervals discussed in Section 1.3 with weights defined by the local linear estimator. They take into account the exact magnitude of the worst-case conditional bias by inflating the critical value according to the ratio  $\bar{t}_n$ , and are therefore valid for any bandwidth choice. The bandwidths for the fixed length local linear intervals are obtained analogously to  $\kappa_{UMSE}$  and  $\kappa_{LE}$  by minimizing the finite sample uniform MSE  $h_{UMSE}$  or the interval half-length  $h_{HL}$ . The variances for interval construction are estimated by the conditional variances of the estimators implied by the weights, using the nearest-neighbor approach of Abadie et al. (2014) based on 10 nearest neighbor matches.

**1.4.2 Monte Carlo Setup.** In order to investigate the performance of the optimized linear confidence intervals relative to the methods introduced above we conduct a simulation study. The setups differ in the conditional mean function, its degree of curvature and the distribution of the assignment variable, yielding a total of 8 different settings. The assignment variable is drawn from an equidistant uniform distribution with support  $\{-1, -1 + \frac{2}{K}, \dots, 1 - \frac{2}{K}, 1\}$ , where the parameter  $K$  controls the number of support points and  $K_\infty$  means the continuous uniform distribution with support  $[-1, 1]$ . The outcome data is generated according to

$$Y_i = \mu_j(X_i) + \varepsilon_i \quad j \in \{1, 2\},$$

where the CEF error  $\varepsilon_i$  is drawn from a mean zero normal distribution with  $\sigma = 0.1$  and

$$\begin{aligned} \mu_1(x) &= D(x)\theta x + \frac{L}{2} \left[ -x^2 + 1.75s_+^2(|x| - 0.15) - 1.25s_+^2(|x| - 0.4) \right] \\ \mu_2(x) &= D(x)\theta x + \frac{L}{2} \left[ (x+1)^2 - 2s_+^2(x+0.2) + 2s_+^2(x-0.2) - 2s_+^2(x-0.4) + 2s_+^2(x-0.6) - 0.92 \right], \end{aligned}$$

where  $s_+^2(x) = D(x)x^2$  denotes square of the plus function. Note that both CEFs are second order splines with maximal second order derivative magnitude  $L$  and thus elements of  $\mathcal{F}(L)$ . The function  $\mu_1$  attains the second order bound only in  $(-0.15, 0.15)$  while  $\mu_2$  attains the bound everywhere on  $[-1, 1]$  with alternating signs in each interval defined by the knots. In the simulation, we set  $\theta = -0.5$  and consider bounds  $L \in \{2, 6\}$ . Figure 1.1 displays the shape of the functions.

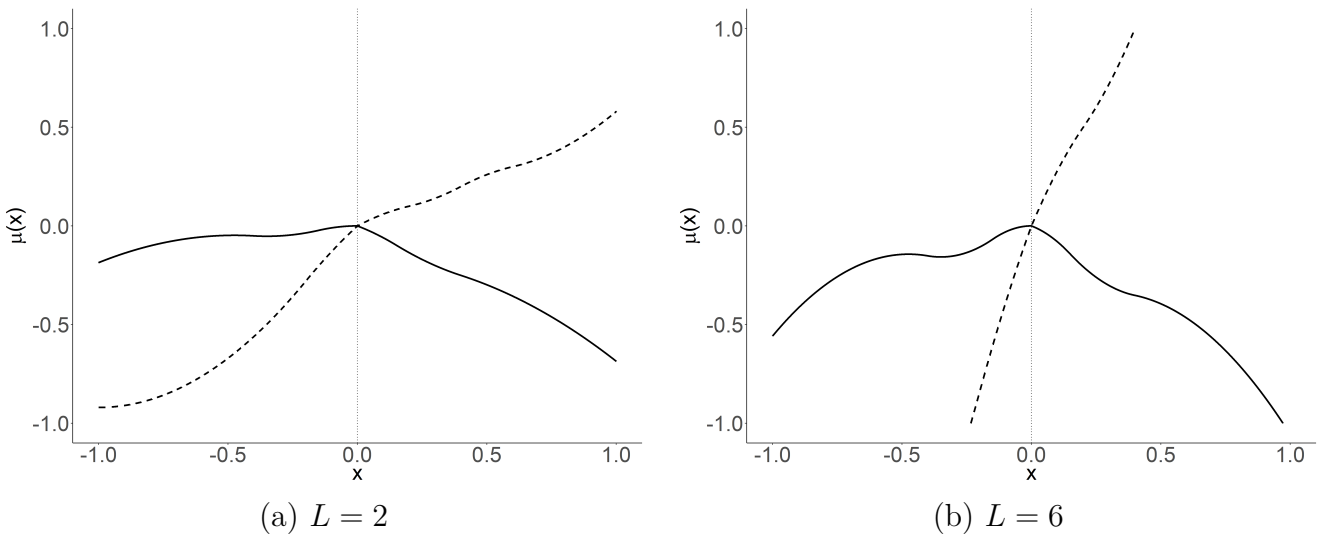


Figure 1.1: Shape of  $\mu_1$  (solid) and  $\mu_2$  (dashed) for  $\theta = -0.5$  and  $L \in \{2, 6\}$  on  $[-1, 1]$ .

**1.4.3 Monte Carlo Results.** Tables 1.1 and 1.2 show the results of 5000 Monte Carlo runs<sup>4</sup> with sample size  $n = 2000$  for  $\mu_1$  and  $\mu_2$  respectively. The top panel in each table displays the results for the case when the assignment variable is drawn from the continuous uniform distribution, while the bottom panel shows the results for 80 equidistant support points. We use a triangular kernel for all local linear methods. In the case that a bandwidth selector chooses a bandwidth for which the respective estimator is not defined, we manually adjust the bandwidth such that it covers three support points on either side of the cutoff<sup>5</sup>. The left panel in each table reports the results for the low curvature version of the CEF, while the right panel reports the results for high curvature. Columns 1 and 2 indicate the method and the tuning target. The curvature bound  $L$  is either chosen by the rule of thumb  $\hat{L}$  or fixed to 2 or 6, as indicated in the tuning subscript. We report the empirical coverage rate at nominal level 95%, the average length relative to the optimal linear interval with correct curvature bound, as well as the average tuning parameter choice of each method.

Unsurprisingly, conventional and undersmoothing confidence intervals show below nominal coverage in all designs, with undercoverage of undersmoothed intervals becoming more severe in the high curvature regime. The performance of robust bias-corrected intervals varies with the tuning target. While both, the default RBC method<sup>6</sup> that picks both bandwidths using the respective asymptotic MSE-optimal estimate and the coverage error optimized RBC interval undercover severely, the RBC interval obtained by setting both bandwidths to the local linear pointwise MSE bandwidth or the UMSE bandwidth under the ROT estimate of  $L$  show close to nominal coverage and are insensitive to the true curvature. However, for the latter this comes at a cost in terms of their length relative to fixed-length and optimized intervals. In all designs, RBC intervals with tuning that attains close to nominal coverage are at least approximately twice as long as the infeasible length-optimal interval and 40% longer than feasible length optimized intervals under the ROT choice of  $L$ . Fixed-length and optimized intervals show above or close to nominal coverage under the correct curvature, with the lowest empirical coverage at 94.7% ( $\mu_{1,2}, K = 80, L = 6$ ). As one would expect, they are conservative when the true curvature is lower than specified and undercover when the true curvature is higher than specified. The rule of thumb choice of  $L$  tends to overestimate the curvature with the exception of  $(\mu_1, K_\infty, L = 6)$ , leading to above nominal coverage of both uniform interval types in most

---

<sup>4</sup>In Appendix 1.I, we provide analogous results for 20,000 Monte Carlo runs that were conducted on a cluster due to the required computational resources. The results reported in Tables 1.1 and 1.2 can be replicated on a desktop computer within a reasonable timeframe using the replication files. See Appendix 1.H for details.

<sup>5</sup>This occurred only in the discrete setting.

<sup>6</sup>This refers to the default in the authors' software implementation of RBC CIs.

Table 1.1: Monte Carlo Results: Coverage and Relative Length I

$\mu_1(x)$		$L = 2$			$L = 6$		
Method	Tuning	Cov.	RL	$h/\kappa$	Cov.	RL	$h/\kappa$
<i>Cont. Design</i>							
Conv.	$h_{PMSE}$	43.0	0.566	0.226	18.3	0.398	0.183
US	$h_{US}$	87.1	0.999	0.154	64.2	0.702	0.125
RBC	$b_{PMSE}, h_{PMSE}$	53.9	0.849	0.226 (0.462)	38.1	0.582	0.183 (0.388)
RBC	$h = b = h_{PMSE}$	95.2	2.140	0.226	95.1	1.510	0.183
RBC	$b_{CE}, h_{CE}$	84.4	1.160	0.154 (0.462)	68.9	0.807	0.125 (0.388)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.5	4.390	0.167	94.7	3.940	0.104
LL-FL	$h_{UMSE, L=\hat{L}}$	95.0	1.710	0.167	95.2	1.530	0.104
LL-FL	$h_{HL, L=\hat{L}}$	96.8	1.610	0.221	93.1	1.440	0.138
Opt	$\kappa_{UMSE, L=\hat{L}}$	94.9	1.700	1.000	95.1	1.520	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	96.7	1.600	0.246	93.0	1.430	0.246
LL-FL	$h_{UMSE, L=2}$	95.1	1.060	0.187	21.1	0.550	0.187
LL-FL	$h_{HL, L=2}$	97.0	1.000	0.247	0.42	0.517	0.247
Opt	$\kappa_{UMSE, L=2}$	95.1	1.060	1.000	21.0	0.549	1.000
Opt	$\kappa_{HL, L=2}$	97.0	1.000	0.246	0.36	0.517	0.246
LL-FL	$h_{UMSE, L=6}$	99.3	2.060	0.121	95.0	1.060	0.121
LL-FL	$h_{HL, L=6}$	100.0	1.940	0.160	95.0	1.000	0.160
Opt	$\kappa_{UMSE, L=6}$	99.4	2.050	1.000	95.1	1.060	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.930	0.245	95.0	1.000	0.245
<i>Disc. Design</i>							
Conv.	$h_{PMSE}$	44.8	0.578	0.225	17.8	0.403	0.183
US	$h_{US}$	87.8	1.040	0.154	65.3	0.723	0.126
RBC	$b_{PMSE}, h_{PMSE}$	54.4	0.869	0.225 (0.463)	37.2	0.593	0.183 (0.389)
RBC	$h = b = h_{PMSE}$	95.4	2.320	0.225	95.2	1.670	0.183
RBC	$b_{CE}, h_{CE}$	84.8	1.200	0.154 (0.463)	69.6	0.832	0.126 (0.389)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.0	4.550	0.170	94.3	4.380	0.108
LL-FL	$h_{UMSE, L=\hat{L}}$	95.4	1.900	0.170	95.4	1.760	0.108
LL-FL	$h_{HL, L=\hat{L}}$	97.3	1.810	0.224	93.4	1.670	0.142
Opt	$\kappa_{UMSE, L=\hat{L}}$	95.4	1.860	1.000	95.0	1.710	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	97.2	1.760	0.246	93.2	1.620	0.243
LL-FL	$h_{UMSE, L=2}$	95.0	1.070	0.189	19.7	0.548	0.189
LL-FL	$h_{HL, L=2}$	97.1	1.000	0.251	0.30	0.515	0.251
Opt	$\kappa_{UMSE, L=2}$	95.4	1.060	1.000	20.9	0.546	1.000
Opt	$\kappa_{HL, L=2}$	97.2	1.000	0.242	0.32	0.513	0.242
LL-FL	$h_{UMSE, L=6}$	99.5	2.100	0.125	95.0	1.080	0.125
LL-FL	$h_{HL, L=6}$	100.0	1.960	0.163	94.9	1.010	0.163
Opt	$\kappa_{UMSE, L=6}$	99.4	2.080	1.000	94.9	1.070	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.950	0.237	94.7	1.000	0.237

*Note:* Empirical coverage rate (Cov.), average length relative to optimized interval with true smoothness bound (RL) and tuning parameter choice ( $h/\kappa$ ) of conventional (Conv.), undersmoothed (US), robust bias-corrected (RBC), fixed length (LL-FL) and optimized linear (Opt) 95% confidence intervals over 5.000 Monte Carlo draws. The pilot RBC bandwidth is reported in parentheses if it is selected separately from  $h$ .

Table 1.2: Monte Carlo Results: Coverage and Relative Length II

$\mu_2(x)$		$L = 2$			$L = 6$		
Method	Tuning	Cov.	RL	$h/\kappa$	Cov.	RL	$h/\kappa$
<i>Cont. Design</i>							
Conv.	$h_{PMSE}$	42.7	0.599	0.221	55.4	0.593	0.143
US	$h_{US}$	84.6	1.060	0.151	79.7	1.050	0.098
RBC	$b_{PMSE}, h_{PMSE}$	62.2	0.908	0.221 (0.439)	82.8	0.778	0.143 (0.336)
RBC	$h = b = h_{PMSE}$	95.4	2.260	0.221	93.9	2.250	0.143
RBC	$b_{CE}, h_{CE}$	87.0	1.230	0.151 (0.439)	86.8	1.150	0.098 (0.336)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.1	5.510	0.134	94.3	4.870	0.088
LL-FL	$h_{UMSE, L=\hat{L}}$	98.1	2.140	0.134	98.7	1.890	0.088
LL-FL	$h_{HL, L=\hat{L}}$	99.3	2.010	0.178	99.6	1.780	0.116
Opt	$\kappa_{UMSE, L=\hat{L}}$	98.1	2.130	1.000	98.7	1.880	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	99.3	2.010	0.246	99.6	1.770	0.246
LL-FL	$h_{UMSE, L=2}$	95.0	1.060	0.187	20.6	0.550	0.187
LL-FL	$h_{HL, L=2}$	95.4	1.000	0.247	0.02	0.517	0.247
Opt	$\kappa_{UMSE, L=2}$	95.1	1.060	1.000	20.5	0.549	1.000
Opt	$\kappa_{HL, L=2}$	95.3	1.000	0.246	0.02	0.517	0.246
LL-FL	$h_{UMSE, L=6}$	99.3	2.060	0.121	95.0	1.060	0.121
LL-FL	$h_{HL, L=6}$	100.0	1.940	0.160	95.0	1.000	0.160
Opt	$\kappa_{UMSE, L=6}$	99.4	2.050	1.000	95.1	1.060	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.930	0.245	95.0	1.000	0.245
<i>Disc. Design</i>							
Conv.	$h_{PMSE}$	45.0	0.621	0.218	54.2	0.607	0.143
US	$h_{US}$	84.4	1.120	0.149	80.2	1.000	0.106
RBC	$b_{PMSE}, h_{PMSE}$	62.1	0.939	0.218 (0.436)	83.2	0.801	0.143 (0.334)
RBC	$h = b = h_{PMSE}$	95.5	2.520	0.218	94.1	2.680	0.143
RBC	$b_{CE}, h_{CE}$	85.8	1.300	0.149 (0.436)	87.5	1.110	0.106 (0.334)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.0	5.990	0.137	94.1	5.730	0.092
LL-FL	$h_{UMSE, L=\hat{L}}$	98.3	2.490	0.137	98.9	2.060	0.092
LL-FL	$h_{HL, L=\hat{L}}$	99.5	2.370	0.181	99.8	1.940	0.121
Opt	$\kappa_{UMSE, L=\hat{L}}$	98.4	2.420	1.000	98.8	2.000	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	99.4	2.300	0.246	99.8	1.890	0.239
LL-FL	$h_{UMSE, L=2}$	95.0	1.070	0.189	18.9	0.548	0.189
LL-FL	$h_{HL, L=2}$	95.6	1.000	0.251	0.04	0.515	0.251
Opt	$\kappa_{UMSE, L=2}$	95.2	1.060	1.000	20.3	0.546	1.000
Opt	$\kappa_{HL, L=2}$	95.3	1.000	0.242	0.06	0.513	0.242
LL-FL	$h_{UMSE, L=6}$	99.5	2.100	0.125	95.0	1.080	0.125
LL-FL	$h_{HL, L=6}$	100.0	1.960	0.163	94.9	1.010	0.163
Opt	$\kappa_{UMSE, L=6}$	99.4	2.080	1.000	94.9	1.070	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.950	0.237	94.7	1.000	0.237

*Note:* Empirical coverage rate (Cov.), average length relative to optimized interval with true smoothness bound (RL) and tuning parameter choice ( $h/\kappa$ ) of conventional (Conv.), undersmoothed (US), robust bias-corrected (RBC), fixed length (LL-FL) and optimized linear (Opt) 95% confidence intervals over 5.000 Monte Carlo draws. The pilot RBC bandwidth is reported in parentheses if it is selected separately from  $h$ .

interval types in most designs with lowest empirical coverage at 93.0%. The resulting intervals thus tend to be conservative but are still substantially shorter than their pointwise counterparts under valid tuning. Interestingly, the length-optimal bias-aware intervals show higher coverage rates than their uniform MSE-optimal counterparts on average. The length-optimal weights of both, fixed length and optimized estimators "oversmooth" relative to the uniform MSE-optimal choice of the respective tuning parameter. The optimally tuned fixed-length intervals demonstrate performance on par with their optimized counterparts, indicating that the high minimax efficiency of local linear estimators under second order bounds demonstrated in [Armstrong and Kolesár \(2020\)](#) for estimating the value of the CEF at a point also holds for the estimation of first derivatives.

**1.4.4 Gains from Optimization.** The popularity of local linear estimators in empirical practice is motivated by their intuitive appeal and a range of attractive theoretical properties of local polynomials, in particular their asymptotic minimax efficiency over the Taylor class of functions ([Fan, 1993](#), [Fan et al., 1997](#), [Cheng et al., 1997](#)). [Armstrong and Kolesár \(2020\)](#) show that local polynomial estimators can also attain high minimax efficiency in the Hölder class of functions defined by derivative bounds among a large class of estimators to which a central limit theorem applies and that have worst-case bias and standard deviation that scale as powers of a bandwidth parameter. This result relies on their observation that, for relevant performance criteria, the asymptotic minimax performance of two estimators in this class does not depend on the criterion but is solely governed by their worst-case biases, their standard deviations and their rate exponents  $r = \gamma_b/(\gamma_b - \gamma_s)$ , where  $\gamma_b$  and  $\gamma_s$  denote the scaling exponents of the worst-case bias and the standard deviation respectively. Moreover, they show that the optimal worst-case bias to standard deviation ratio depends only on the criterion and  $r$  (cf. Theorem 2.1 in [Armstrong and Kolesár, 2020](#)).

Their analytic results provide us with guidance on what to expect with respect to the asymptotic efficiency gains of optimized linear confidence intervals and allow us to state a lower bound for the efficiency gain of our method relative to uniform MSE-optimal fixed length intervals. For the local linear estimator of the kink parameter,  $r = 0.4$  and their calculations imply an efficiency gain of approximately 6% at  $\alpha = 0.05$  for moving from the uniform MSE-optimal to the length-optimal local linear interval (cf. Figure 3 in [Armstrong and Kolesár, 2020](#)), which is consistent with our Monte-Carlo results. Thus, a lower bound for the efficiency gain of the optimized interval relative to the UMSE-optimal fixed length interval is 6%. This

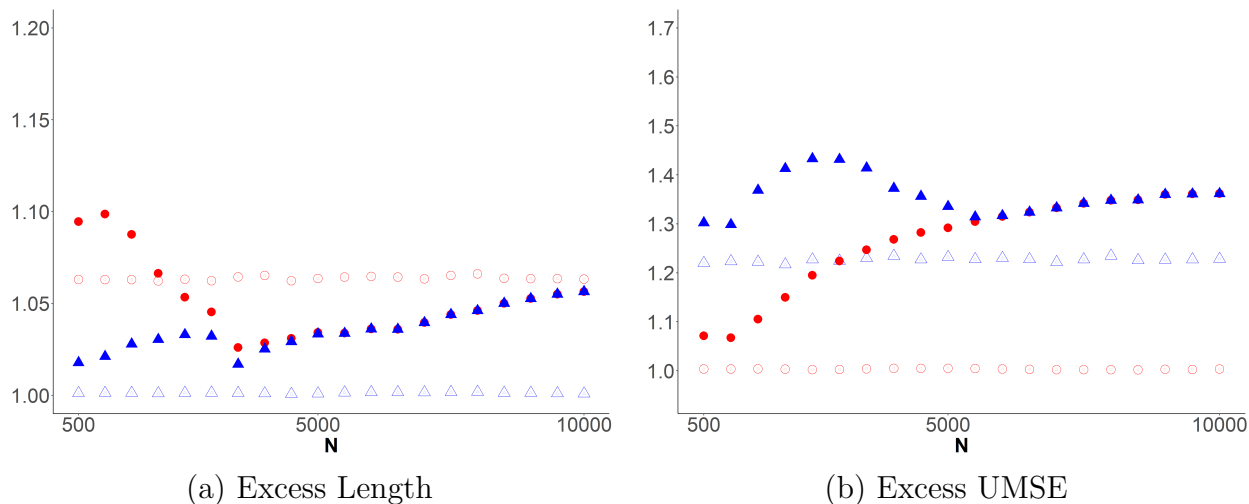


Figure 1.2: Excess length of a 95% confidence interval (a) and excess UMSE (b) of the length-optimal (triangle) and UMSE-optimal (circle) local linear fixed-length interval and estimator relative to the respective optimized linear interval/estimator for continuous (blank) and discrete (filled) designs across different sample sizes.

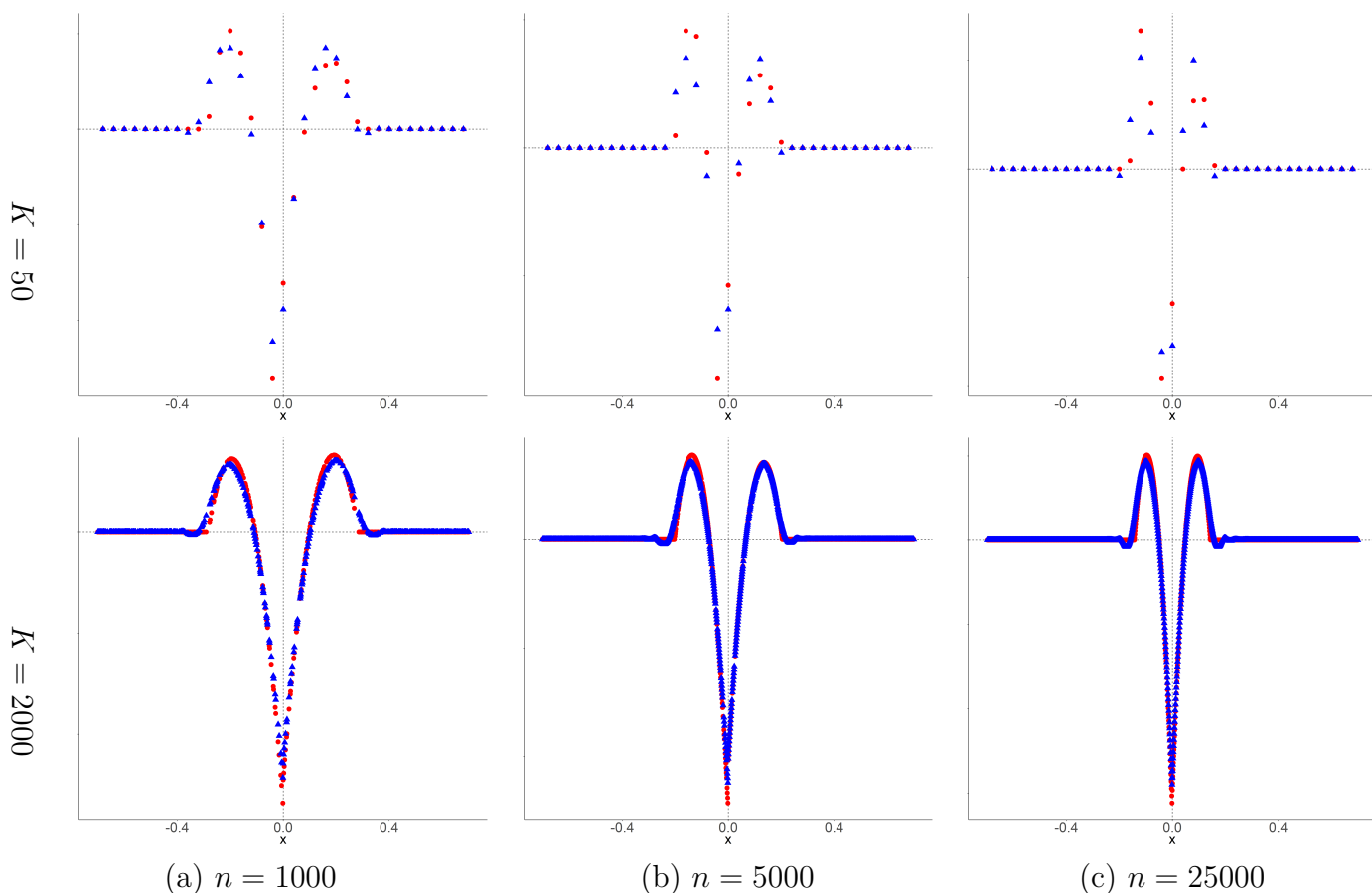


Figure 1.3: Uniform MSE-optimal linear (triangle) and fixed-length local linear (circle) weights for a coarse ( $K = 50$ ) and a dense ( $K = 2000$ ) uniform design on  $[0, 1]$  for different sample sizes,  $L = 2$  and  $\sigma = 0.2$ . The plotted weights are scaled by  $n^{4/5}$  to facilitate comparison across sample sizes.

implies a larger lower bound for the efficiency gain relative to undersmoothed and RBC intervals based on valid bandwidth choices. Figure 1.2 shows the relative risk of fixed length length-optimal and uniform MSE-optimal estimators in terms of interval length (a) and uniform MSE (b) relative to the optimized linear counterparts for the respective criterion for the continuous and discrete ( $K = 40$ ) uniform design for different sample sizes. It illustrates that the efficiency gains from optimization increase as the discreteness of the assignment variable becomes more severe<sup>7</sup>. This is because the shape of the optimized weighting function increasingly deviates from the local linear weights as the discreteness increases. This can be seen in Figure 1.3, which depicts uniform MSE-optimal local linear and optimized weights for different degrees of discreteness. In continuous designs fixed-length and the optimal linear kernels are nearly identical and the only advantage of the optimization based approach is that it can be easily modified to sharpen inference via shape constraints as discussed in Appendix 1.F.

## 1.5 Empirical Illustration

We apply our method to the data of Landais (2015), who estimates the effect of unemployment benefits on the duration of unemployment in a regression kink design. The paper exploits kinks in the schedule of unemployment benefits arising from a hard cap at a maximum benefit amount  $b_{max}$ . In the US, the weekly benefit amount  $b$  received by an eligible unemployed is a fixed fraction  $\gamma$  of a function of previous quarterly earnings  $hqw$  in a base period up to the cap.

$$b = \begin{cases} \gamma hqw, & \text{if } \gamma hqw \leq b_{max} \\ b_{max}, & \text{if } \gamma hqw > b_{max} \end{cases}$$

Landais (2015) reports estimates of  $\tau_{RKD}$  for five US states: Louisiana, Idaho, Missouri, New Mexico and Washington. For the sake of exposition, we focus on the results for Louisiana, which serves as the leading example in the paper. Figure 1.4 displays the benefit schedule for Louisiana for the time period covered by the data. Due to adjustments, the maximum benefit level changed over time, resulting in five distinct kinks. For the period under consideration the weekly benefit rate was fixed at  $\gamma = 0.04$ , which corresponds to a constant replacement rate of 52% up to the respective kink, from where onwards the replacement rate decreases. The paper utilizes data from the Continuous Wage and Benefit History (CWBH), a publicly available administrative UI data set for the US that contains the universe of unemployment spells and

---

<sup>7</sup>The convergence of the risk observed in Figure 1.2 is due to manual bandwidth adjustments to ensure that the estimators are well defined.



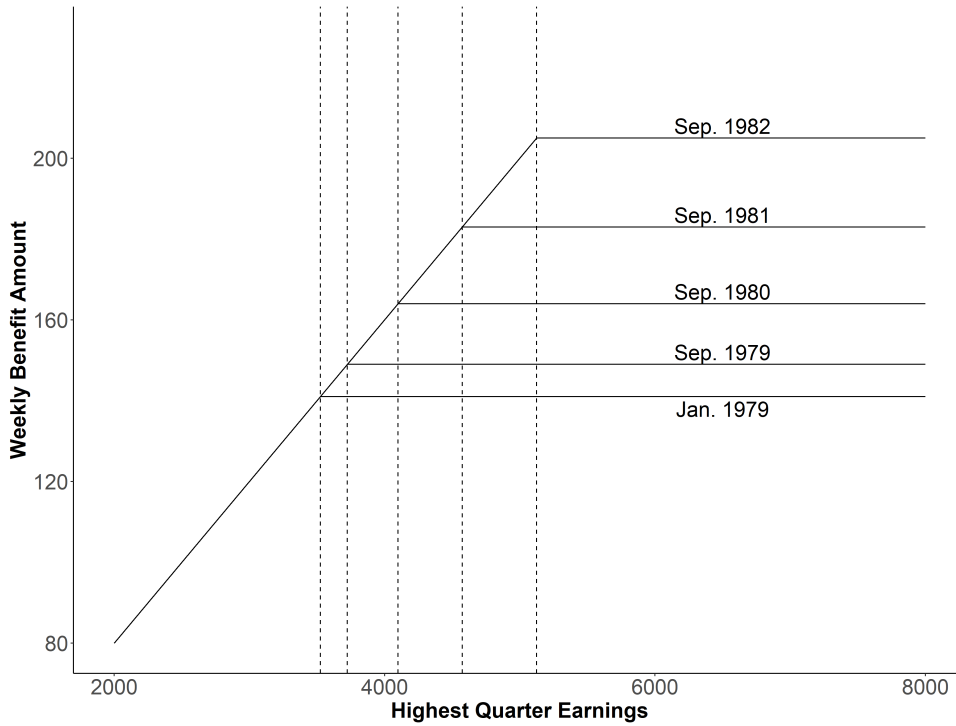


Figure 1.4: Weekly UI Benefit Schedule, Louisiana 1979-1984 (Landais, 2015).

wage records for the five US states from the late 1970s to 1984, with different states starting the recording at different points in time. For Louisiana, the dataset contains  $n = 44702$  unemployment spells for the whole time period. See Landais (2015) Section II.A for a detailed description of the data. Figure 1.5 plots the pooled data for all five time periods, where we have normalized the assignment variable (highest quarterly earnings) by the location of the respective kink and included only data within an 85% interval of the kink  $[0.15, 1.85]$ . This corresponds to bandwidths in the range of 3000-4300 USD. In order to reduce noise, it shows the average unemployment durations, measured in weeks, in 30 equally wide bins. The paper reports RKD estimates of the effect of the benefit level on unemployment duration separately for each time period, with point estimates and standard errors rescaled to the 2010 USD price level. In the main specification, Landais (2015) estimates  $\tau_{RKD}$  using local linear regression with a fixed bandwidth  $h = 2500$ <sup>8</sup> and reports conventional 95% confidence intervals for  $\tau_{RKD}$  based on Eicker-Huber-White standard errors. Table 1.3 replicates<sup>9</sup> the results reported in Table 2 of the paper and additionally presents robust bias-corrected ( $h = b = \hat{h}_{PMSE}$ ) and optimized linear ( $L = \hat{L}_{ROT}$ ) point estimates and 95% confidence intervals computed on the same data. The point estimates reported in Table 1.3 correspond to the estimated effects of a

<sup>8</sup>In the paper's online Appendix, the author reports robustness checks for the pooled sample consisting of the last two periods using bandwidths 1500 and 4500.

<sup>9</sup>The point estimates in the top-left panel for period 3 and 4 deviate by 0.001 from the results reported in the paper. We attribute this difference to rounding.

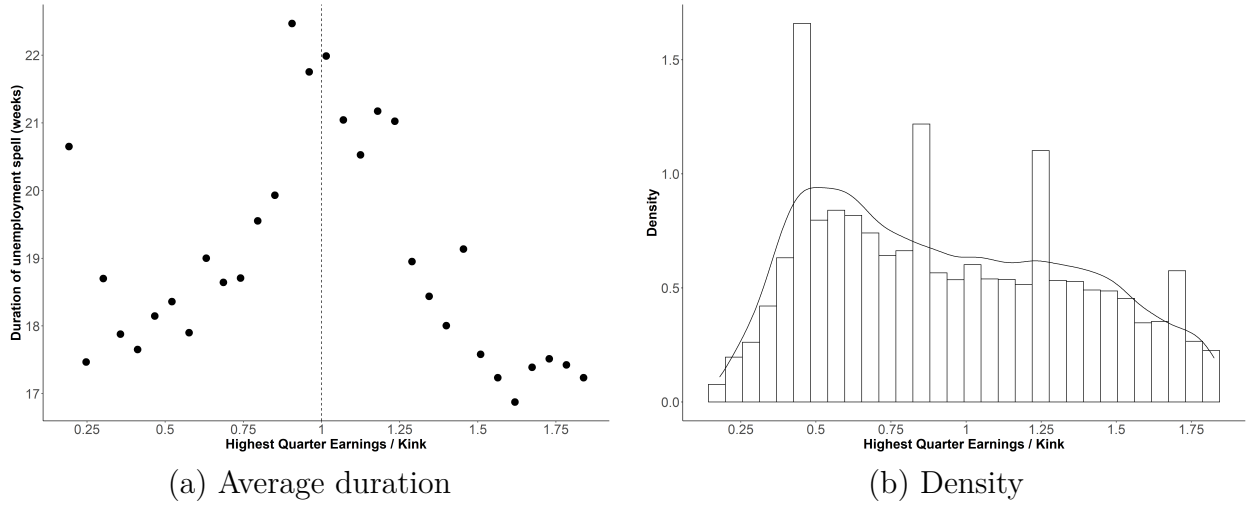


Figure 1.5: Average duration (a) and density (b) in 30 equal width bins over  $[-0.15, 1.85]$ .

1 USD increase in weekly benefits on the duration of paid unemployment in weeks. For example, the point estimate for period 4 reported in Landais (2015) (top-left panel) suggest that a 1 dollar increase in weekly unemployment benefits leads to a 0.043 weeks increase in the duration of unemployment at the kink. The estimates in the top-left panel correspond to (local) elasticities in the range of 0.2 and 0.7, suggesting that a 10% increase in the average weekly

Table 1.3: RKD Estimates: Effect of the Benefit Level on Unemployment Duration

Period	Landais (2015)			RBC		
	Estimate	Interval	$h$	Estimate	Interval	$h$
1	0.006	$[-0.005, 0.018]$	2500	-0.012	$[-0.460, 0.435]$	600
2	0.018	$[0.007, 0.028]$	2500	-0.094	$[-0.291, 0.102]$	1063
3	0.019	$[0.007, 0.030]$	2500	0.058	$[-0.252, 0.368]$	855
4	0.043	$[0.025, 0.061]$	2500	-0.198	$[-0.529, 0.133]$	1024
5	0.047	$[0.035, 0.060]$	2500	0.190	$[0.044, 0.337]$	1410
Period	Optimized			Optimized		
	Estimate	Interval	$\kappa$	Estimate	Interval	$\kappa$
1	0.005	$[-0.006, 0.017]$	0.729	0.006	$[-0.006, 0.017]$	1
2	0.017	$[-0.004, 0.039]$	0.302	0.016	$[-0.016, 0.039]$	1
3	-0.056	$[-0.250, 0.138]$	0.367	-0.041	$[-0.238, 0.156]$	1
4	0.063	$[-0.045, 0.170]$	0.240	0.068	$[-0.046, 0.183]$	1
5	0.059	$[0.010, 0.109]$	0.217	0.072	$[0.020, 0.125]$	1

*Note:* Ad-hoc conventional local linear, robust bias-correction and optimized RKD point estimates and 95% confidence intervals for  $\tau_{RKD}$  in the CWBH data for the periods Jan-Sep 1979 (Period 1), Sep 1979-Sep 1980 (Period 2), Sep 1980-Sep 1981 (Period 3), Sep 1981-Sep 1982 (Period 4) and Sep 1982-Dec 1983 (Period 5).

benefit amount increases unemployment duration by 2 to 7% on average at the kink. As can be seen from the confidence intervals reported in the remaining panels, this finding is sensitive to potential smoothing biases for most of the time periods under consideration, with both RBC and optimized intervals covering zero at the 95% level for periods 1-4. However, while the RBC point estimates differ substantially from those reported in the top-left panel, with confidence intervals that are mostly uninformative for the question at hand, our procedure yields point estimates relatively close to those reported in Landais (2015) and lower bounds of 95% confidence intervals that are marginally below zero for most periods. This is because the estimated curvature is rather low relative to the estimators' standard deviation in all data sets except for period 3, with  $\hat{L}_{ROT} \times 10^5 = (0.0137, 0.058, 1.99, 0.477, 0.145)$ . For the same reason, the difference between length-optimal and UMSE-optimal intervals is rather small in this data set as shown in the lower panel. Overall, these results indicate that the uncertainty associated with the estimated effect of unemployment benefits on unemployment duration is higher than suggested by the top-left panel unless one is certain about the linearity of the CEF in the domain specified by the bandwidth.

As mentioned earlier, the optimization approach to bias-aware RKD inference allows us to readily sharpen inference by imposing shape constraints on the CEF. This is often useful, as the plateaus in the schedules that are typically exploited in regression kink designs often give rise to empirically plausible concavity and convexity restrictions. In Appendices 1.F and 1.D, we discuss this further and explain how shape constraints are implemented in practise. Table 1.4 presents the counterparts of the optimized intervals reported in the lower panel of Table 1.3 under the restriction that  $\mu$  is a concave function, demonstrating that such constraints can substantially shrink the confidence set.

Table 1.4: Optimized Intervals under Shape Constraints: Concavity

Period	$n$	Length-optimal			UMSE-optimal		
		Estimate	Interval	$\kappa$	Estimate	Interval	$\kappa$
1	2493	0.001	[-0.007, 0.009]	1.325	0.001	[-0.007, 0.009]	1
2	4580	0.005	[-0.003, 0.012]	0.990	0.005	[-0.003, 0.012]	1
3	4019	0.002	[-0.005, 0.009]	1.157	0.002	[-0.005, 0.009]	1
4	5577	0.034	[0.025, 0.043]	1.281	0.034	[0.025, 0.043]	1
5	10721	0.023	[0.017, 0.029]	1.214	0.023	[0.017, 0.029]	1

*Note:* Shape constrained (concavity of  $\mu$ ) optimized linear confidence intervals for  $\tau_{RKD}$  in the CWBH data for the periods as in Table 1.3.

## 1.6 Conclusion

Motivated by the finite sample coverage problems of pointwise approaches to nonparametric inference, this paper proposes a robust and efficient alternative method for the construction of nonparametric confidence intervals in regression kink designs. Given a curvature bound, the method is fully data driven, easy to implement, and has excellent finite sample coverage and length properties due to its minimax construction that explicitly takes into account the worst-case smoothing bias in a given data set.

# Appendices to Chapter 1

## 1.A Bias Derivation

The proposition that the weights solving problem (1.3) also solve the program (1.1) relied on two arguments. First, we argued that the weights solving (1.1) must lie in the feasible set of (1.3) by means of simple Taylor expansions. Second, we argued that by the result in (1.2), the worst-case conditional bias of  $\hat{\theta}$  over  $\mathcal{F}(L)$  is proportional to the smoothness bound  $L$  and, by definition of the remainder, the same as the worst-case conditional bias over the normalized class  $\bar{\mathcal{F}}(L)$ . The second argument relied on an explicit formula for the remainder of Taylor expansions near the threshold obtained via integration by parts and the Fubini-Tonelli Theorem.

*Proof.* Since  $\mu \in \mathcal{F}(L)$ ,  $\mu^1$  is Lipschitz by definition and hence absolutely continuous, and has second derivative almost everywhere with  $\sup_{x \in \mathcal{X}} |\mu^2(x)| \leq L$ . By the Fundamental Theorem of Calculus we have that for each  $X_i \in \mathcal{X}_+$  and  $\mu \in \mathcal{F}(L)$

$$\mu_+(X_i) = \mu_+(0) + \int_0^{X_i} \mu_+^1(t) dt.$$

Set  $u(t) = \mu_+^1(t)$  and  $v(t) = X_i - t$ . Integration by parts then yields

$$\begin{aligned} \mu_+(X_i) &= \mu_+(0) - \int_0^{X_i} u(t)v^1(t) dt \\ &= \mu_+(0) - [\mu_+^1(t)(X_i - t)]_0^{X_i} + \int_0^{X_i} \mu_+^2(t)(X_i - t) dt \\ &= \mu_+(0) + \mu_+^1(0)X_i + \int_0^{X_i} \mu_+^2(t)(X_i - t) dt. \end{aligned}$$

Analogously, for each  $X_i \in \mathcal{X}_-$  and  $\mu \in \mathcal{F}(L)$ ,  $\mu_-(X_i) = \mu_-(0) + \mu_-^1(0)X_i - \int_{X_i}^0 \mu_-^2(t)(X_i - t) dt$ .

The integral form of the remainder allows us to write the conditional mean as

$$\begin{aligned} \mathbb{E}[\hat{\theta}|X_N] &= \sum_{i=1}^N w_+ \left[ \mu_+(0) + \mu_+^1(0)X_i + \int_0^{X_i} \mu_+^2(t)(X_i - t) dt \right] \\ &\quad + \sum_{i=1}^N w_- \left[ \mu_-(0) + \mu_-^1(0)X_i - \int_{X_i}^0 \mu_-^2(t)(X_i - t) dt \right], \end{aligned}$$

which we used to confirm the first statement. The constraints

$$\sum_{i=1}^n w_{i,+} = 0 \quad \sum_{i=1}^n w_{i,-} = 0 \quad \sum_{i=1}^n w_{i,+} X_i = 1 \quad \sum_{i=1}^n w_{i,-} X_i = -1$$

are in fact necessary conditions for the worst-case bias over  $\mathcal{F}(L)$  to be finite, as, if they are not all satisfied, we can choose  $(\mu_+, \mu_-, \mu_+^1, \mu_-^1)$  to induce an arbitrary conditional bias. As a consequence, we have that, if the worst-case conditional bias is finite, it is given by

$$\sup_{\mu \in \mathcal{F}(L)} \mathbb{E}[\hat{\theta} - \theta | X_N] = \sup_{\mu \in \mathcal{F}(L)} \left[ \sum_{i=1}^n w_{i,+} \int_0^{X_i} \mu_+^2(t)(X_i - t)dt - \sum_{i=1}^n w_{i,-} \int_{X_i}^0 \mu_-^2(t)(X_i - t)dt \right].$$

In order to derive the result (1.2) we define the two measure spaces  $(\mathbb{R}, \mathcal{L}, \nu_1)$  and  $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \nu_2)$ , where  $\mathcal{L}$  denotes the Lebesgue  $\sigma$ -algebra,  $\nu_1$  the Lebesgue measure,  $\mathcal{P}(\mathbb{N})$  the power set of the natural numbers and  $\nu_2$  the counting measure. Note that the measures are  $\sigma$ -finite on the real numbers and natural numbers respectively. Let  $(\mathbb{N} \times \mathbb{R}, \mathcal{A}, \nu)$  denote the product space, where  $\mathcal{A}$  is the generated product  $\sigma$ -algebra and  $\nu$  is the unique product measure. Moreover, let  $g : \mathbb{N} \times \mathbb{R} \mapsto \mathbb{R}$ ,  $g(i, t) = \tilde{w}(i)\mu^2(t)(\tilde{t}(i) - t)$ , where  $\tilde{w}(i) = w_i \cdot \mathbb{1}_{[1 \leq i \leq n]}$  and  $\tilde{t}(i) = X_i \cdot \mathbb{1}_{[1 \leq i \leq n]}$  and define the sets  $A_+ = \{(i, t) \in \mathbb{N} \times \mathbb{R} \mid 0 \leq t \leq \tilde{t}(i)\}$  and  $A_- = \{(i, t) \in \mathbb{N} \times \mathbb{R} \mid \tilde{t}(i) \leq t < 0\}$ . Note that by definition,  $\sum_{i=1}^n w_{i,+} \int_0^{X_i} \mu_+^2(t)(X_i - t)dt = \int_{\mathbb{N}} \int_0^{\tilde{t}(i)} g(i, t) d\nu_1(t) d\nu_2(i) \equiv T_1$  and  $\sum_{i=1}^n w_{i,-} \int_{X_i}^0 \mu_-^2(t)(X_i - t)dt = \int_{\mathbb{N}} \int_{\tilde{t}(i)}^0 g(i, t) d\nu_1(t) d\nu_2(i) \equiv T_2$ . By the Fubini–Tonelli Theorem it holds<sup>10</sup> that

$$T_1 = \int_{A_+} g(i, t) d\nu = \int_0^\infty \int_{\tilde{t}(i) \geq t} g(i, t) d\nu_2(i) d\nu_1(t) = \int_0^\infty \mu_+^2(t) \sum_{i: X_i \in [t, \infty)} w_{i,+}(X_i - t) dt,$$

$$T_2 = \int_{A_-} g(i, t) d\nu = \int_{-\infty}^0 \int_{\tilde{t}(i) < t} g(i, t) d\nu_2(i) d\nu_1(t) = \int_{-\infty}^0 \mu_-^2(t) \sum_{i: X_i \in (-\infty, t]} w_{i,-}(X_i - t) dt,$$

and it follows that under the constraints,

$$\sup_{\mu \in \mathcal{F}(L)} \mathbb{E}[\hat{\theta} - \theta | X_N] = \sup_{\mu \in \mathcal{F}(L)} \left[ \int_0^\infty \mu_+^2(t) \sum_{i: X_i \in [t, \infty)} w_{i,+}(X_i - t) dt - \int_{-\infty}^0 \mu_-^2(t) \sum_{i: X_i \in (-\infty, t]} w_{i,-}(X_i - t) dt \right].$$

This yields (1.2) and, by the definition of the remainder<sup>11</sup>, an immediate consequence is

$$\bar{B}(w_N) = \sup_{\mu \in \mathcal{F}(L)} \mathbb{E}[\hat{\theta} - \theta | X_N] = L \sup_{\mu \in \mathcal{F}(1)} \mathbb{E}[\hat{\theta} - \theta | X_N] = L \sup_{R \in \bar{\mathcal{F}}(1)} \sum_{i=1}^n w_i R(X_i).$$

with  $\bar{\mathcal{F}}(1) = \{f : f(0) = f^1(0) = 0 \wedge |f^1(x) - f^1(x')| \leq |x - x'|, x, x' \in \mathcal{X}\}$ . This shows the equivalence of problems (1.1) and (1.3) in terms of optimal weights and implies that, by symmetry of  $\mathcal{F}(L)$  with respect to zero,  $r^*L$  is a sharp bound on the magnitude of  $\bar{B}(w_N)$ .

<sup>10</sup>Note that if  $\max_{1 \leq i \leq n} \{|w_i|\} < \infty$  boundedness of  $\mathcal{X}$  and  $\mu^2$  implies  $\int_{A_\pm} |g(i, t)| d\nu < \infty$ .

<sup>11</sup>Note that  $|R^1(x) - R^1(x')| = |\mu^1(x) - \mu^1(x')| \leq |x - x'|$  for any  $x, x' \in \mathcal{X}$  and  $\mu \in \mathcal{F}(L)$ .

*Remark.* Provided that  $\sum_{i=1}^n w_{i,+} X_i^j = -\sum_{i=1}^n w_{i,-} X_i^j = \delta_{v,p}$  for  $j = 1, \dots, p$ , iteratively integrating by parts and applying the same logic yields a general formula for the conditional bias of a linear estimator of  $\theta^v$  as a function of the weights and  $\mu^{p+1}$ :

$$\frac{1}{p!} \left[ \int_0^\infty \mu_+^{p+1}(t) \sum_{i: X_i \in [t, \infty)} w_{i,+} (X_i - t)^p dt - \int_{-\infty}^0 \mu_-^{p+1}(t) \sum_{i: X_i \in (-\infty, t]} w_{i,-} (X_i - t)^p dt \right].$$

## 1.B Constraint Qualification

In Section 1.3.4 we argued that, by strong duality of (1.3), we can recover the solution of (1.3), and thus the weights solving (1.1) as well as the associated worst-case bias over  $\mathcal{F}(L)$ , by solving the dual problem underlying (1.5). Because (1.3) is a strictly convex quadratic problem and our constraints are all linear equalities and inequalities, we can verify strong duality by showing that a refined Slater's condition (cf. Boyd and Vandenberghe, 2004) applies. This constraint qualification states that strong duality holds whenever there exists a feasible point for (1.3), which we demonstrate by showing that the weights of the local polynomial estimator of order  $p \geq 1$  together with  $r$  corresponding to its worst-case conditional bias over  $\mathcal{F}(1)$  satisfy all constraints of (1.3).

*Proof.* Let  $K_h(t) = K(t/h)$  for a bounded kernel function  $K$  with support  $[-1, 1]$  indexed by a bandwidth  $h > 0$  and let  $e_j$  denote the unit column vector of length  $2p + 2$  with a one at position  $j$ . The local polynomial estimator of order  $p \geq v$  for a jump in the  $v$ -th derivative is  $\hat{\theta}_{LP(v,p)} = v! w_{LP(v,p)}^T Y_N$  with weights given by  $w_{LP(v,p)} = e_{p+2+v}^T (X_0^T W_0 X_0)^{-1} X_0^T W_0$ , where the matrices  $X_0$  ( $n \times 2p + 2$ ) and  $W_0$  ( $n \times n$ ) are defined as  $W_0 = \text{diag} \left( K_h(X_1) \ \dots \ K_h(X_n) \right)$  and  $X_0 = [\chi_1, \dots, \chi_n]^T$  with  $\chi_i = (1, X_i, \dots, X_i^p, D(X_i), X_i D(X_i), \dots, X_i^p D(X_i))$ . From this definition, it immediately follows that a generalization of the equality constraints of (1.3) hold, since the local polynomial estimator is unbiased for polynomials of order  $j \leq p$

$$\begin{aligned} \sum_{i=1}^n w_{i,+} X_i^v &= e_{p+2+v}^T (X_0^T W_0 X_0)^{-1} X_0^T W_0 X_0 e_{p+2+v} = 1 \\ \sum_{i=1}^n w_i X_i^v &= e_{p+2+v}^T (X_0^T W_0 X_0)^{-1} X_0^T W_0 X_0 e_{1+v} = 0 \\ \sum_{i=1}^n w_{i,+} X_i^j &= e_{p+2+v}^T (X_0^T W_0 X_0)^{-1} X_0^T W_0 X_0 e_{p+2+j} = 0 \\ \sum_{i=1}^n w_i X_i^j &= e_{p+2+v}^T (X_0^T W_0 X_0)^{-1} X_0^T W_0 X_0 e_{1+j} = 0 \quad \text{for } p \geq j \neq v. \end{aligned}$$

Moreover, the integral form of the remainder immediately provides us with a feasible bound  $r$  as  $\sup_{R \in \bar{\mathcal{F}}(1)} \mathbb{E} [\hat{\theta}_{LP(v,p)} - \theta^v | X_N] = \sup_{R \in \bar{\mathcal{F}}(1)} \mathbb{E} [v! \sum_{i=1}^n w_i R(X_i) | X_N] \leq \sum_{i=1}^n \left| w_i X_i^{p+1} \frac{v!}{(p+1)!} \right|$ . Thus, the problem (1.3) is strictly feasible, and, by Slater's condition, strong duality holds.

It turns out that the simple bound obtained via this approach is sharp for general local polynomial estimators. To show this explicitly, assume that  $\hat{\theta}_{LP(v,p)}$  is well defined and note that

$$\bar{w}_{LP(v,p),+}(t) = \sum_{X_i \geq t} w_{i,LP(v,p)}(X_i - t)^p = e_{2+p+v}^T \left( X_0^T W_0 X_0 \right)^{-1} X_0^T W_0 \begin{pmatrix} (X_1 - t)^p D(X_1 - t) \\ \vdots \\ (X_n - t)^p D(X_n - t) \end{pmatrix}.$$

It follows from regression principles that  $\bar{w}_{LP(v,p),+}(t)$  and  $\bar{w}_{LP(v,p),-}(t)$  can be understood as the  $(v+1)$ -th coefficients from a weighted polynomial regression of  $\Delta_{i,+}(t) = (X_i - t)^p D(X_i - t)$  and  $\Delta_{i,-}(t) = (X_i - t)^p (1 - D(X_i - t))$  on  $M_i = (1, X_i, \dots, X_i^p)$  based on subsets of the data with  $X_i \geq 0$  and  $X_i < 0$  respectively. The corresponding regressions in turn can be understood as Tikhonov regularized least-squares problems, which allows us to show that the regression coefficients  $\bar{w}_{LP(v,p),+}(t)$  and  $\bar{w}_{LP(v,p),-}(t)$  are attenuated versions of coefficients with deterministic sign. To show this formally, we define the following objects

$$\begin{aligned} \mathbf{X}_+(t) &= [M_i^T, \dots] : && \text{Matrix of } M_i \text{'s for units with } 0 \leq t \leq X_i, \\ \Gamma_+(t) &= [M_i^T, \dots] : && \text{Matrix of } M_i \text{'s for units with } 0 \leq X_i < t, \\ W_{\mathbf{X}_+}(t) &= \text{diag}(K_h(X_i), \dots) : && \text{Matrix of weights for units with } 0 \leq t < X_i, \\ W_{\Gamma_+}(t) &= \text{diag}(K_h(X_i), \dots) : && \text{Matrix of weights for units with } 0 \leq X_i < t, \\ \Delta_{X,+}(t) &= [\Delta_{i,+}(t), \dots] : && \text{Vector of } \Delta_{i,+}(t) \text{'s for units with } X_i \geq 0. \end{aligned}$$

Analogously,

$$\begin{aligned} \mathbf{X}_-(t) &= [M_i^T, \dots] : && \text{Matrix of } M_i \text{'s for units with } X_i \leq t \leq 0, \\ \Gamma_-(t) &= [M_i^T, \dots] : && \text{Matrix of } M_i \text{'s for units with } t \leq X_i < 0, \\ W_{\mathbf{X}_-}(t) &= \text{diag}(K_h(X_i), \dots) : && \text{Matrix of weights for units with } X_i \leq t \leq 0, \\ W_{\Gamma_-}(t) &= \text{diag}(K_h(X_i), \dots) : && \text{Matrix of weights for units with } t \leq X_i < 0, \\ \Delta_{X,-}(t) &= [\Delta_{i,-}(t), \dots] : && \text{Vector of } \Delta_{i,-}(t) \text{'s for units with } X_i < 0. \end{aligned}$$



Let  $\|\cdot\|_2$  denote the Euclidean norm. In this notation,  $\bar{w}_{LP(v,p),+}(t)$  and  $\bar{w}_{LP(v,p),-}(t)$  are the  $(v+1)$ -th elements of the solutions to the "regularized" least-squares problems

$$\min_{\gamma \in \mathbb{R}^{p+1}} \left\| W_{\mathbf{X}_{\pm}}(t)^{\frac{1}{2}} [\mathbf{X}_{\pm}(t)\gamma - \Delta_{X,\pm}(t)] \right\|_2^2 + \left\| W_{\Gamma_{\pm}}(t)^{\frac{1}{2}} \Gamma_{\pm}(t)\gamma \right\|_2^2.$$

Assume without loss of generality that  $\Delta_{X,+}(t)$  and  $\Delta_{X,-}(t)$  are sorted in ascending and descending order respectively. By least squares algebra, it then follows that  $\bar{w}_{LP(v,p),\pm}(t)$  is

$$\hat{\gamma}_v = e_{v+1}^T \left( \mathbf{X}_{\pm}(t)^T W_{\mathbf{X}_{\pm}}(t) \mathbf{X}_{\pm}(t) + \Gamma_{\pm}(t)^T W_{\Gamma_{\pm}}(t) \Gamma_{\pm}(t) \right)^{-1} \mathbf{X}_{\pm}(t)^T \begin{bmatrix} W_{\Gamma_{\pm}}(t) & \mathbf{0} \\ \mathbf{0} & W_{\mathbf{X}_{\pm}}(t) \end{bmatrix} \Delta_{X,\pm}(t).$$

Note that, for any  $t$ , the "unregularized" subproblem  $\min_{\gamma \in \mathbb{R}^{p+1}} \left\| W_{\mathbf{X}_{\pm}}(t)^{\frac{1}{2}} [\mathbf{X}_{\pm}(t)\gamma - \Delta_{X,\pm}(t)] \right\|_2^2$  corresponds to the simple weighted least squares regressions of  $(X_i - t)^p$  on  $(1, X_i, \dots, X_i^p)$  on the data sets defined by  $X_i \geq t$  and  $X_i < t$  respectively. By the binomial theorem, it holds for any  $p \geq 0$

$$(x - t)^p = \sum_{k=0}^p \binom{p}{k} x^{p-k} (-t)^k = \binom{p}{0} x^p (-t)^0 + \binom{p}{1} x^{p-1} (-t)^1 + \dots + \binom{p}{p} x^0 (-t)^p,$$

which implies that there always exists a perfect solution (in the mean squared error sense) to the "unregularized" subproblem, with coefficients that do not depend on the weights  $W_{\mathbf{X}_{\pm}}(t)$  or the data  $\mathbf{X}_{\pm}(t)$ . More importantly, the sign of the coefficients does only depend on the sign of  $t$  and the orders  $p$  and  $v$ . Note that, for  $t > 0$ , the coefficients  $\left\{ \binom{p}{k} (-t)^k \right\}_{k=0}^p$  in the polynomial expansion alternate in sign, starting with a positive coefficient on  $x^p$ . For  $t < 0$  all coefficients are positive.

It follows that for  $t \geq 0$ ,  $\bar{w}_{LP(p,v)}(t)$  is positive if  $p - v$  is even, and negative if  $p - v$  is odd, while  $\bar{w}_{LP(p,v)}(t)$  is positive for all  $t < 0$ . As a consequence, the conditional bias  $\int_{\mathbb{R}} \mu^{p+1}(t) \bar{w}(t) dt$  is maximized by setting  $\mu^{p+1}(t) = L$  if  $p - v$  is even and  $\mu^{p+1}(t) = -L \text{sign}(t)$  if  $p - v$  is odd. This yields the following general formula of the worst-case bias of a local polynomial estimator under  $p+1$  order bounds

$$\bar{B}_n(w_{LP(v,p)}) = \frac{v!L}{(p+1)!} \left[ (-1)^{p-v} \sum_{i=1}^n w_{i,+} X_i^{p+1} + \sum_{i=1}^n w_{i,-} X_i^{p+1} \right].$$

## 1.C Dual Optimization

Appendix 1.A justifies the equivalence of (1.1) and (1.3) by establishing (1.2). Appendix 1.B shows strong duality of (1.3) and the original dual underlying (1.5), implying that the solution to (1.1) can be obtained by solving the dual problem. To derive the objective in (1.5) and to obtain (1.4), which we used to recover the solution of (1.3), we relied on a reformulation of the dual objective that gave rise to a quadratic problem nested in the dual objective  $q(\nu, \lambda)$ . The problem was obtained by interchanging the order of the supremum and infimum in the original dual objective

$$\begin{aligned} q(\nu, \lambda) &= \inf_{(w_N, r)} L(w, r, \nu, \lambda) \\ &= \inf_{(w_N, r)} \sup_{R \in \bar{\mathcal{F}}(1)} \sum_{i=1}^n w_i^2 \sigma_i^2 + \kappa L^2 r^2 + \nu \left( \sum_{i=1}^n w_i R(X_i) - r \right) + \lambda_1 \left( \sum_{i=1}^n w_{i,+} \right) \\ &\quad + \lambda_2 \left( \sum_{i=1}^n w_{i,-} \right) + \lambda_3 \left( \sum_{i=1}^n w_{i,+} X_i - 1 \right) + \lambda_4 \left( \sum_{i=1}^n w_{i,-} X_i + 1 \right). \end{aligned}$$

*Proof.* To see that this is admissible, first note that the set of functions  $\bar{\mathcal{F}}(1)$  is convex. Let  $z \in [0, 1]$  and consider any two  $f_1, f_2 \in \bar{\mathcal{F}}(1)$ , with  $(x, x') \in \mathcal{X}_\pm$ . By definition of  $\bar{\mathcal{F}}(1)$ ,

$$\begin{aligned} z f_1(0) + (1 - z) f_2(0) &= z f_1^1(0) + (1 - z) f_2^1(0) = 0 \\ z |f_1^1(x) - f_1^1(x')| + (1 - z) |f_2^1(x) - f_2^1(x')| &\leq |x - x'| \end{aligned}$$

which implies that  $z f_1(x) + (1 - z) f_2(x) \in \bar{\mathcal{F}}(1)$ . Moreover, any sequence of functions  $f_n(x)$  in  $\bar{\mathcal{F}}(1)$  is uniformly bounded and uniformly equicontinuous. The first statement follows from arguments analogous to those in Appendix 1.A since  $|f_n(x)| \leq |f_0^x(x - t) dt| = 0.5x^2$  for all  $f_n \in \bar{\mathcal{F}}(1)$ . Uniform equicontinuity holds since for any  $(x, y)$  in  $[\underline{x}, \bar{x}]$  and sequence  $f_n$  in  $\bar{\mathcal{F}}(1)$

$$|f_n(x) - f_n(y)| \leq \left| \int_0^x (x - t) dt - \int_0^y (y - t) dt \right| = |0.5(x^2 - y^2)| \leq \max\{|\bar{x}|, |\underline{x}|\} |x - y|,$$

so that for  $\delta = \frac{\varepsilon}{\max\{|\bar{x}|, |\underline{x}|\}}$  it holds that  $|f_n(x) - f_n(y)| \leq \varepsilon$  whenever  $|x - y| < \delta$ . It follows from the Arzelà-Ascoli Theorem that the set  $\bar{\mathcal{F}}(1)$  is relatively compact in  $C([\underline{x}, \bar{x}])$ , the space of continuous real-valued functions on  $[\underline{x}, \bar{x}]$ . To obtain the result, define  $g : \mathbb{R}^{n+1} \times \bar{\mathcal{F}}(1) \mapsto \mathbb{R}^1$

$$\begin{aligned} g(\mathbf{x}, R | \nu, \lambda) &= \sum_{i=1}^n w_i^2 \sigma_i^2 + \kappa L^2 r^2 + \nu \left( \sum_{i=1}^n w_i R(X_i) - r \right) + \lambda_1 \left( \sum_{i=1}^n w_{i,+} \right) \\ &\quad + \lambda_2 \left( \sum_{i=1}^n w_{i,-} \right) + \lambda_3 \left( \sum_{i=1}^n w_{i,+} X_i - 1 \right) + \lambda_4 \left( \sum_{i=1}^n w_{i,-} X_i + 1 \right), \end{aligned}$$

where  $\mathbf{x} = (w_1, \dots, w_n, r)^T$ . Clearly,  $\mathbb{R}^{n+1}$  is convex and  $g$  is continuous in both arguments. Moreover, the function is, for any given  $R \in \bar{\mathcal{F}}(1)$ , convex over  $\mathbb{R}^{n+1}$  and, for any given  $\mathbf{x} \in \mathbb{R}^{n+1}$ , linear (concave) over  $\bar{\mathcal{F}}(1)$ . It then follows from Sion's minimax theorem (Sion, 1958) that

$$\inf_{\mathbf{x} \in \mathbb{R}^{n+1}} \sup_{R \in \bar{\mathcal{F}}(1)} g(\mathbf{x}, R|\nu, \lambda) = \sup_{R \in \bar{\mathcal{F}}(1)} \inf_{\mathbf{x} \in \mathbb{R}^{n+1}} g(\mathbf{x}, R|\nu, \lambda).$$

Interchanging the infimum and supremum in the dual objective substantially simplifies the minimax problem and allows us to obtain the result in (1.4). To make this explicit, define  $\mathbf{Q} = \text{diag}(2\sigma_1^2, \dots, 2\sigma_n^2, 2\kappa L^2)$ , let  $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x}$  and denote by  $f^*(\mathbf{y}) = \sup_{\mathbf{x}} [\mathbf{x}^T \mathbf{y} - f(\mathbf{x})]$  the conjugate function of  $f(\mathbf{x})$ . Moreover, let

$$\mathbf{C} = \begin{bmatrix} D(X_1) & \cdots & D(X_n) & 0 \\ 1 - D(X_1) & \cdots & 1 - D(X_n) & 0 \\ D(X_1)X_1 & \cdots & D(X_n)X_n & 0 \\ (1 - D(X_1))X_1 & \cdots & (1 - D(X_n))X_n & 0 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \quad \mathbf{A}^T = \begin{bmatrix} R(X_1) \\ \vdots \\ R(X_n) \\ -1 \end{bmatrix} \quad \mathbf{b} = 0.$$

In this notation, the dual objective after the exchange of supremum and infimum is given by

$$\begin{aligned} q(\nu, \lambda) &= \sup_{R \in \bar{\mathcal{F}}(1)} \inf_{\mathbf{x}} [f(\mathbf{x}) + \nu(\mathbf{A}\mathbf{x} - \mathbf{b}) + \lambda^T(\mathbf{C}\mathbf{x} - \mathbf{d})] \\ &= \sup_{R \in \bar{\mathcal{F}}(1)} -\lambda^T \mathbf{d} + \inf_{\mathbf{x}} (f(\mathbf{x}) + (\nu \mathbf{A} + \lambda^T \mathbf{C}) \mathbf{x}) \\ &= \sup_{R \in \bar{\mathcal{F}}(1)} -\lambda^T \mathbf{d} - f^*(-(\mathbf{A}^T \nu + \mathbf{C}^T \lambda)) \\ &= \sup_{R \in \bar{\mathcal{F}}(1)} -\lambda^T \mathbf{d} - \frac{1}{2} (\mathbf{A}^T \nu + \mathbf{C}^T \lambda)^T \mathbf{Q}^{-1} (\mathbf{A}^T \nu + \mathbf{C}^T \lambda) \\ &= \sup_{R \in \bar{\mathcal{F}}(1)} -\frac{1}{4} \sum_{i=1}^n \frac{[\lambda_1 D(X_i) + \lambda_2 (1 - D(X_i)) + \lambda_3 D(X_i) X_i + \lambda_4 (1 - D(X_i)) X_i + \nu R(X_i)]^2}{\sigma_i^2} \\ &\quad - \frac{1}{4} \frac{\nu^2}{\kappa L^2} - \lambda_3 + \lambda_4, \end{aligned}$$

Note that for  $\sigma_i > 0, L > 0, \kappa > 0$ , the objective function of the inner minimization problem in the dual objective is a quadratic form with positive definite Hessian. As a consequence, we obtain the analytical expressions (1.4) by  $(w_N, r) = -\mathbf{Q}^{-1}(\mathbf{A}^T \nu + \mathbf{C}^T \lambda)$  and can write  $q(\nu, \lambda)$  using its conjugate function. Finally, since by (1.2) the maximum exists, we can write the dual problem as a maximization problem over the space  $\bar{\mathcal{F}}(1)$  which yields (1.5).

*Remark.* In order to pass this problem to a numeric solver, it is useful to turn it into a minimization problem and to reparameterize it to avoid products of optimization parameters,  $\tilde{R}(X_i) \leftarrow \nu R(X_i)$ . This yields the baseline problem that we solve via discrete approximation.

$$\begin{aligned} \underset{\nu, \lambda, \tilde{R}}{\text{minimize}} \quad & \sum_{i=1}^n \frac{[\lambda_1 D(X_i) + \lambda_2(1 - D(X_i)) + \lambda_3 D(X_i)X_i + \lambda_4(1 - D(X_i))X_i + \tilde{R}(X_i)]^2}{4\sigma_i^2} \\ & + \frac{\nu^2}{4\kappa L^2} + \lambda_3 - \lambda_4 \end{aligned} \quad (1.7)$$

subject to  $\nu \in \mathbb{R}_+, \lambda \in \mathbb{R}^4, \tilde{R} \in \bar{\mathcal{F}}(\nu)$ .

## 1.D Implementation

Our R implementation solves the QP (1.7) by approximating the continuous argument  $\tilde{R}$  on a discrete equidistant grid with distance  $\Delta = h$  between any two adjacent grid points. Let  $\underline{x}$  and  $\bar{x}$  denote the minimal and maximal elements of  $\{X_i - c\}_{i=1}^n$ . In a first step, we divide the interval  $I = (\underline{x} - \frac{h}{2}, \bar{x} + \frac{h}{2}]$  in  $J = (\bar{x} - \underline{x})/h$  disjoint intervals  $I_j = (\underline{x} + (j - \frac{3}{2})h, \underline{x} + (j - \frac{1}{2})h]$ . Let  $x_j$  denote the center of interval  $j$ , so that  $x_1 = \underline{x}$  and  $x_J = \bar{x}$ . In a second step, we assign each data point  $X_i$  to its closest grid point. For example, if  $X_i$  falls into  $I_j$  we assign  $j \leftarrow i$  and store this mapping. Moreover, we keep track of the two grid points  $(x_{c,-}, x_{c,+})$  that are closest to the normalized cutoff zero (ties are possible only if zero itself is a grid point).

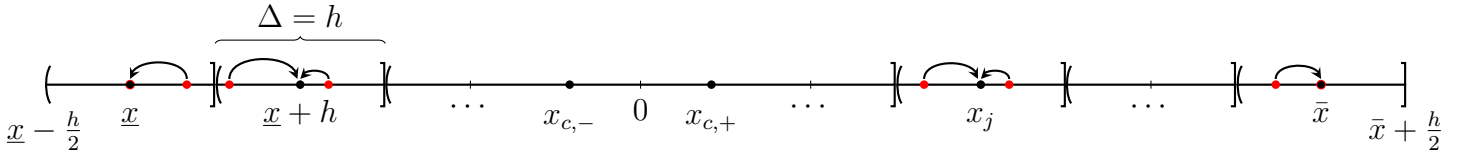


Figure 1.D.1: Illustration of the discretization strategy. Red (black) dots indicate data (grid) points.

Note that by construction  $0 \geq x_{c,-} \xrightarrow{J \rightarrow \infty} 0$  and  $0 \leq x_{c,+} \xrightarrow{J \rightarrow \infty} 0$  since  $h \xrightarrow{J \rightarrow \infty} 0$ . Since  $\mu \in \mathcal{F}(L)$ ,

$$\begin{aligned} \tilde{R}(0) - \tilde{R}(x_{c,-}) &= O(|x_{c,-}|) \text{ as } x_{c,-} \rightarrow 0 & \tilde{R}^1(0) - \frac{\tilde{R}(0) - \tilde{R}(x_{c,-})}{|x_{c,-}|} &= O(|x_{c,-}|) \text{ as } x_{c,-} \rightarrow 0 \\ \tilde{R}(0) - \tilde{R}(x_{c,+}) &= O(|x_{c,+}|) \text{ as } x_{c,+} \rightarrow 0 & \tilde{R}^1(0) - \frac{\tilde{R}(x_{c,+}) - \tilde{R}(0)}{|x_{c,+}|} &= O(|x_{c,+}|) \text{ as } x_{c,+} \rightarrow 0 \end{aligned}$$

and we take into account the constraint  $\tilde{R}(0) = \tilde{R}^1(0) = 0$  by enforcing  $\tilde{R}(x_{c,-}) = \tilde{R}(x_{c,+}) = 0$ . In order to take into account the second order derivative constraint  $|\tilde{R}^2(x)| \leq L$ , we utilize the equidistance of grid points to impose the constraints via second order central difference

approximations  $\tilde{R}^2(x) - [\tilde{R}(x+h) - 2\tilde{R}(x) + \tilde{R}(x-h)]h^{-2} = O(h^2)$  as  $h \rightarrow 0$ . This yields 2 linear equality and  $2J$  linear inequality constraints<sup>12</sup> that we enforce during the optimization

$$\begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}^T \begin{bmatrix} \tilde{R}(\underline{x}) \\ \vdots \\ \tilde{R}(x_{c,-}) \\ \tilde{R}(x_{c,+}) \\ \vdots \\ \tilde{R}(\bar{x}) \end{bmatrix} = \mathbf{0}_2 \quad \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ -1 & 2 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & 0 \end{bmatrix} \begin{bmatrix} \tilde{R}(\underline{x}) \\ \tilde{R}(\underline{x}+h) \\ \vdots \\ \tilde{R}(\bar{x}) \end{bmatrix} \leq \mathbf{1}_{2J}\nu h^2.$$

Let  $\mathbf{A}_1$  and  $\mathbf{A}_2$  denote the two matrices above. In order to write the quadratic program (1.7) in standard form, we define  $\tilde{\mathbf{x}} = (\tilde{w}_J, \nu, \lambda, \tilde{R}_J)^T$ , where  $\tilde{w}_j = -2\sigma(x_j)^2 w_j$  and  $\tilde{R}_j = \nu R(x_j)$  are the reparameterized weights and remainders evaluated at the approximation points. The quadratic and linear part of the objective are  $\tilde{\mathbf{H}} = \text{diag}(0.5\sigma(\underline{x})^{-2}, \dots, 0.5\sigma(\bar{x})^{-2}, 0.5\kappa^{-1}L^{-2}, \mathbf{0}_{J+4}^T)$  and  $\tilde{d} = (\mathbf{0}_J^T, 0, 0, 0, 1, -1, \mathbf{0}_J^T)^T$ . In this notation, (1.7) under discretization of  $\tilde{R} \in \bar{\mathcal{F}}(\nu)$  is

$$\begin{aligned} & \underset{\tilde{\mathbf{x}}}{\text{minimize}} && \frac{1}{2}\tilde{\mathbf{x}}^T \tilde{\mathbf{H}} \tilde{\mathbf{x}} + \tilde{d}^T \tilde{\mathbf{x}} \\ & \text{subject to} && [-\mathbf{I}_J, \mathbf{0}_{J \times 1}, \tilde{\mathbf{C}}^T \lambda, \mathbf{I}_J] \tilde{\mathbf{x}} = \mathbf{0}_J \\ & && [\mathbf{0}_{2 \times J+5}, \mathbf{A}_1] \tilde{\mathbf{x}} = \mathbf{0}_2 \\ & && [\mathbf{0}_{2J \times J+5}, \mathbf{A}_2] \tilde{\mathbf{x}} \leq \mathbf{1}_{2J}\nu h^2, \nu \in \mathbb{R}_+, \end{aligned} \tag{1.8}$$

where  $\tilde{\mathbf{C}}$  is defined in analogy to the matrix  $\mathbf{C}$  in Appendix 1.C, with the approximation points  $x_j$  ( $j = 1 \dots, J$ ) playing the role of the data points  $X_i$  ( $i = 1, \dots, n$ ). In the implementation, we reformulate (1.8) in terms of the equivalent Second Order Cone Program to avoid issues of standard solvers associated with the semidefiniteness of  $\tilde{\mathbf{H}}$  induced by the  $J+4$  zero rows. To this end, let  $q$  be an additional parameter and note that since  $\tilde{\mathbf{H}}$  is positive semidefinite, we can write  $\tilde{\mathbf{x}}^T \tilde{\mathbf{H}} \tilde{\mathbf{x}} = \tilde{\mathbf{x}}^T \mathbf{R}^T \mathbf{R} \tilde{\mathbf{x}} = \|\mathbf{R} \tilde{\mathbf{x}}\|_2^2$ , allowing us to equivalently write the problem with an objective that is linear in  $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}, q)^T$

$$\begin{aligned} & \underset{\tilde{\mathbf{x}}}{\text{minimize}} && e_{2J+6}^T \tilde{\mathbf{x}} + [\tilde{d}, 0]^T \tilde{\mathbf{x}} \\ & \text{subject to} && q + 1 \geq \sqrt{\|\mathbf{R} \tilde{\mathbf{x}}\|_2^2 + q^2} \\ & && \vdots \end{aligned}$$

<sup>12</sup>Note that we need to enforce the second order constraints only at grid points that have data points in their respective interval. We can thus increase precision without linearly growing the number of constraints.

Given a solution to (1.8), we assign weights to data points according to the mapping between the index sets of  $X_i$  and  $x_j$ , and rescale each weight to satisfy the moment conditions exactly.

## 1.E Proof of Proposition 1

The honesty property of the interval (1.6) for any given bound  $L$  relies on an (asymptotic) upper bound for the worst-case smoothing bias and uniform convergence of  $s_n^{-1}[\hat{\theta} - \mathbb{E}[\hat{\theta}|X_n]] \xrightarrow{D} N(0, 1)$  over  $\mathcal{F}(L)$ . It then follows from standard arguments that a confidence interval based on the approximation  $(\hat{\theta} - \theta)/s_n \sim |N(\bar{t}_n, 1)|$  yields asymptotically uniformly valid confidence intervals.

*Proof.* We first show that under Assumption 1, conditionally on  $X_N$ , Lyapunov's condition applies uniformly over all permitted CEFs. By Lyapunov's Central Limit Theorem and assumptions A1 (i)-(ii) this requires us to show that for all  $\mu \in \mathcal{F}(L)$

$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} \left[ |w_i(Y_i - \mu(X_i)) - \mathbb{E}[w_i(Y_i - \mu(X_i))|X_N]|^{2+\delta} |X_N \right] = o_p(1).$$

By the definition of  $\mu$ ,  $\mathbb{E}[w_i(Y_i - \mu(X_i))|X_N] = 0$ , and the Lyapunov condition holds since

$$\begin{aligned} \frac{\sum_{i=1}^n \mathbb{E} \left[ |w_i(Y_i - \mu(X_i))|^{2+\delta} |X_N \right]}{\left[ \sqrt{\sum_{i=1}^n \sigma_i^2 w_i^2} \right]^{2+\delta}} &\leq \frac{\sum_{i=1}^n |w_i|^{2+\delta} \mathbb{E} \left[ |Y_i - \mu(X_i)|^{2+\delta} |X_N \right]}{\left[ \sqrt{\sum_{i=1}^n \sigma_i^2 w_i^2} \right]^{2+\delta}} \\ &\leq C \frac{\sum_{i=1}^n |w_i|^{2+\delta}}{\left[ \sqrt{\sum_{i=1}^n w_i^2 \sigma_i^2} \right]^{2+\delta}} && \text{by A1 (iv)} \\ &\leq C \frac{\max_i \{w_i\}^2 \sum_{i=1}^n |w_i|^\delta}{\left[ \sigma_{\min} \sqrt{\sum_{i=1}^n w_i^2} \right]^{2+\delta}} && \text{by A1 (iii)} \\ &= \frac{C}{\sigma_{\min}^{2+\delta}} \frac{\max_i \{w_i\}^2}{\sum_{i=1}^n w_i^2} = o_p(1) && \text{by A1 (v)} \end{aligned}$$

where the second and third line utilize the uniform moment bounds and the last equation relies on the limit behavior of the ratio  $\bar{w}_R$ . It then follows from standard arguments that a uniform asymptotic upper bound on undercoverage of (1.6) is given by

$$\liminf_{n \rightarrow \infty} \left( \inf_{\mu \in \mathcal{F}(L)} \Pr \left[ \left| s_n^{-1}[\hat{\theta} - \mathbb{E}[\hat{\theta}|X_N]] + t_n \right| \leq \text{cv}_{1-\alpha}(\bar{t}_n) \mid X_N \right] - \inf_{\mu \in \mathcal{F}(L)} \Pr \left[ |Z + t_n| \leq \text{cv}_{1-\alpha}(\bar{t}_n) \mid X_N \right] \right) = 0,$$

where  $Z$  denotes a standard normal variable. Then by the definition of  $\text{cv}_{1-\alpha}$  and  $\bar{t}_n$  it holds

$$\liminf_{n \rightarrow \infty} \inf_{\mu \in \mathcal{F}(L)} \Pr \left[ \mu_+^1(0) - \mu_-^1(0) \in \mathcal{I}_\alpha \right] \geq 1 - \alpha.$$

## 1.F Extensions

**1.F.1 Fuzzy Discontinuity Design.** The inference problem discussed in the main body of the paper is a special case of the general problem of inference on the ratio of jumps in the  $\nu$ -th derivative of two CEFs at a point (normalized to zero). Let  $\mu_Y(x) = \mathbb{E}[Y_i | X_i = x]$  and  $\mu_T(x) = \mathbb{E}[T_i | X_i = x]$ . The general parameter of interest in discontinuity designs is

$$\tau_{RD} = \frac{\mu_{Y,+}^\nu(0) - \mu_{Y,-}^\nu(0)}{\mu_{T,+}^\nu(0) - \mu_{T,-}^\nu(0)} = \frac{\theta_Y^\nu}{\theta_T^\nu}, \quad (1.9)$$

under the assumption that  $\theta_T^\nu \neq 0$ , with  $\nu = 0$  corresponding to RDDs and  $\nu = 1$  to RKDs. First note that, provided a bound on the magnitude of the respective derivatives of order  $p > \nu$ , we can, irrespective of the order  $\nu$ , employ the optimization approach discussed in Section 1.3 separately to  $\theta_Y^\nu$  and  $\theta_T^\nu$  by restricting the feasible set via  $\sum_{i=1}^n w_{i,+} X_i^k = -\sum_{i=1}^n w_{i,-} X_i^k = \delta_{\nu,k}$  as in Appendix 1.A. This is because, by the arguments in Appendix 1.B, we can always derive a suitable constraint qualification to employ the logic of Section 1.3.

In the sharp case, that is if the institutional rule  $T$  governing treatment assignment is deterministic  $\mathbb{V}[T(X_i) | X_i] = 0$  and fully implemented  $\Pr\{T_i = T(X_i)\} = 1$ , it is sufficient to do so for just  $\theta_Y^\nu$  and to conduct inference on  $\tau_{RD}$  in complete analogy to Section 1.3. However, in the fuzzy setting, that is if the rule is probabilistic, or if there are deviations from the assignment rule, the denominator needs to be estimated and inference on  $\tau_{RD}$  requires a method that addresses the nonlinearity induced by the ratio. As a consequence, additional complications for inference arise and our optimization approach can not be directly employed. Fuzzy designs can arise for various reasons such as noncompliance, multivariate rules with unobserved assignment variables or measurement errors, and occur frequently in applications.

The standard approach to inference in fuzzy settings is to estimate  $\theta_Y^\nu$  and  $\theta_T^\nu$  separately and rely on a linearization of the estimator ratio to build confidence intervals based on a delta method argument. The delta method approach to inference in discontinuity designs has three important limitations. The first two concern the validity of the distributional approximation in settings with discrete assignment variables or weak identification, that is if the denominator  $\theta_T$  is close to zero. In both cases, the validity of the approximation breaks down, as otherwise

asymptotically negligible terms in the expansion have non-zero probability limits. A third problem is that, due to the nonlinearity of the estimator, the bias-aware approach can not control the exact worst-case smoothing bias.

Motivated by these shortcomings of delta method based inference, [Noack and Rothe \(2019\)](#) propose a method to conduct inference based on an Anderson-Rubin type inversion argument ([Anderson and Rubin, 1949](#)). This approach is known to allow for valid pointwise inference under weak identification and importantly also allows them to apply bias-aware methods to construct honest confidence sets, as the method avoids linearization. In order to extend our optimization based approach to fuzzy designs, we employ their strategy.

The basic idea is to consider the parameter  $\theta^v(\tau) = \theta_Y^v - \tau\theta_T^v$  and to construct honest  $(1 - \alpha)$  confidence intervals for  $\theta^v(\tau)$  for different values of  $\tau$ . Note that by definition of  $\theta_Y^v$  and  $\theta_T^v$ , this corresponds to constructing CIs for the jump in the  $v$ -th derivative of  $\mu_\tau(x) = \mathbb{E}[Y_i - \tau T_i | X_i = x]$  at the discontinuity point for any given value of  $\tau$ . A confidence set for  $\tau_{RD}$  is then obtained by collecting all values of  $\tau$  for which the auxiliary interval covers zero, and adjusting this set depending on whether an honest confidence interval for  $\theta_T^v$  contains zero or not. The procedure thus reduces the fuzzy inference problem to repeated sharp problems and allows us to immediately employ optimized intervals. The only particularity that arises under this strategy relates to the shape of the confidence sets, which varies depending on whether the auxiliary confidence interval for  $\theta_T^v$  covers zero or not. Let  $\mathcal{I}_T = [\underline{\theta}_T^v, \bar{\theta}_T^v]$  denote an optimized interval for  $\theta_T^v$  and let  $\mathcal{I}_{\tau_{RD}} = [\underline{\tau}_{RD}, \bar{\tau}_{RD}]$  denote the confidence interval<sup>13</sup> obtained by employing the above method. The following five cases are possible for some  $a < b$ .

Table 1.F.1: Shape of Anderson-Rubin Confidence Sets for  $\tau_{RD}$

	Case	Shape of Confidence Set
i)	$0 \notin \mathcal{I}_T$	$[a, b]$
ii)	$0 \in \mathcal{I}_T, 0 \notin \mathcal{I}_\tau$	$(-\infty, a] \cup [b, \infty)$ .
iii)	$0 \in \mathcal{I}_T, 0 \in \mathcal{I}_\tau$	$(-\infty, \infty)$
iv)	$\underline{\theta}_T = 0, \underline{\tau}_{RD} > 0$	$[a, \infty)$
v)	$\bar{\theta}_T = 0, \bar{\tau}_{RD} < 0$	$(-\infty, b]$

See [Noack and Rothe \(2019\)](#) for a detailed discussion of the construction and properties of bias-aware Anderson-Rubin Confidence Sets based on local polynomial estimators. As discussed in their paper, the cases iv) and v) are empirically irrelevant, as the events  $\underline{\theta}_T = 0$  and  $\bar{\theta}_T = 0$  occur with probability zero. Adapting their findings to our setting, we impose the following

<sup>13</sup> $\mathcal{I}_\tau$  has this shape by continuity of  $w^*$ ,  $\bar{B}(w_N)$  and  $s_n$  in  $\tau$ , and the folded normal quantile function in  $\bar{t}_n$ .



modification of Assumption 1 to ensure that the Anderson-Rubin Confidence Sets obtained via optimization are well defined and honest.

**Assumption 1.F.1** Let  $(C, \delta, \sigma_{\min}, \sigma_{\max}) \in \mathbb{R}_+^4$  be some fixed vector.

(i)  $\{Y_i, X_i, T_i\}_{i=1}^n$  is an i.i.d. random sample of size  $n$  from a fixed population.

(ii)  $(\mu_Y, \mu_T) \in \mathcal{F}(L) \times \mathcal{F}_T(L_T)$  for some  $L \geq 0$  and  $L_T \geq 0$  with

$$\mathcal{F}_p(L) = \{f : |f_{\pm}^{p-1}(x) - f_{\pm}^{p-1}(x')| \leq L|x - x'|, (x, x') \in \mathcal{X}_{\pm}\}$$

$$, \quad \mathcal{F}_T(L) = \{f : f \in \mathcal{F}(L_T) \wedge |f_+^v(0) - f_-^v(0)| > 0\}.$$

(iii) For all  $x \in R_X$ ,  $(\mu_Y, \mu_T) \in \mathcal{F}(L) \times \mathcal{F}_T(L_T)$  and  $\tau \in \mathbb{R}$ ,

$$0 < \sigma_{\min}^2 \leq E[(Y_i - \tau T_i - \mu_{\tau}(X_i))^2 | X_i = x] \leq \sigma_{\max}^2.$$

(iv) For all  $x \in R_X$ ,  $(\mu_Y, \mu_T) \in \mathcal{F}(L) \times \mathcal{F}_T(L_T)$  and  $\tau \in \mathbb{R}$ ,

$$\mathbb{E}[|Y_i - \tau T_i - \mu_{\tau}(X_i)|^{2+\delta} | X_i = x] \leq C.$$

(v) For each  $\tau \in \mathbb{R}$ , the solution  $w^*$  satisfies

$$\frac{\max_i w_i^2}{\sum_{i=1}^n w_i^2} \xrightarrow{P} 0$$

The adjustments in Assumption 1.F.1 ensure that (1.9) is well defined and that we can construct honest CIs for  $\theta_T^v$  and  $\theta_{\tau RD}^v$  (for each value of  $\tau$ ), by utilizing the bound on the  $p$ -th derivatives encoded in  $\mathcal{F}_p(L)$  and  $\mathcal{F}_T(L_T)$ . Note that by the logic discussed in Section 1.3, the worst-case bias of our estimator over  $\mathcal{F}(L) \times \mathcal{F}_T(L_T)$  is proportional to  $L + |\tau|L_T$ . In order to compute the confidence sets, it is thus required to construct optimized intervals for different values of  $\tau$ , taking into account the relevant functional constraint.

In order to implement this procedure, we first compute a length-optimized interval for  $\theta_T^v$ . In a second step, we compute UMSE optimized intervals on a prespecified grid of values of  $\tau$  by calling our method on data with modified outcome variable  $\tilde{Y}_i = Y_i - \tau T_i$ , adjusting the smoothness bound as needed. If the grid contained a sufficiently large range of values of  $\tau$ , this yields approximate values for  $\underline{\tau}_{RD}$  and  $\bar{\tau}_{RD}$ . We then compute length-optimized intervals starting at the two approximate bounds until we found the roots of  $\hat{\theta}^v(\tau) \pm s_n(\tau)cv_{1-\alpha}(\bar{t}_n(\tau))$ . Finally, we report confidence sets according to Table F.1.

The method of [Noack and Rothe \(2019\)](#) thus allows us to extend the optimization approach to fuzzy RKDs and RDDs. The latter is a special case of the multivariate RDD problem considered in [Imbens and Wager \(2019\)](#), which we extend to the (univariate) fuzzy setting. Given the results in [Imbens and Wager \(2019\)](#) and our simulations for RKDs one would expect these intervals to be particularly useful in fuzzy settings with discrete running variables.

**1.F.2 Shape Constraints.** An attractive feature of the honest optimization approach to RD inference in sharp designs is that it is rather simple to utilize additional structural information on the CEF under consideration. In our view, this is a noteworthy feature of combining the bias-aware approach with optimization techniques, as utilizing such information for inference is typically rather difficult. This is because in general the distribution of a restricted estimator depends in a non-trivial fashion on whether and where the shape constraints are binding, which is typically not known a priori (Freyberger and Reeves, 2018).

Under the bias-aware approach, this dependence operates through the worst-case bias and is rather simple in structure. In principle, any additional constraint that can be approximated in terms of finite differences of Taylor remainders and that does not break the convexity and compactness of  $\bar{\mathcal{F}}(1)$  can be directly utilized to sharpen inference. Such information can easily be incorporated by adding suitable constraints to (1.8). For example, concavity of  $\mu$  could be utilized by the following modification to the feasible set of (1.8)

$$\begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ -1 & 2 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & 0 \end{bmatrix} \begin{bmatrix} \tilde{R}(\underline{x}) \\ \tilde{R}(\underline{x} + h) \\ \vdots \\ \tilde{R}(\bar{x}) \end{bmatrix} \leq \begin{pmatrix} \mathbf{0}_J \\ \mathbf{1}_J \nu h^2 \end{pmatrix}.$$

Analogously, monotonicity of  $\mu$  can be imposed via the constraint

$$\begin{bmatrix} 1 & 0 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \end{bmatrix} \begin{bmatrix} \tilde{R}(\underline{x}) \\ \tilde{R}(\underline{x} + h) \\ \vdots \\ \tilde{R}(\bar{x}) \end{bmatrix} > (<) \mathbf{0}_J.$$

This is particularly useful for applied work that utilizes regression kink designs, as the "plateau"-schedules that often form the basis for kink designs tend to generate empirical CEFs that plausibly satisfy shape constraints. Note that this applies directly only to the sharp setting, as the properties of  $\theta_\tau(c)$  depend on both  $\mu_Y$  and  $\mu_T$  and even  $c$ . In the fuzzy setting one therefore needs to be more careful when thinking about imposing shape constraints in this fashion.

## 1.G Runtime Estimates

Table 1.G.1 shows average runtimes of the procedure proposed in this paper for 100 simulated runs for each of the sample sizes reported in the first column. The data was drawn from a continuous uniform distribution on  $[-1, 1]$  with  $L = 2$  and  $\sigma^2 = 0.1^2$ . The computations were conducted on a standard desktop computer using R 4.1.0. and the implementation provided in the replication files.

Table 1.G.1: Runtime Estimates

Sample Size	UMSE-optimal	HL-optimal
500	0.115	2.738
1000	0.109	2.813
1500	0.108	2.849
2000	0.116	3.076
2500	0.120	3.030
3000	0.117	3.141
3500	0.120	3.226
4000	0.129	3.286
4500	0.129	3.115
5000	0.127	3.304
5500	0.130	3.568
6000	0.129	3.590
6500	0.133	3.419
7000	0.132	3.458
7500	0.136	3.365
8000	0.132	3.320
8500	0.138	3.419
9000	0.141	3.641
9500	0.145	3.806
10000	0.144	3.899

*Note:* Runtime estimates in seconds for UMSE and length optimized linear confidence intervals for different sample sizes.

## 1.H Replication Files and Programs

Table 1.H.1 contains all R packages required to run the replication files that reproduce the numerical results and figures reported in this paper, as well as additional packages (\*) that are not required for the replication but necessary to utilize all configurations of the provided functions. The second column states the package versions under which the results were produced.

Table 1.H.1: Dependencies

Package	Version	Purpose
RMosek	9.2	optimization
Matrix	1.3-4	computation
foreach	1.5.1	parallelization
doParallel	1.0.16	parallelization
doSNOW	1.0.19	parallelization
doRNG	1.8.2	replication
FRD*	1.0.1	uniformly valid nn-variance estimator
rdrobust	0.98	robust bias-correction bandwidths
reshape2	1.4.4	reshaping data
ggplot2	3.3.5	visualization
latex2exp	0.5.0	visualization

*Note:* R packages required to run the replication files .

In order to replicate the simulation results reported in Tables 1.1 and 1.2 in a reasonable time frame it is necessary to install RMosek, an R interface for the Mosek software library designed to solve large-scale convex optimization problems. This is generally recommended for implementing the proposed optimized linear confidence intervals on medium to large datasets. If RMosek is not installed, the `optrkd` function provided in the replication files will rely on a solver that is integrated in base R, which tends to take longer, especially on large datasets, and has lower accuracy. The installation of RMosek is thus required to match the results numerically. The following link explains how RMosek can be downloaded and installed: <https://docs.mosek.com/9.2/rmosek/install-interface.html>. Academic licenses are available free of charge and can be requested here: <https://www.mosek.com/products/academic-licenses/>. The FRD package is the companion package of Noack and Rothe (2019) and can be downloaded [here](#). All other packages can be downloaded and installed from CRAN via the standard commands.

The files described in Table 1.H.2 contain the relevant functions and programs that replicate the results in the order in which they appear in this paper. All output is precomputed and stored in the `results_local` and `figures` folders. The simulation study reported in Tables 1.1 and 1.2 requires approximately 14 hours at 6 kernels to complete. Numerically similar results can be obtained by reducing the number of Monte Carlo runs  $M$  to 2500.

Table 1.H.2: Replication Files

Name	Description
<i>Functions</i>	
<code>srd.R</code>	functions for optimized linear RD intervals
<code>frd.R</code>	functions for optimized linear FRD intervals
<code>lp_unif.R</code>	functions for fixed length local polynomial RD intervals
<code>utils.R</code>	utility functions
<code>cvar.R</code>	functions to estimate the conditional variance
<code>rot.R</code>	functions to compute rule of thumbs for $L$
<code>hpc_sim.R</code>	function to conduct the simulation study in Appendix 1.I
<code>rel_risk.R</code>	functions to conduct a simulation study
<i>Programs</i>	
<code>plt_cefs.R</code>	plots the CEFs displayed in Figure 1
<code>num_tbls12.R</code>	conducts the simulation study reported in Table 1 and 2
<code>plt_relrisk.R</code>	computes and plots the results reported in Figure 2
<code>plt_discweights.R</code>	computes and plots the weights displayed in Figure 3
<code>plt_application.R</code>	schedule and binscatter plots displayed in Figures 4 and 5
<code>num_application.R</code>	computes the estimates and intervals reported in Section 5

*Note:* Description of the R files that reproduce the results reported in the main body.

The code in the file `srd.R` is a modified and adapted version of [Imbens and Wager \(2019\)](#) original code for optimized inference in regression discontinuity designs. In particular, the discrete approximation of the second order constraint on the remainder is due to them and theoretically justified by Proposition 2 in their paper. All other code was produced autonomously.

## 1.I Additional Simulation Results

Tables 1.I.1 and 1.I.2 contain the results of a more extensive Monte Carlo study based on 20,000 Monte Carlo runs. The setup is the same as in Section 1.4.2 and the results remain qualitatively the same. The tables, including additional columns, are available in the `results_cluster` folder as `tableX_panelX_large20.rda` and are computed using `hpc_sim.R`.

Table 1.I.1: Monte Carlo Results: Coverage and Relative Length I

$\mu_1(x)$		$L = 2$			$L = 6$		
Method	Tuning	Cov.	RL	$h/\kappa$	Cov.	RL	$h/\kappa$
<i>Cont. Design</i>							
Conv.	$h_{PMSE}$	43.4	0.567	0.225	18.4	0.398	0.183
US	$h_{US}$	87.1	1.000	0.154	64.0	0.704	0.125
RBC	$b_{PMSE}, h_{PMSE}$	54.3	0.850	0.225 (0.462)	38.0	0.583	0.183 (0.388)
RBC	$h = b = h_{PMSE}$	95.2	2.150	0.225	95.0	1.510	0.183
RBC	$b_{CE}, h_{CE}$	84.6	1.160	0.154 (0.462)	69.6	0.808	0.125 (0.388)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.3	4.590	0.166	94.7	4.030	0.104
LL-FL	$h_{UMSE, L=\hat{L}}$	95.1	1.770	0.166	94.9	1.560	0.104
LL-FL	$h_{HL, L=\hat{L}}$	96.6	1.610	0.219	93.1	1.470	0.138
Opt	$\kappa_{UMSE, L=\hat{L}}$	95.0	1.750	1.000	95.1	1.550	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	96.6	1.650	0.246	93.1	1.460	0.246
LL-FL	$h_{UMSE, L=2}$	95.1	1.060	0.187	21.1	0.550	0.187
LL-FL	$h_{HL, L=2}$	96.7	1.000	0.247	0.38	0.517	0.247
Opt	$\kappa_{UMSE, L=2}$	95.0	1.060	1.000	21.0	0.549	1.000
Opt	$\kappa_{HL, L=2}$	96.7	1.000	0.246	0.34	0.517	0.246
LL-FL	$h_{UMSE, L=6}$	99.2	2.060	0.121	94.9	1.060	0.121
LL-FL	$h_{HL, L=6}$	100.0	1.940	0.160	94.8	1.000	0.160
Opt	$\kappa_{UMSE, L=6}$	99.3	2.050	1.000	95.0	1.060	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.940	0.244	94.8	1.000	0.244
<i>Disc. Design</i>							
Conv.	$h_{PMSE}$	45.2	0.579	0.224	18.0	0.404	0.183
US	$h_{US}$	87.6	1.040	0.153	65.9	0.725	0.126
RBC	$b_{PMSE}, h_{PMSE}$	54.7	0.870	0.224 (0.438)	37.5	0.594	0.183 (0.389)
RBC	$h = b = h_{PMSE}$	95.4	2.320	0.224	95.2	1.670	0.183
RBC	$b_{CE}, h_{CE}$	85.1	1.210	0.153 (0.438)	70.4	0.834	0.126 (0.389)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.1	4.600	0.169	94.9	4.420	0.108
LL-FL	$h_{UMSE, L=\hat{L}}$	95.4	1.890	0.169	95.4	1.750	0.108
LL-FL	$h_{HL, L=\hat{L}}$	97.2	1.790	0.223	93.5	1.660	0.141
Opt	$\kappa_{UMSE, L=\hat{L}}$	95.4	1.850	1.000	95.3	1.710	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	97.1	1.750	0.245	93.5	1.620	0.243
LL-FL	$h_{UMSE, L=2}$	95.0	1.070	0.189	19.8	0.548	0.189
LL-FL	$h_{HL, L=2}$	96.8	1.000	0.251	0.28	0.515	0.251
Opt	$\kappa_{UMSE, L=2}$	95.1	1.060	1.000	20.9	0.546	1.000
Opt	$\kappa_{HL, L=2}$	96.9	1.000	0.242	0.32	0.513	0.242
LL-FL	$h_{UMSE, L=6}$	99.4	2.100	0.125	94.9	1.080	0.125
LL-FL	$h_{HL, L=6}$	100.0	1.960	0.163	94.9	1.010	0.163
Opt	$\kappa_{UMSE, L=6}$	99.3	2.080	1.000	95.0	1.070	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.950	0.237	94.8	1.000	0.237

*Note:* Empirical coverage rate (Cov.), average length relative to optimized interval with true smoothness bound (RL) and tuning parameter choice ( $h/\kappa$ ) of conventional (Conv.), undersmoothed (US), robust bias-corrected (RBC), fixed length (LL-FL) and optimized linear (Opt) 95% confidence intervals over 20,000 Monte Carlo draws. The pilot RBC bandwidth is reported in parentheses if it is selected separately from  $h$

Table 1.1.2: Monte Carlo Results: Coverage and Relative Length II

$\mu_2(x)$		$L = 2$			$L = 6$		
Method	Tuning	Cov.	RL	$h/\kappa$	Cov.	RL	$h/\kappa$
<i>Cont. Design</i>							
Conv.	$h_{PMSE}$	43.1	0.600	0.221	56.0	0.595	0.143
US	$h_{US}$	84.6	1.060	0.151	80.2	1.050	0.098
RBC	$b_{PMSE}, h_{PMSE}$	62.5	0.911	0.221 (0.439)	83.2	0.781	0.143 (0.336)
RBC	$h = b = h_{PMSE}$	95.4	2.270	0.221	94.0	2.260	0.143
RBC	$b_{CE}, h_{CE}$	86.8	1.240	0.151 (0.439)	86.6	1.150	0.098 (0.336)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.1	5.710	0.134	94.4	4.970	0.088
LL-FL	$h_{UMSE, L=\hat{L}}$	98.3	2.200	0.134	98.7	1.920	0.088
LL-FL	$h_{HL, L=\hat{L}}$	99.3	2.070	0.177	99.6	1.180	0.115
Opt	$\kappa_{UMSE, L=\hat{L}}$	98.3	2.190	1.000	98.7	1.910	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	99.3	2.060	0.246	99.5	1.800	0.246
LL-FL	$h_{UMSE, L=2}$	95.0	1.060	0.187	20.8	0.550	0.187
LL-FL	$h_{HL, L=2}$	95.1	1.000	0.247	0.04	0.517	0.247
Opt	$\kappa_{UMSE, L=2}$	95.0	1.060	1.000	20.7	0.549	1.000
Opt	$\kappa_{HL, L=2}$	95.0	1.000	0.246	0.05	0.517	0.246
LL-FL	$h_{UMSE, L=6}$	99.2	2.060	0.121	94.9	1.060	0.121
LL-FL	$h_{HL, L=6}$	100.0	1.940	0.160	94.8	1.000	0.160
Opt	$\kappa_{UMSE, L=6}$	99.3	2.050	1.000	95.1	1.060	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.940	0.244	94.9	1.000	0.244
<i>Disc. Design</i>							
Conv.	$h_{PMSE}$	44.8	0.622	0.218	54.1	0.608	0.143
US	$h_{US}$	84.7	1.120	0.149	81.1	1.000	0.106
RBC	$b_{PMSE}, h_{PMSE}$	62.3	0.941	0.218 (0.436)	83.4	0.803	0.143 (0.334)
RBC	$h = b = h_{PMSE}$	95.5	2.530	0.218	94.1	2.690	0.143
RBC	$b_{CE}, h_{CE}$	86.0	1.300	0.149 (0.436)	87.8	1.120	0.106 (0.334)
RBC	$h = b = h_{UMSE, L=\hat{L}}$	95.2	6.010	0.137	94.9	5.740	0.092
LL-FL	$h_{UMSE, L=\hat{L}}$	98.4	2.430	0.137	99.0	2.060	0.092
LL-FL	$h_{HL, L=\hat{L}}$	99.4	2.310	0.180	99.8	1.940	0.121
Opt	$\kappa_{UMSE, L=\hat{L}}$	98.4	2.380	1.000	98.9	2.010	1.000
Opt	$\kappa_{HL, L=\hat{L}}$	99.4	2.250	0.244	99.8	1.890	0.237
LL-FL	$h_{UMSE, L=2}$	94.9	1.070	0.189	19.0	0.548	0.189
LL-FL	$h_{HL, L=2}$	95.2	1.000	0.251	0.03	0.515	0.251
Opt	$\kappa_{UMSE, L=2}$	95.0	1.060	1.000	20.4	0.546	1.000
Opt	$\kappa_{HL, L=2}$	95.2	1.000	0.242	0.04	0.513	0.242
LL-FL	$h_{UMSE, L=6}$	99.4	2.100	0.125	94.9	1.080	0.125
LL-FL	$h_{HL, L=6}$	100.0	1.960	0.163	94.9	1.010	0.163
Opt	$\kappa_{UMSE, L=6}$	99.3	2.080	1.000	95.0	1.070	1.000
Opt	$\kappa_{HL, L=6}$	100.0	1.950	0.237	94.9	1.000	0.237

*Note:* Empirical coverage rate (Cov.), average length relative to optimized interval with true smoothness bound (RL) and tuning parameter choice ( $h/\kappa$ ) of conventional (Conv.), undersmoothed (US), robust bias-corrected (RBC), fixed length (LL-FL) and optimized linear (Opt) 95% confidence intervals over 20,000 Monte Carlo draws. The pilot RBC bandwidth is reported in parentheses if it is selected separately from  $h$





## Chapter 2

# Social Mobility in Germany

### Abstract

This chapter proposes and implements a strategy for the measurement of social mobility in Germany that allows for the estimation of mobility statistics using census data. The reported mobility statistics characterize intergenerational social mobility in Germany by the association between a child's probability of obtaining an A-Level degree, an important educational qualification in the German institutional framework, and its parents' position in the national income distribution. We document that a 10 percentile increase in the parental income rank is associated with a 5.2 percentage point increase in the probability of obtaining an A-Level degree. This parental income gradient has not changed for the birth cohorts of 1980-1996, despite a large-scale policy of expanding upper secondary education. We document substantial variation in our mobility estimates across regions and show that the estimated regional disparities are unlikely to be driven by sorting of households.

## 2.1 Introduction

Social mobility is an important indicator for both fairness and economic efficiency in a society, as it captures the extent to which individuals of different socioeconomic backgrounds are offered equal economic and social opportunities. Next to violating widely held fairness ideals, a low level of social mobility indicates the misallocation of resources, as it suggests that talented individuals from disadvantaged backgrounds are impeded from realizing their potential.

While, in many European countries, social mobility concerns and equality of opportunity principles feature prominently in the public debate, and political preferences on important issues such as redistributive policies are documented to be driven by beliefs about intergenerational mobility (Alesina et al., 2018), reliable mobility statistics are often not available.

The scarcity of empirical evidence on social mobility can be explained by the relatively high data requirements necessary for the construction of mobility statistics, which require representative data that allows to link the outcomes of parents to a measure of opportunities for children. While the household panel studies that are available in many countries often contain this information, they are typically too small to deliver sufficiently precise estimates to facilitate regional comparisons or the analysis of time trends (cf. Lee and Solon, 2009; Mazumder, 2018). An attractive alternative are large-scale administrative data sources, such as linked tax records, which researchers were able to obtain for some countries (e.g. Chetty et al., 2014). However, such data is not available for many countries, including Germany, where to date no large-scale empirical study of social mobility across time and space exists.

With the aim to fill this gap, this paper proposes and implements a measurement strategy for social mobility in Germany that allows for the use of census data to document time trends and regional differences in social mobility at a higher level of detail than previously possible. Motivated by Germany's early tracking system of secondary education, which allocates children into different tracks at the end of primary school, our mobility statistics measure the association between parental income and the educational opportunities of children. Our measure of opportunities captures whether a child will obtain the A-Level (Abitur), the highest secondary schooling degree in Germany which grants direct access to the tuition-free national university system and marks an important sign of social distinction in the German society. Since secondary school aged children and adolescents typically still live in their parental household, we are able to link them to their parents in the German census data. Our data covers one percent of the German population in every year from 1997 to 2018, providing detailed information on the educational activities of 526,000 children and the socioeconomic status of their parents.

We present three main findings. First, relative social mobility at the national level has remained constant for recent birth cohorts. On average, a 10 percentile rank increase in the parental income distribution was associated with a 5.2 percentage point increase in the probability to obtain an A-Level degree, corresponding to a top-bottom gap of approximately 50 percentage points. For the 1980-1996 birth cohorts, this parental income gradient has not changed despite the *Bildungsexpansion*, a large-scale expansion of upper secondary education in Germany. This long-term expansion was in parts a policy response to a public debate on social mobility (cf. Dahrendorf, 1965; Hadjar and Becker, 2006) and increased the A-Level share from 39% for children born in 1980 to 53% for the 1996 birth cohort. We document that the *Bildungsexpansion* took place uniformly across the income distribution, with almost identical increases in the share of A-Level educated children in all quintiles of the national parental income distribution. This enhanced the odds ratio for disadvantaged children but left the parental income gradient unaffected. The same pattern emerges when estimating mobility trends for subpopulations often emphasized in the public debate on social mobility, such as children in single parent households or children of parents with low levels of formal education.

Second, we document substantial geographical variation in our social mobility measures across German states, cities and local labor markets. For example, the top-bottom gap in the probability of attaining an A-Level degree is approximately 20 percentage points larger in Bremen than in Hamburg, two city states in north-west Germany approximately 100 kilometers apart. Interestingly, we also find significant and meaningful differences within states. For example, the top-bottom gap is approximately 8 percentage points larger in Cologne than in Duesseldorf, two large cities in North Rhine-Westphalia located approximately 40 kilometers apart. Similarly, the share of children obtaining an A-Level from the bottom quintile of the income distribution is 8 percentage points smaller in Nuremberg than in Munich, two large cities in Bavaria. We consider this noteworthy, as education policies, which the prior literature has suspected to be a key determinant of mobility, vary mainly at the state level in Germany.

Third, we show that observable household characteristics are not suited to explain the variation in mobility measures across local labor markets. This is important as, abstracting from estimation uncertainty, differences in our mobility measures can arise either due to structural differences between places or due to systematic sorting of different households into different local labor markets. Which answer prevails has important implications for the usefulness of place-based policies intended to promote social mobility, a topic of ongoing debate in the academic literature. The census data employed in this paper contains rich information on the structure and characteristics of households, allowing us to directly test the importance of sorting

by conditioning on an extensive set of household characteristics. We find that the mobility ranking between local labor markets is largely unchanged when conditioning on household characteristics, indicating that sorting is unlikely to explain the observed regional differences.

In addition, we provide mobility statistics for population subgroups and explore which regional characteristics are most predictive of our mobility estimates. We find that parental education is highly predictive of the educational opportunities of children, highlighting the fact that the interpretability advantages of income based measures of parental background come at the cost of missing information that could be used to characterize social mobility more comprehensively. In our prediction exercise, we find that local labor market conditions, social characteristics and the quality of local schools are best suited to predict our mobility statistics.

The remainder of this paper is structured as follows. Section 2.2 discusses the related literature and relevant aspects of the German institutional framework. In Section 2.3, we describe our data and measurement strategy. Section 2.4 presents our results at the national level. Our regional estimates, including our analysis of local labor markets, are presented in Section 2.5. Section 2.6 concludes.

## **2.2 Related Literature and Institutional Background**

### **2.2.1 Related Literature**

The study of social mobility has a long tradition in economics, sociology and educational research. Across disciplines, efforts to understand and describe the association between the opportunities of children and their parents' socioeconomic status have been made. Since opportunities are difficult to measure, empirical studies of social mobility have generally aimed at the joint distribution of outcomes, with different disciplines emphasizing different outcomes. While early sociological studies focused on occupational transitions between generations, educational research studied intergenerational correlations in educational attainment. In economics, the most common measure of social mobility is the intergenerational elasticity of (lifetime) earnings (IGE), which can be derived from standard intergenerational life-cycle models of human capital accumulation (cf. [Becker and Tomes, 1986](#); [Zimmerman, 1992](#); [Solon, 1992](#); [Mazumder, 2005](#)). Since estimates of the IGE are sensitive to non-linearities and measurement issues at the bottom of the income distribution, recent empirical work relies on rank-rank correlations in lifetime income ([Dahl and DeLeire, 2008](#); [Chetty et al., 2014](#)) to produce more robust mobility statistics. A major step forward in terms of data quality has been achieved by [Chetty et al. \(2014\)](#), who were able to obtain linked administrative tax records. The sample size of this study

allowed the authors to produce reliable estimates of rank-rank correlations across regions in the US, opening up the field for new strategies aimed at understanding the causes of social mobility (cf. [Chetty and Hendren, 2018a](#); [Chetty and Hendren, 2018b](#)). This approach was recently replicated for other countries, including Italy ([Acciari et al., 2019](#)), Switzerland ([Chuard and Grassi, 2020](#)), Canada ([Corak, 2020](#)) and Australia ([Deutscher and Mazumder, 2020](#)).

While measures of social mobility based on the joint distribution of lifetime incomes are attractive, as they allow for easy cross-country comparisons and have a natural interpretation in terms of consumption, they have important limitations. First, since they rely on estimates of children's lifetime income, they are only feasible for relatively old birth cohorts, as reliable estimates of lifetime income require data on children's earnings in the age range 30-40. Thus, such measures are not suited to investigate relatively recent developments in social mobility. Second, since individuals value non-monetary qualities of jobs (cf. [Kalleberg, 1977](#); [Mottaz, 1985](#); [Kalleberg, 2011](#)) and parental income is documented to be positively associated with the non-monetary compensation that children receive from their work (cf. [Boar and Lashkari, 2021](#)), measures based on the joint distribution of incomes may overestimate the degree of inter-generational mobility. Finally, large-scale linked tax data are not available in many countries. Thus, in such countries, mobility measures based on the joint distribution of incomes can only be estimated with sufficient precision at the aggregate level, preventing further analysis of time and geographic variation that is possibly informative about the determinants of social mobility.

For example, in Germany, it is not possible to link individual tax returns. For that reason, most empirical evidence on income mobility is based on the German Socio Economic Panel (GSOEP), the German counterpart of the Panel Study of Income Dynamics (PSID). Like the PSID, the GSOEP provides detailed information about child outcomes and parental background, but suffers from a small sample size. In the GSOEP it is therefore not possible to document time trends or more fine-grained geographical variation in social mobility with a sufficient degree of statistical confidence. [Schnitzlein \(2016\)](#) shows that existing estimates of the national IGE based on the GSOEP are sensitive to small variations in sampling criteria, resulting in a wide range of plausible estimates. It is therefore not surprising that the empirical evidence regarding the level of social mobility in Germany is mixed. Studies that investigate intergenerational income mobility in the GSOEP include [Eisenhauer and Pfeiffer \(2008\)](#), [Riphahn and Heineck \(2009\)](#), [Eberharter \(2013\)](#) and [Bratberg et al. \(2017\)](#). These studies typically find a higher level of income mobility in Germany as compared to the US, and lower levels of mobility in East than in West Germany, albeit with high statistical uncertainty.

Our measurement approach circumvents the data requirements imposed by life-cycle bias concerns by focusing on children’s educational opportunities, while retaining the interpretability advantages of income based measures of parental socioeconomic status. This allows us to draw on the German census data, providing us with the statistical power necessary to conduct a more comprehensive study of social mobility in Germany.<sup>1 2</sup>

A similar approach was followed by Hilger (2015) for the US, who reports mobility statistics based on census data that measure the association between children’s years of schooling and parental income. However, while we also rely on the co-residency of children and their parents, the outcome studied in Hilger (2015) manifests much later in life, when most children have already left the parental household, exacerbating sample selection concerns that necessitate an imputation procedure. Focusing on the years of schooling is necessary in the US context, as almost all children attend academic high school programs. In contrast, the German system of secondary education is separated in an academic and a vocational track, making it better suited for a census based analysis of social mobility as we outline below.

## 2.2.2 Institutional Background

The salient feature of Germany’s system of secondary education is early age tracking, where only the successful completion of the highest track results in the award of an A-Level degree (Abitur), which grants direct access to the tuition-free national university system.

After finishing the four-year elementary school around the age of 10, children are allocated into one of three tracks.<sup>3</sup> While the highest track, the Gymnasium (grades 5-12/13), provides general academic education that aims to prepare children for college, the lower two tracks (grades 5–9/10) provide vocational training with a focus on preparing students for an apprenticeship. In contrast to the US, where only a small share of students attends vocational schools, around 50% of children in Germany are enrolled in a vocational program. The rigor of the tracking system is mediated by the possibility to switch tracks and academic components in the curricula of vocational programs. In particular, it is common that academically talented students from the vocational track switch to the general high track or attend a specialized high track after they finish their vocational degree when they are around 16 years old.

---

<sup>1</sup>A less comprehensive version of the German Census data has previously been used to document differences in the intergenerational correlation in educational attainment between East and West Germany (cf. Riphahn and Trübswetter (2013) and Klein et al. (2019)).

<sup>2</sup>The idea to rely on educational outcomes of children that can be measured early in life has recently been popularized in a small but growing literature on social mobility in developing countries (cf. Alesina et al., 2021; Asher et al., 2020 Muñoz, 2021).

<sup>3</sup>In the states of Berlin and Brandenburg, elementary school lasts six years. The number of vocational tracks varies over time and between states, with some states having adopted a single vocational track design.

Since the early educational careers of children have important consequences for the choices available to them at later stages, and early track choices are heavily influenced by parental characteristics (cf. [Dustmann, 2004](#)), the German institutional framework<sup>4</sup> is particularly suited to study social mobility through the lens of educational opportunities. The importance of track choices for social mobility is reinforced by the fact that almost all primary and secondary schools as well as universities are state-funded, mostly based on student headcounts, resulting in a comparatively large equality in the endowments and quality between different schools and universities. While the exact implementation of the education system can vary across regions, as the responsibility for the education system falls under the jurisdiction of the 16 German states, there are only minor differences in the state provided financing and there are no legal differences between the educational qualifications obtained in different states. In particular, the Standing Conference of State Education Secretaries has the stated goal to ensure a high degree of comparability of educational qualifications across German states and there are no legal differences between the A-Level degrees issued from different states.

The tracking system of secondary education is complemented by a system of dual education that combines apprenticeships in a company and vocational education, integrating school-based learning with company-based practice. The apprenticeships are standardized across the country, preventing large disparities by school or company, and offer occupation-specific practical knowledge and training of skills that allow children to directly enter a specific occupation. Apprenticeships offer a popular pathway into the labor market, and many young adults choose to enter vocational programs despite being eligible for further study. While typically not a strict requirement, the A-Level constitutes a beneficial factor for obtaining vocational training in many popular white-collar occupations ([Klein et al., 2019](#)).

In consequence, the A-Level degree is by far the most important qualification in the German education system and individuals who obtain it enjoy substantially above-average economic outcomes. Using data on full-time workers aged 30-45, we find an A-Level wage premium of 42% for monthly net income.<sup>5</sup> This estimate mirrors [Schmillen and Stüber \(2014\)](#) who report a 44% A-Level wage premium for total gross lifetime earnings. An A-Level degree is also associated with a lower risk of being unemployed ([Hausner et al., 2015](#)) and a higher life expectancy ([Gärtner, 2002](#)) and marks an important sign of social distinction in the German

---

<sup>4</sup>A more detailed description of the German tracking system and track switching in Germany can be found in [Biewen and Tapalaga \(2017\)](#) and [Dustmann et al. \(2017\)](#).

<sup>5</sup>The A-Level wage premium is estimated using the waves 1997-2018 of the German Mikrocensus (described in the next section) by regressing the log of net monthly personal income of full-time workers aged 30-45 on an A-Level indicator, as well as a set of age and year indicators to implicitly account for job experience.



society. Overall, this illustrates that, for children in Germany, the A-Level degree is a compelling measure for their social and economic opportunities.

## 2.3 Data and Measurement Strategy

Our analysis is based on data of the German Microcensus (Mikrozensus, hereafter MZ), a large-scale annual representative survey of the German population administered by the Federal Statistical Office of Germany. The MZ is comparable to, but more detailed than, the American Community Survey and constitutes the largest survey program of official statistics in Europe. The survey was first administered in West Germany in 1957 and includes East Germany since 1991. It contains individual level data on a wide range of topics, including family status and linkage within the household, citizenship, labor market participation, income as well as information on educational activities and attainment for all members of the sampled households.

The MZ survey has several features that make it particularly suited for our research question. First, it allows us to reliably match children to their parents as long as they are still registered at their parents' household. This is because, by law, it is compulsory for individuals living in Germany to register at their household and the sampled households are obliged to provide information on every person registered at their respective household. Second, it contains fine-grained information on the location of households and is sufficiently large to permit the estimation of mobility statistics for single cohorts and regions. Moreover, since its inception, the survey was continuously improved and its institutional embeddedness and design offer several advantages over comparable national surveys as we outline below.

**Sampling Design.** Each year, a randomly selected 1% sample of the persons and households in Germany is asked to participate in the survey. By law, participation is mandatory for members of the selected households, which remain in the survey for at most four subsequent years. The primary sampling units consist of clusters of neighbouring buildings and all households belonging to a sampled cluster are interviewed. The non-response rate is approximately 3%.<sup>6</sup> Each year, one quarter of the initially sampled clusters are replaced by new clusters, resulting in partial overlap of sampling units. The detailed nature of the questionnaire together with the low non-response rate and partial panel dimension of the data allow us to mitigate measurement and sample selection concerns often brought forward in the context of survey data.

---

<sup>6</sup>The non-response rate is driven by households that could not be reached and residents in shared accommodations (cf. [Statistisches Bundesamt, 2018](#)), which we exclude from our sample. Appendix 2.A contains additional information on the survey and sampling design of the MZ.



### 2.3.1 Variable Definitions

**Measuring Opportunities of Children.** Motivated by the importance of the A-Level degree for children’s future educational and labor market opportunities in the German institutional framework, we measure opportunities by a binary variable  $Y_i$  that is equal to one if a child has obtained, or is on track to obtain, a degree that is equivalent to an A-Level, and zero otherwise. Specifically, our outcome variable is equal to one if (i) a child has obtained a degree that qualifies for tertiary education<sup>7</sup> or if (ii) a child is enrolled in the last 2-3 years of a track which leads to such a degree at the successful completion of school.<sup>8</sup> In the following, we refer to this outcome as an A-Level degree and characterize intergenerational mobility in terms of the conditional probabilities of obtaining an A-Level degree for children of different parental backgrounds.

Our outcome definition takes into account three considerations. First, while the MZ survey is conducted on a rolling basis, A-Level degrees are typically awarded in the second quarter of the calendar year. Back of the envelope calculations suggest that, if we only count children who have already obtained an A-Level degree, we would miss-measure our outcome for around 40% of the graduating cohort in each survey year. Second, since the share of children failing the final examination in a given year is low<sup>9</sup>, including upper stage students allows us to capture children that can reasonably be expected to obtain an A-Level degree but rotate out of the survey before they do so. Finally, including younger children disproportionately increases our sample size, as younger children are more likely to live with their parents. Table 2.1 displays the share of children living with at least one parent by age of the child, calculated from our data. In our data, virtually all children younger than 15 still live with at least one parent. However, the share of children co-residing with their parents is decreasing with child age with steep drops after the legal age of 18. While 92% of 18 year olds are living with at least one of their parents, this fraction drops to 44% for individuals at the age of 23. In Section 2.3.3, we discuss how the co-residency and move-out patterns observed in the MZ data affect the interpretation of our results.

---

<sup>7</sup>We classify educational qualifications as equivalent to an A-Level if they grant access to the tuition-free national university system. This includes *Allgemeine Hochschulreife (Abitur)*, *Fachgebundene Hochschulreife* and *Fachhochschulreife*.

<sup>8</sup>The MZ data contains information on the type of school and grade level attended by all sampled children. Our definition subsumes all students on *Allgemeinbildende Schulen* enrolled in the *Gymnasiale Oberstufe* as well as students from specialized tracks like *Berufliches Gymnasium* or *Fachoberschule* which award an A-Level degree.

<sup>9</sup>The national average failure rate is approximately 3 percent on average for the years 2010-2020. For an overview of the share of children failing the final examination see <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html>.

Table 2.1: Co-Residence Rate by Child Age

Child Age	15	16	17	18	19	20	21	22	23
Share Living with Parents	0.99	0.98	0.97	0.92	0.84	0.72	0.62	0.52	0.44

*Notes:* This table reports the fraction of individuals which live in the same household as at least one of their parents in the MZ waves 1997 to 2018 by age at observation.

**Measuring Parental Background.** We measure parental background by a household’s self-reported monthly net income, excluding the income of all dependent children. Our income measure covers all sources of income, including labor income, business profits and social security transfers. To account for differences in the structure of costs encountered by households of different compositions, we scale all household incomes by the OECD-modified equivalence scale.<sup>10</sup> Following [Dahl and DeLeire \(2008\)](#) and [Chetty et al. \(2014\)](#), we then compute the households’ percentile ranks in the sample distribution<sup>11</sup> of equivalized household income, and assign each child the rank of their respective household, which we refer to as the parental income rank  $R_i$ .

We emphasize that our aim is *not* to estimate some causal effect of relative parental income on children’s educational attainment. Instead, our measure intends to capture relative advantages in family circumstances of some children relative to others in a fashion that allows for the construction of robust and easy to interpret mobility statistics. To that end, parental income ranks are conceptually attractive, as the relevance of financial resources and costly enrichment activities for different aspects of child development is widely recognized and there exists empirical evidence of significant disparities in child-related expenditures across the income distribution in Germany. [Table 2.2](#) reports estimates of monthly child-related expenditures in different categories based on data of the 2018 Income and Consumption Survey (EVS) for dual parent households with single children in the top and bottom decile of the national income distribution. The estimates reveal substantial gaps in monthly expenditures on child-enrichment activities in categories such as education, health as well as culture and leisure activities, suggesting that parental income ranks are a suitable measure of parental background for the construction of mobility statistics in Germany. Our approach of computing parental income ranks based on the sample distribution of equivalized household income reflects the idea

<sup>10</sup>In [Appendix 2.B](#), we show that the choice of the scaling factor is not influential for our results at the aggregate level. However, the quality of the linear approximation to the empirical CEF of our A-Level indicator conditional on parental income ranks is improved when computing ranks based on equivalized incomes.

<sup>11</sup>In [Appendix 2.B.1](#), we provide additional information on the sample income distributions and details on the construction of our rank variable.

that we want to capture a household’s available resources for child-related expenditures.

Table 2.2: Monthly Child-related Expenditures of Single Child Households

Category	Total	Education	Health	Food	Culture	Mobility	Other
Top Decile	1212	83	113	156	205	85	244
Bottom Decile	424	28	11	104	47	29	65
Ratio	2.85	2.96	10.27	1.5	4.36	2.93	3.75

*Notes:* This table reports estimates of the monthly child-related expenditures in Euro of dual parent, single child households in the top and bottom decile of the German national income distribution for different expenditure categories. The data is reported in the 2018 Income and Consumption Survey (EVS) of the German Federal Statistical Agency ([Statistisches Bundesamt, 2021](#)).

The measure of household income provided in the MZ data that we use to compute the parental income ranks has two important limitations. First, it is not asked directly in the survey, but imputed by the Statistical Office. The survey respondents report their personal income in 24 predefined bins, as well as their household’s total net income. While the top bin is defined only by a lower bound, all other bins are defined by half open intervals. The bins were updated over time to account for changes in the income distribution such that less than 1% of all households fall into the top bin. The right-skewed shape of the income distribution is reflected by increasing bin-widths ([Statistisches Bundesamt, 2019](#)). The Statistical Office transforms the personal binned income into a continuous variable, essentially randomizing individuals uniformly within each bin. In a second step, these values are summed up to a continuous measure of household income. Our rank computations are based on an equalized version of the imputed household income. We divide the household’s total income by the sum of weights assigned to each member of a household. The first person in each household is assigned a weight of 1. Each additional member above the age of 14 is assigned a weight of 0.5. Children below the age of 14 are assigned a weight of 0.3, yielding a measure of income that takes into account differences in the household’s size and composition. Second, since the underlying bin data is self-reported, it is likely subject to misreporting. As a consequence, our rank computations are based on a measure of household income that is subject to two types of measurement error. While the measurement error induced by the imputation procedure is independent of household characteristics and thus has well understood statistical implications, we can not rule out that the errors due to self-reporting are systematically related to household attributes, with less obvious implications for the statistical properties of our mobility statistics. In Section 2.3.3, we discuss to what extent we can address and mitigate these measurement issues.

### 2.3.2 Mobility Statistics

The central building block of all mobility statistics reported in this paper are estimates of the conditional probability of attaining an A-Level degree for children with parental income rank in a given set  $A$ ,  $E[Y_i|R_i \in A]$ . Consequently, all estimands are descriptive, in the sense that all uncertainty about our mobility statistics stems only from the fact that we do not observe the full population of Germany.

Following the recent literature, we define two sets of mobility statistics with the aim to distinguish between two mobility concepts: absolute and relative mobility. While measures of absolute mobility are informative about the level of opportunities for disadvantaged children, relative mobility measures seek to capture differences in opportunities between children of disadvantaged backgrounds relative to those of more advantaged backgrounds.

**Absolute Mobility.** Our preferred measure of absolute mobility is the probability of obtaining an A-Level degree for a child from the bottom quintile of the parental income distribution:

$$Q1 = E(Y_i|R_i \leq 20). \quad (2.1)$$

We refer to this estimand as the  $Q1$  measure and estimate it by its sample analogue  $\overline{Q1}$ . A high value of the  $Q1$  measure implies high absolute mobility, as it indicates that a large share of disadvantaged children are eligible to enter the university system.

**Relative Mobility.** In contrast, relative mobility measures are concerned with differences in opportunities between children from low-income families relative to children from high-income families. A simple measure of relative mobility is the  $Q5/Q1$  ratio:

$$Q5/Q1 = \frac{E(Y_i|R_i > 80)}{E(Y_i|R_i \leq 20)}, \quad (2.2)$$

which captures the odds ratio of obtaining an A-Level degree for children from the top quintile relative to those in the bottom quintile of the parental income distribution. A high value of the  $Q5/Q1$  ratio implies low relative mobility. For example, a ratio of  $Q5/Q1 = 2$  means that children from the top quintile of the income distribution are twice as likely to obtain an A-Level degree as children from the bottom quintile of the income distribution. We estimate the  $Q5/Q1$  ratio by its sample analogue  $\overline{Q5}/\overline{Q1}$ .

**Parametric Mobility Statistics.** In order to produce precise estimates of the conditional probabilities underlying the above mobility statistics using sample analogues, it is necessary to observe sufficiently many children of parents in the relevant regions of the income distribution. While this is not a concern for estimates at the national or state level, where sufficiently many observations are available, producing precise estimates in smaller subsets of the data, such as single cities or regions, is challenging.

An important feature of our data that we rely on to produce informative estimates for such partitions is that the empirical conditional expectation function,  $\widehat{E}[Y_i|R_i]$ , of our outcome given the parental income rank can be well approximated by a linear function in various partitions of our data. As a consequence, we can use a parsimonious parametric model to extrapolate towards regions with few observations and characterize mobility in terms of the model parameters. Formally, we do so by approximating the respective conditional expectation function (CEF) by its best linear predictor, which is defined as

$$\theta = \arg \min_{\theta} E[(Y_i - Z_i'\theta)^2],$$

with  $Z_i = (1, R_i)'$  and  $\theta = (\alpha, \beta)$ . In practise, we estimate the model parameters by running an OLS regression of our outcome indicator on the parental income rank variable. If the CEF is linear, the model parameters have a natural interpretation in terms of conditional probabilities: The intercept  $\alpha$  corresponds to the probability of attaining an A-Level degree for children from the bottom of the income distribution, while the slope coefficient  $\beta$  measures the gap in the probability of obtaining an A-Level between the children at the top and the bottom of the income distribution. As such, they represent meaningful measures of absolute and relative mobility. We refer to the slope coefficient as the *parental income gradient* and report estimates of  $\beta \times 100$ , which captures the gap in percentage points, for improved readability.

While the "true" CEFs are unlikely to be exactly linear, our parametric mobility statistics provide meaningful approximations that can be estimated with sufficient precision to allow for meaningful comparisons. In the results section of this paper, we provide evidence that the empirical CEF is close to linear in many different partitions of the data, which, in our view, together with complimentary evidence, lends credence to the parametric mobility statistics we report for our regional analysis in Section 2.5, where we rely heavily on the parental income gradient to characterize regional relative mobility. In Section 2.3.3, we explain how we quantify the uncertainty associated with our estimates.

### 2.3.3 Sample Definition and Limitations

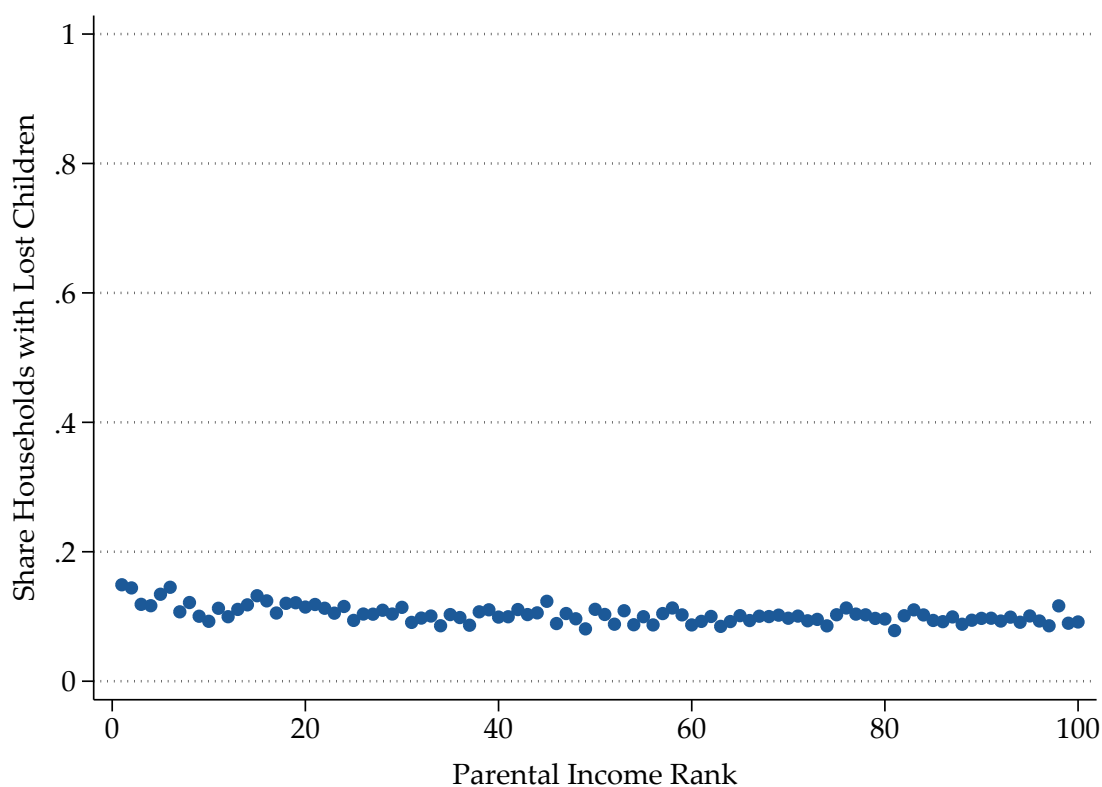
In our analysis, we use the MZ survey waves from 1997 to 2018, for which a consistent definition of all relevant variables is available.<sup>12</sup> Our primary sample contains all sampled children aged 17 to 21 which are observed in the same single-family household as at least one of their parents. The age range is chosen to balance the following trade-off: For older children, our outcome is measured more precisely, i.e. we do not need to rely on upper-stage enrollment but are more likely to observe the completed degree. At the same time, the fraction of children in our sample that has already moved out of the parental household, and thus can not be matched to their parents, increases with age, which guides our choice for the upper bound. The lower bound is chosen to account for track switching, as children enrolled in the final stage of an A-Level track are typically at least 17 years old.

**Sample Selection.** An immediate concern caused by the observed move-out patterns in the MZ data relates to the representativeness of our sample. If the observed move-out decisions were systematically related to parental income and the educational attainment of children, the external validity of our estimates would be undermined, in the sense that our statistics would not measure social mobility in the general population of interest. An advantage of the MZ survey is that the partial panel dimension of the data set allows us to investigate the empirical plausibility of such concerns. While we acknowledge that dependencies of this type are generally plausible, we do not find evidence of sample selection in our data. Figure 2.1 displays the share of observed move-outs of children by parental income rank for the subsample of households in our data that is observed in the survey in multiple years. It shows that move-outs occur uniformly across the income distribution and are thus uncorrelated with the parental income rank. This can also be seen in Table 2.3, which documents how time-constant characteristics of the children in our sample change with the age at observation. The average parental income and the associated income rank of children in the age range 17-21 are essentially constant. Over the full age-range, we find that females and children of parents with an A-Level degree tend to move out at earlier ages. While this exercise suggests that sample selection is not a major concern for our analysis, we also demonstrate in the next section that choosing alternative age ranges barely affects our results at the aggregate level.

---

<sup>12</sup>For our national and regional estimates, we restrict our sample to the survey waves 2011-2018 (230,000 children) to produce recent mobility statistics and avoid ambiguities caused by a series of administrative reforms that changed county boundaries. The mobility statistics by birth cohort reported in Section 2.4.2 are computed based on the observed children of the 1980-1996 birth cohorts (526,000 children).

Figure 2.1: Move-out Frequency by Parental Income Rank



*Notes:* This figure shows the relative frequency of move-outs of children younger than 21 by parental income rank. It is computed using the subset of households observed in the survey more than once.

Table 2.3: Average Characteristics of Children by Age at Observation

Child Age	Share Female	Mean Parental Income (Equiv.)	Parental Income Rank	Share Parents with A-Level
13	0.49	1153	47	0.35
14	0.49	1127	45	0.34
15	0.49	1167	47	0.34
16	0.49	1161	46	0.33
17	0.49	1244	50	0.33
18	0.48	1245	50	0.32
19	0.47	1245	50	0.32
20	0.44	1239	50	0.31
21	0.42	1243	50	0.31
22	0.41	1162	46	0.31
23	0.39	1174	46	0.30

*Notes:* This table reports average attributes of children in the MZ waves 1997 to 2018 that are observed in the same household as at least one of their parents by age at observations. The ranks are computed based upon the sample distribution of equalized household income as described in Section 2.3.1.

**Measurement Error.** A second concern relates to the quality of the data underlying our estimates. While the accuracy of the information on educational activities and qualifications underlying our outcome variable is documented (Schimpl-Neimanns, 2006), the information on household income reported in the MZ data is subject to measurement error as explained in Section 2.3.1. Our measurement approach reflects concerns regarding the quality of the MZ income data in that we rely on rank-based measures of social mobility, documented to have favorable bias properties in the presence of measurement errors relative to other approaches (Nybom and Stuhler, 2017). Moreover, we demonstrate that our results do not change when we compute income ranks based on multi-year averages as explained below.

**Life-cycle Bias.** As discussed in Section 2.3.2, our definition of household income does not seek to capture lifetime income but the household resources available for enrichment activities during childhood. In our baseline, we therefore compute parental income ranks based on the earliest observed household income. To the extent that parental incomes fluctuate from year to year due to transitory income shocks, our estimates could overstate mobility. To address this issue, we again exploit the partial panel dimension of our data to compute multi-year averages of parental income before assigning ranks. We demonstrate below that our results are insensitive to this procedure. While such fluctuations could affect our rank computations, our education-based measure of opportunities does not suffer from life-cycle biases. In contrast to traditional measures that rely on the labor market incomes of children, such as the IGE, we can therefore study recent birth cohorts without compromising the quality of our estimates.

**Standard Errors.** The standard errors reported alongside our estimates in the results section of this paper abstract from the fact that we estimate the cutoffs defining the percentile ranks. For our parametric mobility statistics, we report Liang-Zeger standard errors (Liang and Zeger, 1986) clustered at the level of the sampling district, the primary sampling unit of the MZ. The standard errors reported alongside our estimates of the quintile measures are based on the variance of the sample averages in the respective subsamples, effectively treating our data as a simple random sample. For the  $Q1$  measure, we report the standard error of the corresponding sample average of outcomes. For the  $Q5/Q1$  ratio, we report plug-in standard errors based on a "delta method argument", that is we linearize the ratio of averages which yields the following approximation for the variance of the sampling distribution of the  $\overline{Q5}/\overline{Q1}$  sample ratio

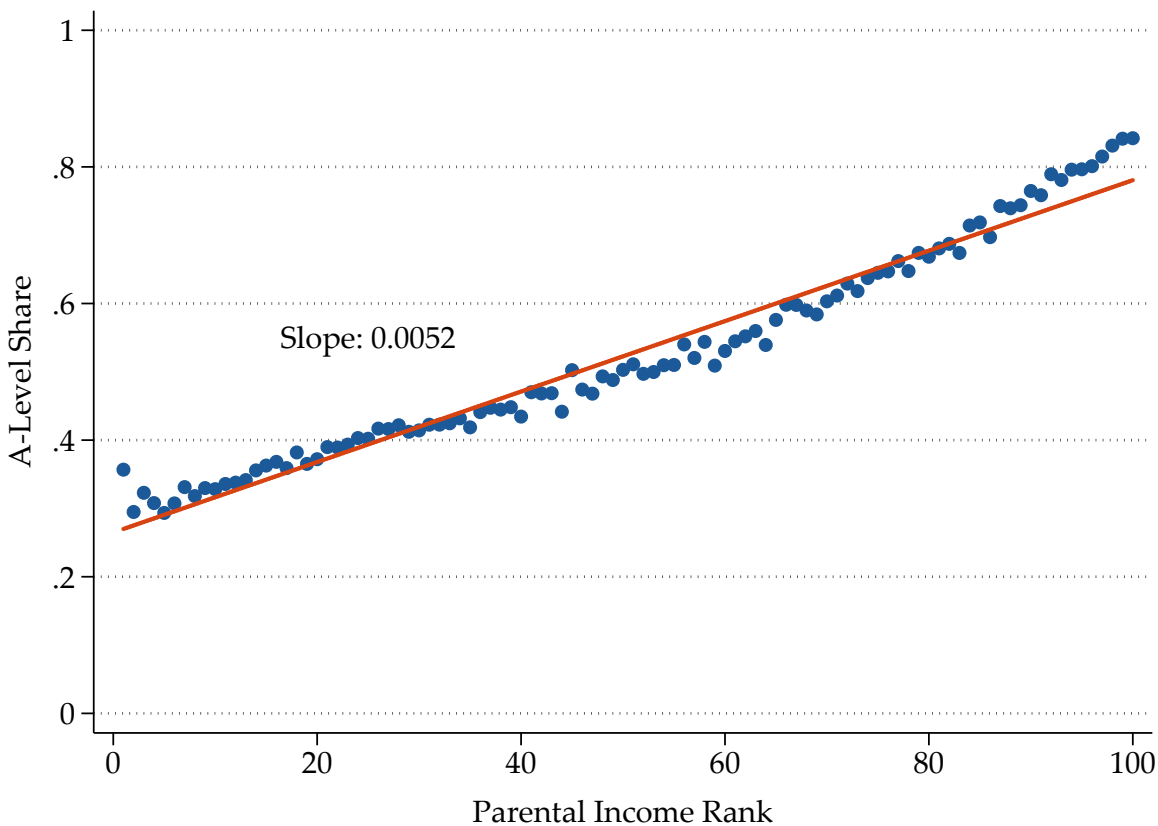
$$V(\overline{Q5}/\overline{Q1}) \approx \frac{1}{(\overline{Q1})^2} \left( V(\overline{Q5}) + \left[ \frac{\overline{Q5}}{\overline{Q1}} \right]^2 V(\overline{Q1}) - 2 \frac{\overline{Q5}}{\overline{Q1}} Cov(\overline{Q5}, \overline{Q1}) \right).$$



## 2.4 National Estimates

We begin our empirical analysis by characterizing social mobility at the national level. Figure 2.2 shows the share of children with an A-Level degree by parental income rank in our data, as well as the best linear approximation to the empirical CEF. As can be seen, a linear model provides a reasonable approximation to the empirical CEF, a regularity that we observe in essentially all considered partitions of our data.

Figure 2.2: Social Mobility at the National Level



*Notes:* This figure shows the fraction of children aged 17-21 that are either enrolled in the upper stage of an A-Level track or have already attained an A-Level degree by the percentile rank of their parents in the national income distribution based on the MZ waves 2011-2018, as well as the best linear approximation to the empirical CEF. The reported slope coefficient of 0.0052 (SE 0.004) is estimated by OLS using the underlying individual level data, with LZ standard errors clustered at the level of the primary sampling unit.

In the national data, we estimate the parental income rank gradient at  $\beta \times 100 = 0.52$ , implying a gap of roughly 50% in the probability of obtaining an A-Level degree between children from the top and the bottom of the income distribution. Our preferred measure of absolute mobility in the national data suggests that roughly one third of the children from the bottom quintile of the income distribution complete an A-Level degree, with Q1 estimated at 0.34. Both,

parametric and nonparametric mobility statistics imply that the odds ratio in the probability of obtaining an A-Level degree between children from the top quintile relative to the bottom quintile is greater than 2, with Q5/Q1 estimated at 2.24. Consistent with our discussion of sample selection concerns and measurement issues in Section 2.3.3, we find that our estimates are robust to variations of the age restriction defining our sample, as well as the computation of income ranks based on multi-year averages of parental income, as shown in Table 2.4.

Table 2.4: National Mobility Estimates

Age	Gradient	Q1	Q5	Q5/Q1	A-Level Share	N
17-21	0.52 (0.004)	0.34 (0.002)	0.76 (0.002)	2.24 (0.007)	0.52	230,972
Averaged	0.52 (0.004)	0.34 (0.003)	0.77 (0.003)	2.26 (0.010)	0.52	230,972
17	0.53 (0.007)	0.31 (0.005)	0.71 (0.004)	2.30 (0.016)	0.49	53,324
18	0.51 (0.007)	0.34 (0.005)	0.76 (0.004)	2.26 (0.014)	0.54	51,278
19	0.51 (0.008)	0.35 (0.005)	0.76 (0.004)	2.20 (0.015)	0.53	46,747
20	0.51 (0.008)	0.35 (0.005)	0.76 (0.005)	2.20 (0.016)	0.53	42,396
21	0.52 (0.008)	0.35 (0.005)	0.76 (0.005)	2.20 (0.017)	0.53	37,227

*Notes:* This table reports our national mobility statistics for different age restrictions with corresponding standard errors in parentheses. The upper panel corresponds to our primary sample. The rows in the lower panel report analogous estimates for samples containing only children of a given age at measurement, as indicated in the first column. The standard errors are computed as described in Section 2.3.3.

Do these estimates depict Germany as a country of high or low relative mobility? While a cross-country comparison of our results is not straightforward, as the German system of upper secondary education and university funding is rather peculiar, we are aware of two US studies which report comparable mobility statistics. Using data from the Census 2000, [Hilger \(2015\)](#) reports a parental income rank gradient of 3.6 percentage points in attending college for children aged 19-21. A higher point estimate is reported in [Chetty et al. \(2014\)](#), who estimate the rank gradient in college enrollment at 6.7 percentage points for children aged 18-21 based on tax registry data. Under the assumption that college enrollment conditional on having obtained an A-Level degree is weakly increasing in parental income ranks, our estimate of 5.2 percentage

points implies a college enrollment gradient that falls into the range of point estimates reported for the US. Abstracting from differences in the distributions of college quality and the selection of students of different parental backgrounds into colleges of different quality, our estimates suggest that educational mobility, as we measure it, in Germany is similar to the US. We consider this finding noteworthy, as (after tax) income inequality is more pronounced in the US than in Germany, suggesting that one should expect steeper rank gradients in the US. Our finding contrasts with cross-country comparisons in relative income-mobility, which typically report higher levels of relative mobility in Germany, highlighting the conceptual difference between income and education based measures of social mobility. Similar results were obtained by Landersø and Heckman (2017), who find that Denmark, a society that is characterized by high levels of income mobility, is similar to the US in terms of measures of educational social mobility.

### 2.4.1 Subgroup Estimates

A natural question to ask is whether our national estimates mask meaningful differences in our mobility measures across subpopulations. Table 2.5 reports our mobility statistics for selected subsamples of our data defined by discrete attributes of children and their households contained in the MZ survey.

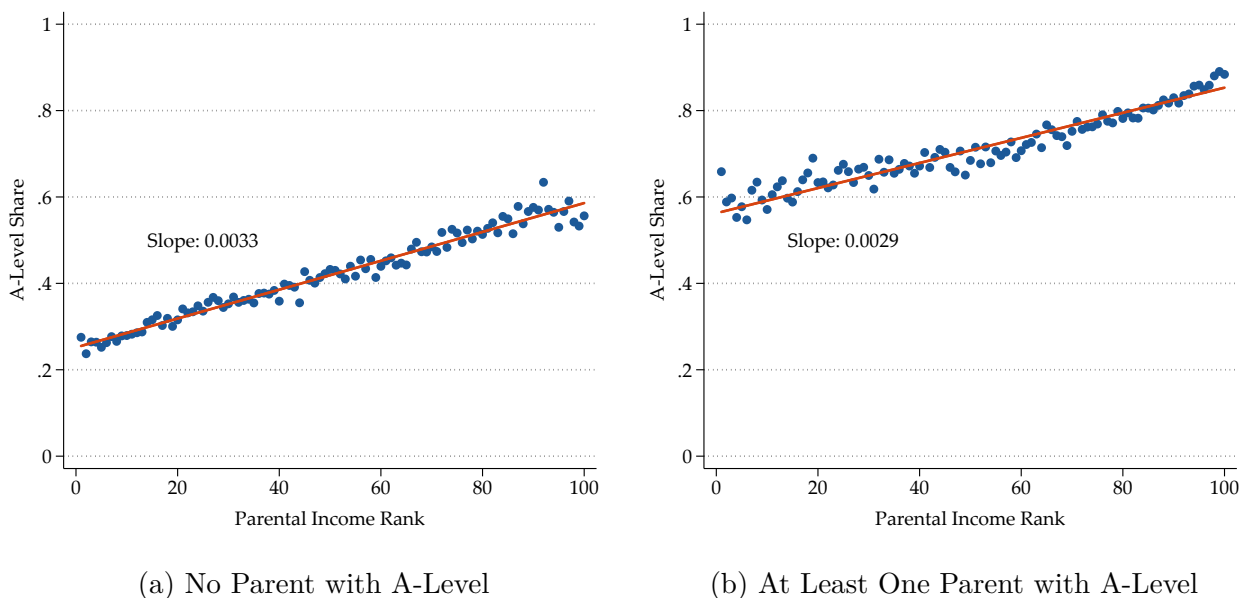
We document a few interesting patterns. Most importantly, we find substantial differences by parental education. Figure 2.3 displays the A-Level share of children by parental income rank and the associated parental income gradient separately for children from households where no parent has an A-Level degree (a) and for children from households where at least one parent has an A-Level degree (b). We find that the A-Level share amongst children of parents without an A-Level degree at the top of the income distribution is comparable to the A-Level share amongst children with at least one A-Level educated parent at the bottom of the income distribution. The conditional rank gradients are attenuated due to the positive correlation between parental education and income ranks with point estimates of approximately 0.3 in both groups, narrowing the top-bottom gap from approximately 50% to around 30%. Roughly speaking, the empirical conditional distribution for children of A-Level educated parents is shifted upwards by approximately 30 percentage points, uniformly across ranks. The intergenerational correlation in A-Level attainment in our data is 0.54. This finding highlights that the interpretability advantages of income-only based measures of parental background come at the cost of missing

Table 2.5: Mobility Statistics for Subgroups

Attribute	Subgroup	Gradient	Q1	Q5	Q5/Q1	A-Level Share	N
<b>Parental Education</b>	No A-Level	0.33 (0.005)	0.28 (0.002)	0.55 (0.005)	1.94 (0.023)	0.39	145,892
	A-Level	0.29 (0.006)	0.61 (0.006)	0.84 (0.002)	1.36 (0.012)	0.75	85,080
<b>Parenting Status</b>	Single Parent	0.50 (0.008)	0.34 (0.003)	0.72 (0.007)	2.13 (0.031)	0.47	50,622
	Two Parents	0.54 (0.004)	0.34 (0.003)	0.76 (0.002)	2.26 (0.021)	0.54	179,715
<b>Parents Married</b>	Not Married	0.47 (0.007)	0.33 (0.003)	0.69 (0.007)	2.11 (0.030)	0.46	51,018
	Married	0.54 (0.007)	0.35 (0.003)	0.77 (0.003)	2.21 (0.024)	0.54	172,999
<b>Gender</b>	Male	0.53 (0.005)	0.29 (0.003)	0.72 (0.003)	2.49 (0.027)	0.47	123,649
	Female	0.50 (0.008)	0.40 (0.003)	0.81 (0.003)	2.02 (0.019)	0.58	107,323
<b>Migration Background</b>	Native	0.54 (0.004)	0.32 (0.002)	0.76 (0.002)	2.35 (0.020)	0.55	164,018
	Migrant	0.48 (0.006)	0.36 (0.003)	0.75 (0.006)	2.11 (0.029)	0.50	60,908
<b>Region</b>	West Germany	0.50 (0.004)	0.34 (0.002)	0.76 (0.002)	2.19 (0.016)	0.52	201,684
	East Germany	0.60 (0.009)	0.40 (0.005)	0.81 (0.006)	2.02 (0.049)	0.51	29,288
<b>Siblings</b>	Yes	0.54 (0.004)	0.35 (0.003)	0.79 (0.003)	2.29 (0.018)	0.52	156,960
	No	0.50 (0.006)	0.32 (0.004)	0.72 (0.003)	2.27 (0.034)	0.52	74,012
<b>Birth Order</b>	First Child	0.51 (0.004)	0.34 (0.003)	0.76 (0.002)	2.22 (0.019)	0.53	165,336
	Second Child	0.52 (0.007)	0.34 (0.004)	0.77 (0.004)	2.27 (0.031)	0.51	56,996
	Later Child	0.57 (0.017)	0.31 (0.008)	0.78 (0.015)	2.48 (0.076)	0.45	8,640

*Notes:* This table reports estimates of our relative and absolute mobility measures for selected groups of children observed in the MZ survey waves 2011-2018, for which a consistent definition of the attributes used to partition the sample is available. Parental marital status indicates if both parents are married. Migration background subsumes all individuals who immigrated to Germany after 1949, as well as all foreigners born in Germany and all individuals born in Germany with at least one parent who immigrated after 1949 or was born in Germany as a foreigner. The standard errors reported in parentheses below each point estimate are computed as described in Section 2.3.3.

Figure 2.3: Empirical Distribution by Parental Education



*Notes:* This figure shows the fraction of children aged 17-21 observed in the MZ survey waves 2011-2018 that are either enrolled in the upper stage of an A-Level track or have already attained an A-Level degree by the parental income rank, separately for children of parents who have not obtained an A-Level degree (a) and children of parents where at least one of the parents has obtained an A-Level degree (b) as well as the corresponding best linear approximations to the empirical CEFs. The reported estimates of the parental income gradient are based on the underlying micro data. Standard errors are reported in the first panel of Table 2.5.

observable attributes of households that could be used to characterize social mobility more comprehensively. The estimates reported in Table 2.5 reveal a few more interesting discrepancies. We find that, at the bottom of the income distribution, females and children of migrants<sup>13</sup> are approximately 11 and 4 percentage points more likely to obtain an A-Level degree than their respective male and native counterparts. While the gender-gap is roughly constant across the income distribution, the difference between migrant and native children vanishes in the top-quintile. Moreover, we document larger income rank gradients for children of married and cohabiting couples, as well as natives and children living in East Germany. The East-West gap in parental income gradients is approximately 0.1, implying a 10 percentage points larger top-bottom gap in the probability of attaining an A-Level degree in East Germany as compared to West Germany. We investigate such regional patterns in our mobility measures in more detail in Section 2.5. Figure 2.B.4 in Appendix 2.B displays the empirical CEF for all subgroups, showing that the linearity assumption underlying our parental income gradient estimates provides a reasonable approximation to our data.

<sup>13</sup>In the MZ, migrants are defined as all individuals who immigrated to Germany after 1949, as well as all foreigners born in Germany and all individuals born in Germany with at least one parent who immigrated after 1949 or was born in Germany as a foreigner.

## 2.4.2 Time Trends

A second interesting question to ask is how our mobility measures have evolved over time. While our descriptive approach does not allow us to attribute changes in mobility measures to specific policies, our measurement strategy enables us to provide novel evidence on the evolution of social mobility in Germany for relatively recent birth cohorts. The period we study is particularly interesting, as it covers the second half of the arguably most significant educational reform in post-war Germany, the "*Bildungsexpansion*", a large-scale policy of expanding upper secondary and higher education that, starting in the early 1970s, increased the A-Level share from around 20% to approximately 50% for the birth cohorts since the mid 1990s. This expansion was a policy response to a heated public debate on social mobility (cf. Dahrendorf, 1965) and the increasing importance of education for economic growth (cf. Picht, 1964) at the time (Hadjar and Becker, 2006). We ask whether the large-scale expansion of upper-secondary education in Germany was accompanied by changes in social mobility as defined by our mobility measures.

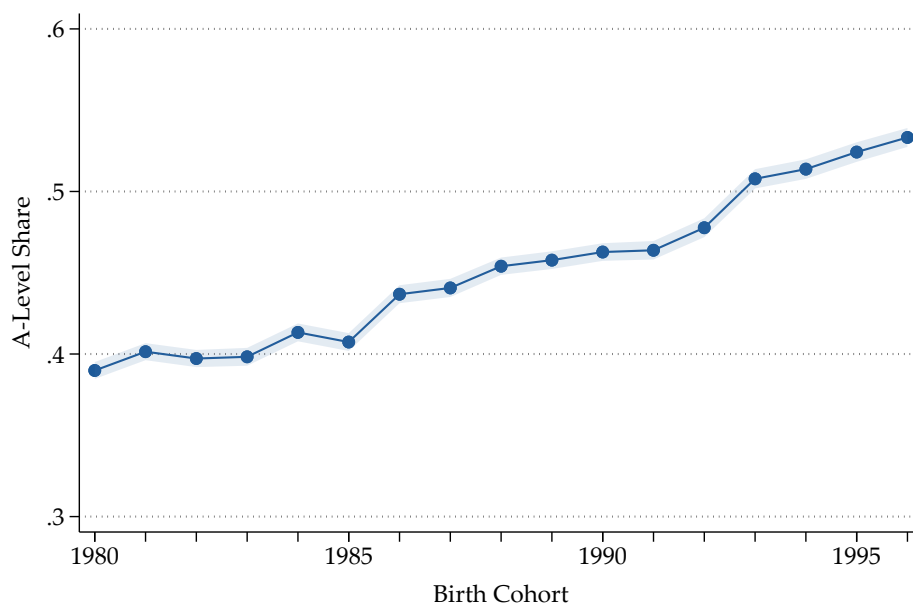
To this end, we focus on a sample of 526,000 children born between 1980-1996.<sup>14</sup> At the time of writing, the children of the respective birth cohorts are 25-40 years old and constitute a significant part of the German working population. Including relatively young cohorts in our analysis is feasible, as our education-based measure of opportunities, in contrast to traditional measures that rely on the labor market incomes of children, does not suffer from life-cycle biases. Figure 2.4 depicts the evolution of the A-Level share amongst 17-21 year old children in the MZ data for the birth cohorts under consideration. For the considered cohorts, the A-Level shares in our data closely mirror those reported in the official school statistics, suggesting that our sample is representative.

As shown in Figure 2.4, our data covers roughly the second half of the expansion, with an observed increase in the A-Level share of around 15% from 39% for the 1980 birth cohort to 53% for children born in 1996. Figures 2.5 and 2.6 display the estimates of our mobility measures for the same cohorts. The figures document that, while the parental income gradient has remained roughly constant at around 0.52, the point estimate that we report at the national level based on more recent data, the odds ratio captured by the Q5/Q1 ratio decreased by approximately one third, from around 3 for the 1980 birth cohort, to slightly above 2 for the 1996 cohort. At the same time, absolute mobility as measured by the Q1 share increased substantially, from approximately 0.22 in 1980 to 0.35 in 1996, which is similar to the share of around one third

---

<sup>14</sup>We restrict our attention to these cohorts to rule out that our estimates are affected by differences in the distribution of age at measurement. For the considered cohorts, the share of 17, 18-, 19-, 20- and 21-year-olds in our data is constant.

Figure 2.4: A-Level Share by Cohort 1980-1996

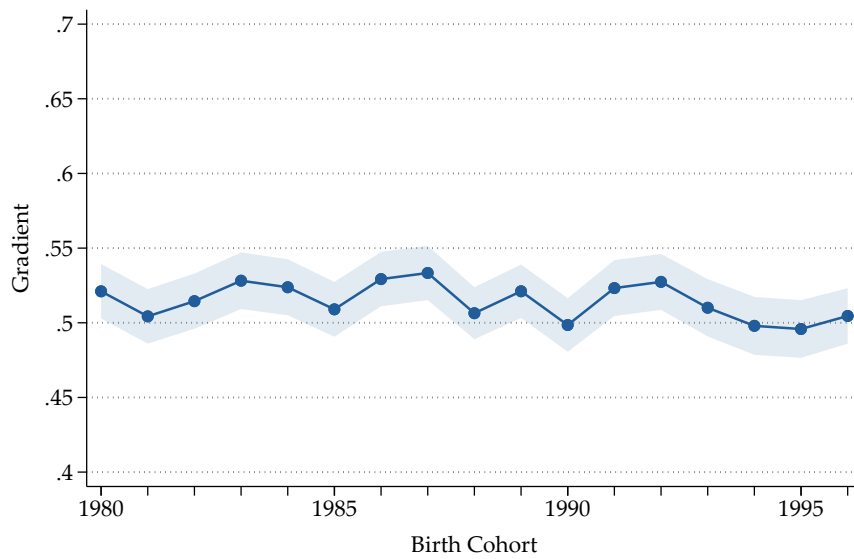


*Notes:* This figure shows the fraction of children born between 1980 and 1996 and observed at ages 17-21 that are either enrolled in the upper stage of an A-Level track or attained an A-Level degree in the MZ data. The shaded area depicts pointwise 95% confidence intervals computed as described in Section 2.3.3.

for children from the bottom quintile of the income distribution on track to an A-Level degree reported in Section 2.4.1 based on more recent cohorts. The same overall pattern emerges when estimating mobility trends by the subgroups studied in Section 2.4.1 as we report in Figures 2.B.6 and 2.B.5 in the Appendix.

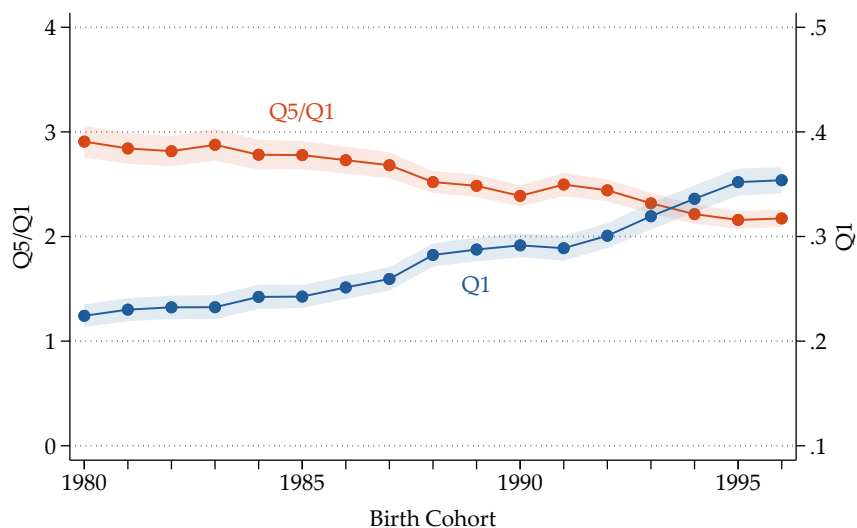
The picture that emerges is best summarized by Figure 2.7, which depicts the A-Level share by quintile across birth cohorts: The *Bildunsexpansion* took place uniformly across the income distribution with increases of about 14 percentage points in the A-Level share in all parts of the distribution from the 1980 cohort to the 1996 cohort. Did the *Bildunsexpansion* achieve its goal of fostering social mobility in Germany? While the expansion unquestionably increased absolute mobility as we measure it, the question whether relative mobility increased or remained stagnant since 1980 depends on the measure. While the attenuation of the Q5/Q1 odds ratio caused by the uniform increases in A-Level shares suggests an increase in relative mobility according to a proportional notion of the concept, the unaltered top-bottom gap in the probability of attaining an A-Level captured by the parental income gradient emphasizes stagnation in absolute differences. Since both, absolute and relative disparities often form the normative basis for interventions, both readings are justifiable.

Figure 2.5: Parental Income Gradient by Cohort 1980-1996



*Notes:* This figure shows the evolution of our estimates of the parental income gradient by birth cohort. The shaded area depicts pointwise 95% confidence intervals computed based on LZ standard errors as described in Section 2.3.3.

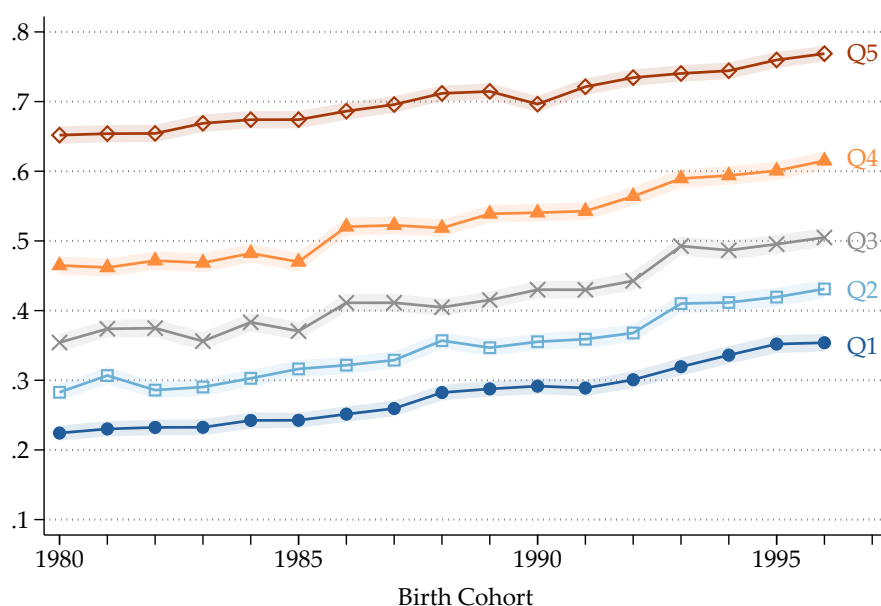
Figure 2.6: Quintile Measures by Cohort 1980-1996



*Notes:* This figure shows the evolution of our estimates of the quintile based measures of social mobility by birth cohort. While the left axis corresponds to the Q5/Q1 ratio, the right axis corresponds to the Q1 measure. The shaded areas depict pointwise 95% confidence intervals computed based on the standard errors of the underlying sample averages as described in Section 2.3.3.



Figure 2.7: A-Level Share by Cohort Quintile 1980-1996



*Notes:* This figure shows the share of children born between 1980 and 1996 who obtained an A-Level degree by birth cohort and quintile of the parental income distribution in the MZ data. The shaded areas depict pointwise 95% confidence intervals computed based on the standard errors of the underlying sample averages as described in Section 2.3.3.

## 2.5 Regional Estimates

An interesting empirical observation documented in the recent empirical literature on social mobility is that there exists substantial geographic variation in social mobility measures within politically homogenous entities, suggesting that within comparisons can be used to gain a better understanding of the causes of social mobility (cf. Chetty et al., 2014; Acciari et al., 2019; Corak, 2020; Deutscher and Mazumder, 2020; Chuard and Grassi, 2020). This idea is appealing, as attributing discrepancies in social mobility to differences in single characteristics or policies of otherwise heterogeneous entities is difficult to justify. Complementary to well-designed evaluations of political reforms that rely on within variation across time (e.g. Bertrand et al., 2021), geographic variation can be helpful in understanding the causal mechanisms fostering or impeding social mobility by identifying exposure effects (cf. Chetty and Hendren, 2018a; Chetty and Hendren, 2018b). Moreover, pronounced regional differences can suggest mechanisms that warrant investigation. Understanding the sources of regional variation is also relevant for policy makers, as the efficiency of place-based programs designed to foster social mobility locally depends on whether the local differences are driven by amendable place characteristics or the composition of local households, a topic of ongoing academic debate.

The regional analysis conducted in this section is motivated by these considerations. In a first step, we present evidence of meaningful geographic variation in our mobility measures across regions in Germany. In a second step, we then ask what we can learn from the observed differences. We structure our regional analysis by disaggregating our data in a stepwise fashion, lending credence to our parametric mobility statistics while taking into account the political and economic landscape of Germany.

### 2.5.1 States

A natural starting point for our regional analysis are the 16 federal states of Germany. This is because, by constitutional law, the responsibility for the design and implementation of the education system falls under the jurisdiction of the German states and not the federal government.<sup>15</sup> As a consequence, state law-makers and the executive branch have considerable discretion in the design of the state education systems, leading to distinctions in the rigor of the tracking system, the capacities of each track, the types of schools and curricula as well as examination standards and other important features of the education system. In particular, the states differ with respect to the tracking age, that is the duration of primary school after which all children are allocated into the different tracks, the number of tracks (2 or 3), as well as the importance of teacher recommendations for admitted track choices. While in all states teachers provide recommendations on the suggested track for each child at the end of primary school, the track recommendations are binding in some states while other states allow parents to freely choose tracks for their children. These parameters of the state education systems and their suspected consequences for social mobility are often at the center of the public debate on educational mobility in Germany.

Table 2.6 reports our mobility estimates for the 16 states, sorted by our point estimates of the parental income gradient in ascending order. We document significant and economically meaningful differences in both, absolute and relative mobility measures between states. For example, the top-bottom gap in the probability of attaining an A-Level degree is approximately 20 percentage points larger in Bremen than in Hamburg, two city states in north-west Germany approximately 100 kilometers apart. Similarly, the share of children obtaining an A-Level degree from the bottom quintile of the parental income distribution is 10 percentage points larger in Baden-Wuerttemberg than in Bavaria, the two southmost states of Germany. The table also

---

<sup>15</sup>These differences do not extend to the A-Level degree. The Standing Conference of State Education Secretaries has the stated goal to ensure a high degree of comparability of educational qualifications across German States and there are no legal differences between the A-Level degrees issued from different states.

Table 2.6: Social Mobility at the State Level

State	Gradient	Q1	Q5	Q5/Q1	A-Level Share	Number of Tracks	Tracking Grade	Binding Teacher Rec.
<b>Hamburg (HH)</b>	0.45 (0.026)	0.43 (0.017)	0.80 (0.014)	1.85 (0.082)	0.60	2	5	No
<b>Rhineland-Palatinate (RP)</b>	0.50 (0.015)	0.36 (0.009)	0.76 (0.009)	2.11 (0.064)	0.53	2	5	No
<b>North Rhine-Westphalia (NW)</b>	0.51 (0.007)	0.41 (0.005)	0.82 (0.014)	2.02 (0.024)	0.59	3	5	R
<b>Hesse (HE)</b>	0.52 (0.012)	0.39 (0.008)	0.81 (0.014)	2.07 (0.047)	0.59	3	5	R
<b>Baden-Wuerttemberg (BW)</b>	0.52 (0.009)	0.34 (0.006)	0.76 (0.005)	2.25 (0.045)	0.53	3	5	R
<b>Saarland (SL)</b>	0.53 (0.032)	0.33 (0.019)	0.74 (0.021)	2.27 (0.145)	0.54	2	5	R
<b>Schleswig-Holstein (SH)</b>	0.53 (0.019)	0.32 (0.012)	0.76 (0.011)	2.35 (0.094)	0.52	2	5	No
<b>Lower Saxony (NI)</b>	0.55 (0.010)	0.29 (0.006)	0.73 (0.007)	2.54 (0.060)	0.48	3	5	No
<b>Bavaria (BY)</b>	0.55 (0.009)	0.25 (0.006)	0.67 (0.005)	2.72 (0.064)	0.42	3	5	Yes
<b>Berlin (BE)</b>	0.56 (0.017)	0.39 (0.010)	0.85 (0.010)	2.19 (0.064)	0.59	2	7	No
<b>Brandenburg (BB)</b>	0.57 (0.022)	0.36 (0.016)	0.84 (0.012)	2.35 (0.109)	0.60	2	7	R
<b>Saxony-Anhalt (ST)</b>	0.57 (0.028)	0.25 (0.014)	0.72 (0.021)	2.82 (0.175)	0.43	2	5	R
<b>Saxony (SN)</b>	0.61 (0.002)	0.27 (0.011)	0.78 (0.013)	2.85 (0.124)	0.48	2	5	Yes
<b>Mecklenburg-Vorpommern (MV)</b>	0.63 (0.034)	0.25 (0.016)	0.76 (0.025)	3.01 (0.214)	0.45	2	5	No
<b>Bremen (HB)</b>	0.64 (0.036)	0.32 (0.019)	0.86 (0.023)	2.65 (0.176)	0.55	2	5	No
<b>Thuringia (TH)</b>	0.65 (0.026)	0.25 (0.014)	0.76 (0.019)	3.07 (0.185)	0.46	2	5	Yes

*Notes:* This table reports estimates of our relative and absolute mobility measures for each federal state of Germany based on all children observed in the MZ survey waves 2011-2018. The standard errors reported in parentheses below each point estimate are computed as described in Section 2.3.3. The states are sorted in ascending order by the point estimate of the parental income gradient. The classification of the state education systems is based on the description of educational reforms in [Helbig and Nikolai \(2015\)](#). The entry "R" in the last column indicates that teacher recommendations were reformed during the time period relevant for our analysis. The reforms of teacher recommendations relevant for the birth cohorts in our sample were conducted as follows: North Rhine-Westphalia (NW): binding until 1997, non-binding until 2006, then binding until 2010, and non-binding since. Hesse (HE): non-binding since 1993. Baden-Wuerttemberg (BW): non-binding since 2011. Saarland (SL): non-binding until 2000, then binding until 2009, then non-binding. Brandenburg (BB): binding since 2007. Sachsen-Anhalt (ST): non-binding until 2005, then binding, then non-binding since 2012.

reiterates the east-west gap documented in Section 2.4.1: with the exception of Bremen, the five least mobile states are all located in East Germany.

While we find that the differences in our measure of absolute mobility can be well explained by differences in the states' A-Level shares, that is the relative capacity of the highest track, there is no clear pattern in our estimates with respect to the aforementioned characteristics of the state education systems displayed in the last three columns of the table. Our findings suggest that, while certainly important, the design of the tracking system is not suited to explain the pronounced differences in our mobility measures between states.

Figure 2.C.1 shows that the estimated differences are not driven by differences in the shape of the empirical CEFs by displaying the rank-binned data underlying our state-level estimates. While we observe some degree of convexity in Bavaria and to a smaller extent in Baden-Wuerttemberg, we find that the linearity assumption underlying our parametric mobility estimates is supported by the data.

## 2.5.2 Cities

A similar picture emerges when we restrict our analysis to urban regions of Germany. Table 2.7 reports our mobility estimates for the 15 largest labor markets of Germany, consisting of cities and their catchment areas as defined by commuting flows.

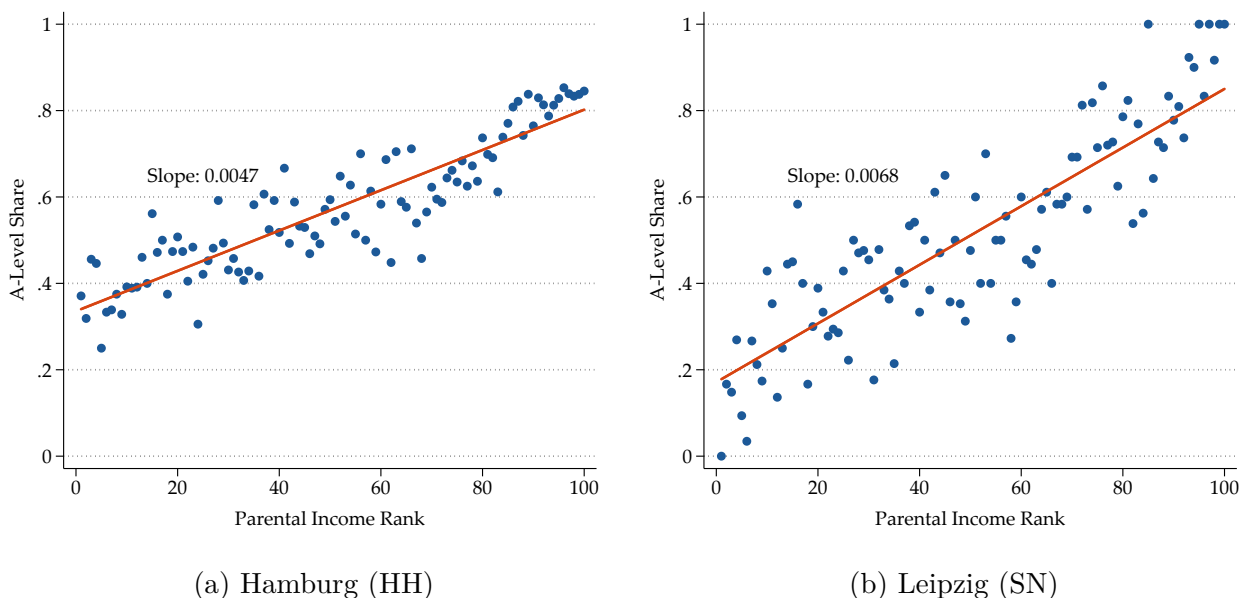
On average, the largest urban regions of Germany show lower levels of relative, but higher levels of absolute social mobility (and A-Level shares), as compared to the national estimates reported in Table 2.4. At the same time, the table shows that the regional differences observed at the state-level can also be found within states. For example, the top-bottom gap is approximately 8 percentage points larger in Cologne than in Duesseldorf, two large cities in North Rhine-Westphalia (NW) located approximately 40 kilometers apart. Similarly, our estimates of absolute mobility differ by 8 percentage points between Nuremberg and Munich, two large cities in Bavaria (BY). The most striking discrepancy between cities in our data is observed for Hamburg and Leipzig, with a difference of approximately 20 percentage points in the estimated top-bottom gap, as well as 15 percentage points in our estimate of the Q1 measure. Figure 2.8 displays our raw data for the two cities. Similar to the previously considered partitions of our data, we find that the empirical CEFs are reasonably well approximated by a linear function. Overall, our city-level findings suggest that the (relative) opportunities of (disadvantaged) children can differ meaningfully across politically similar and geographically close regions of Germany. Figure 2.C.2 shows the empirical CEFs for each urban labor market.

Table 2.7: Social Mobility in the 15 Largest Urban Labor Markets

City	State	Gradient	Q1	Q5	Q5/Q1	A-Level Share
<b>Hamburg</b>	HH	0.47 (0.019)	0.41 (0.013)	0.79 (0.010)	1.94 (0.071)	0.58
<b>Duesseldorf</b>	NW	0.47 (0.023)	0.45 (0.018)	0.84 (0.012)	1.87 (0.080)	0.65
<b>Muenster</b>	NW	0.47 (0.032)	0.47 (0.023)	0.84 (0.018)	1.78 (0.096)	0.62
<b>Gelsenkirchen</b>	NW	0.50 (0.027)	0.40 (0.015)	0.80 (0.019)	2.01 (0.090)	0.57
<b>Stuttgart</b>	BW	0.50 (0.018)	0.34 (0.013)	0.75 (0.009)	2.19 (0.080)	0.55
<b>Bonn</b>	NW	0.50 (0.028)	0.44 (0.021)	0.86 (0.013)	1.94 (0.096)	0.65
<b>Duisburg</b>	NW	0.51 (0.025)	0.42 (0.016)	0.84 (0.015)	2.02 (0.085)	0.58
<b>Frankfurt</b>	HE	0.52 (0.019)	0.42 (0.015)	0.83 (0.009)	1.97 (0.071)	0.62
<b>Munich</b>	BY	0.54 (0.019)	0.31 (0.016)	0.71 (0.009)	2.32 (0.124)	0.53
<b>Dortmund</b>	NW	0.55 (0.025)	0.40 (0.016)	0.86 (0.015)	2.16 (0.095)	0.59
<b>Cologne</b>	NW	0.55 (0.021)	0.38 (0.015)	0.85 (0.012)	2.25 (0.094)	0.60
<b>Hanover</b>	NI	0.56 (0.027)	0.30 (0.018)	0.76 (0.016)	2.51 (0.155)	0.53
<b>Berlin</b>	BE	0.56 (0.016)	0.39 (0.010)	0.85 (0.009)	2.20 (0.064)	0.59
<b>Nuremberg</b>	BY	0.60 (0.028)	0.23 (0.018)	0.70 (0.018)	3.01 (0.246)	0.43
<b>Leipzig</b>	SN	0.68 (0.034)	0.26 (0.021)	0.80 (0.024)	3.11 (0.266)	0.48

*Notes:* This table reports estimates of absolute and relative mobility for the 15 largest urban local labor markets in Germany, as measured by their total population in 2017, based on the MZ waves 2011-2018. The local labor markets are sorted in ascending order by the point estimate of the parental income gradient. Standard errors are computed as described in Section 2.3.3. The point estimates for the city-states can differ from those reported in Table 2.6, as the urban labor markets typically also include surrounding towns and villages.

Figure 2.8: Mobility in Hamburg and Leipzig



*Notes:* This figure shows the fraction of children aged 17-21 observed in the MZ survey waves 2011-2018 that are either enrolled in the upper stage of an A-Level track or have already attained an A-Level degree in Hamburg (a) and Leipzig (b), as well as the best linear approximation to the empirical CEF. Standard errors are reported in Table 2.7.

### 2.5.3 Local Labor Markets

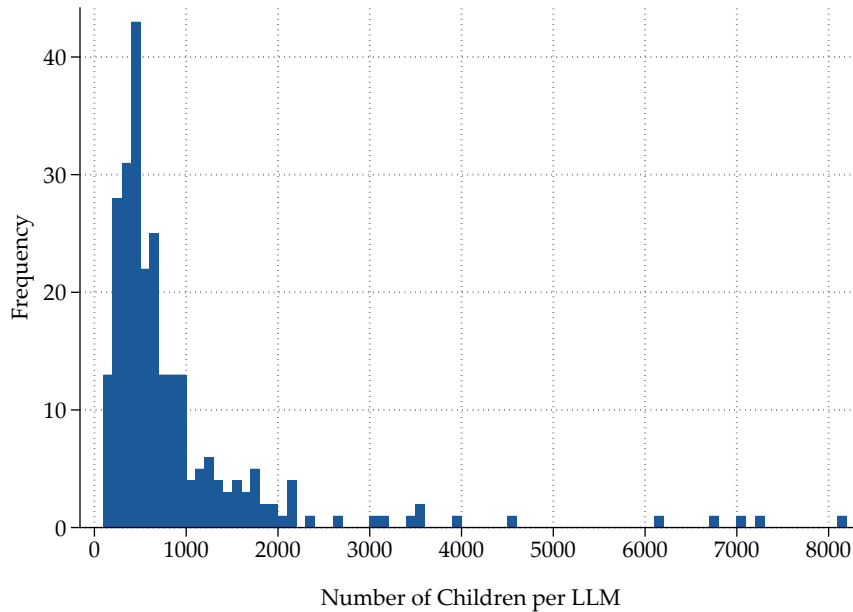
With the aim to learn more about the regional variation in mobility measures across Germany, we disaggregate our data once more to the level of 258 local labor markets (LLMs).<sup>16</sup> The 258 LLMs represent aggregations of counties based on commuting flows, comparable to the commuting zones in the US, and provide a geographic partition of Germany into areas in which people live and work. As such, the properties of local labor markets likely influence the decisions of households and the opportunities of children. With the exceptions of five local labor markets (Bremen, Bremerhaven, Hamburg, Mannheim and Ulm), all counties aggregated into LLMs belong to a single federal state.

While disaggregating our data to the LLM level allows us to ask several interesting questions, it makes it substantially harder to distinguish meaningful variation from noise. Figure 2.9 displays the frequencies of the numbers of children observed by local labor market in bins of 100 observations. The median number of children in our sample (observations) per LLM is 552 (mean: 895). The lowest number of observations across all LLMs is 100 (LLM Sonneberg) and the largest number of observations is 8159 (LLM Stuttgart). In the following, we present our LLM level results using maps of Germany. In Appendix 2.C.4 we show analogous maps,

<sup>16</sup>We assign households to the LLM of their current place of main residence as reported in our data.

aggregated to the level of spatial planning regions, a higher-level aggregation of commuting zones, which demonstrate that the general patterns are not driven by noise.

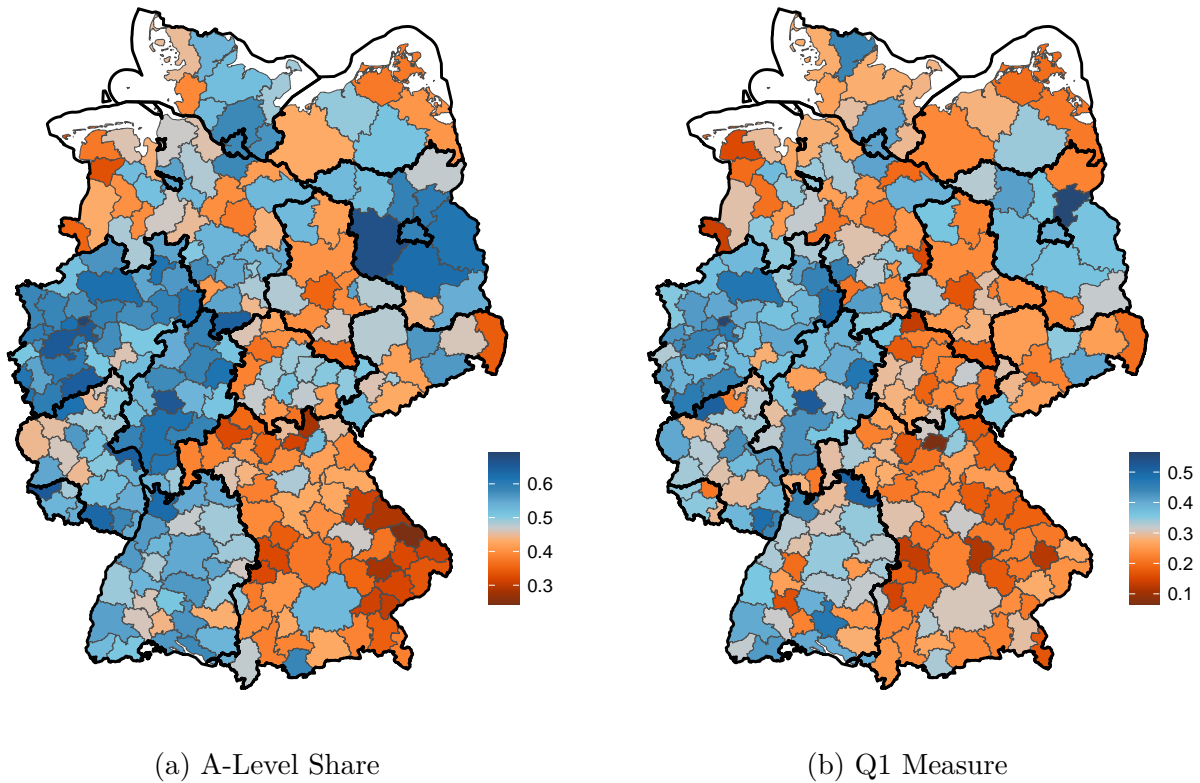
Figure 2.9: Histogram of the Number of Observations by Local Labor Market



*Notes:* This figure shows the frequencies of number of children (observations) by local labor market in bins of 100 observations in the MZ waves 2011-2018.

**Regional Patterns in Absolute Mobility.** We begin our local labor market-level analysis by studying the regional variation in our absolute mobility estimates. Figure 2.10 shows the A-Level Share (a) and our estimates of the Q1 measure (b) in each of the 258 LLMs. As indicated by the legends, red areas correspond to regions with low, and blue areas to regions with high values of the respective statistic. For both statistics, state-level clusters are clearly visible. Panel (a) shows that the A-Level share is uniformly higher in the local labor markets of states with high average A-Level capacities, such as North Rhine-Westphalia or Hesse. In line with our city-level estimates, we find that across Germany, cities typically demonstrate higher A-Level shares than the surrounding local labor markets. Comparing the two panels demonstrates that, unsurprisingly, our measure of absolute mobility is closely linked to the local A-Level share. Consequently, we observe lower levels of absolute mobility in regions with low A-Level shares, such as Bavaria. Overall, there is substantial variation in absolute mobility. In some regions, less than 15% of children from the bottom quintile of the national income distribution obtain an A-Level degree, whereas in other regions this number exceeds 50%. We find that 42% of the variation in the Q1 measure and 57% of the variation in the A-Level share can be attributed to state level differences. In Table 2.C.1 we report the correlations between our mobility estimates.

Figure 2.10: A-Level Share and Q1 Measure by Local Labor Market



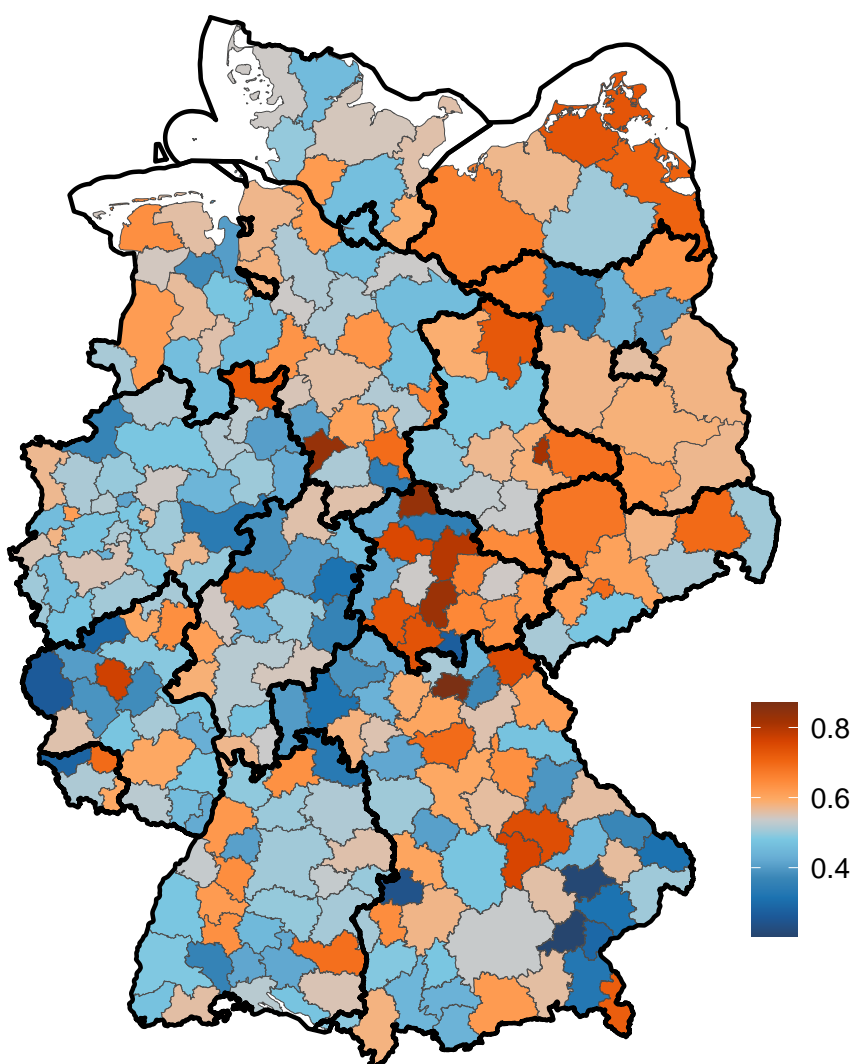
*Notes:* This figure presents heat maps of the estimated A-Level shares (a) and Q1 measures (b) by local labor market. Bold black lines indicate state borders. Children are assigned to local labor markets according to their place of residence at observation. The estimates are based on children aged 17-21 in the MZ waves 2011-2018.

**Regional Patterns in Relative Mobility.** While the variation in our absolute mobility measure can be well explained by state A-Level shares, the regional patterns in our estimates of relative mobility are less obvious. Figure 2.11 presents a heat map of our parental income gradient estimates.<sup>17</sup> Blue areas represent areas of high mobility (low gradients), whereas red areas indicate less mobile regions. The interpretation of the displayed regional patterns is complicated by the fact that our LLM-level gradient estimates are estimated less precise than the sample averages displayed in Figure 2.10. However, we document a few interesting regularities. First, in line with our state-level estimates, LLMs in the former GDR (East Germany) typically exhibit lower levels of relative mobility. Second, urban labor markets are typically adjacent to clusters of more relatively mobile rural labor markets. Moreover, we find multiple clusters of local labor markets with similar estimated levels of relative mobility within each (non-city) state. In contrast to our estimates of absolute mobility, some of the observed

<sup>17</sup>The corresponding heatmap for the Q5/Q1 ratio is displayed in Figure 2.C.3 in the Appendix.



Figure 2.11: Parental Income Gradient by Local Labor Market



*Notes:* This figure presents a heat map of our estimates of the parental income gradient by local labor market. Children are assigned to labor markets according to their current place of residence. The estimates are based on MZ waves 2011-2018. The estimate of the parental income gradient is obtained as the slope coefficient of a regression of the A-Level indicator on a constant and the parental income rank, multiplied by 100, as described in Section 2.3.2.

clusters extend beyond state borders. In some rural labor markets, the parental income gradient is estimated below 0.3, whereas in the least mobile areas the gradient exceeds 0.8. While these differences in point estimates can partially be driven by noise, the evidence we present in Figure 2.C.4 in Appendix 2.C shows that a similar pattern emerges when aggregating data to the level of spatial planning regions. Our maps demonstrate that there exists substantial variation in relative mobility estimates across local labor markets in Germany as a whole, as well as within individual states. The LLMs with the highest gradient (Lichtenfels) and the lowest gradient (Mühldorf) are both located in Bavaria but similar disparities exist in other states.

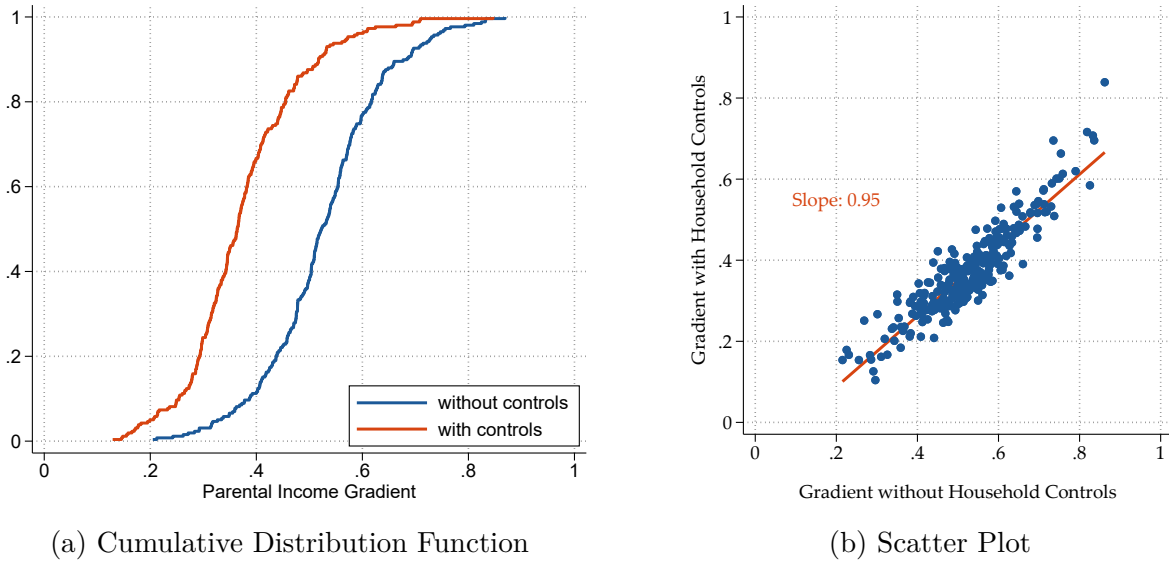
**Sorting** What can we learn from the estimated regional differences across local labor markets? A first potential insight relates to the debate on the potential of place-based mobility policies. An active literature argues that places shape economic outcomes and that place-based policies can be an effective and cost-efficient tool to improve outcomes by amending local conditions (cf. [Kline and Moretti, 2014](#); [Neumark and Simpson, 2015](#)). In the context of educational policies and social mobility, it is often argued that the government should allocate additional resources to the local public school systems of socially immobile regions to enhance mobility. However, such a policy is unlikely to achieve its objective if social mobility in the respective regions is low for reasons other than the quality of local schools. For example, if a region exhibits a high degree of inequality in parental educational attainment, the patterns we document in Section 2.4.1 would likely result in low levels of relative mobility as measured by the parental income gradient. Such systematic sorting mechanisms are at the center of the academic debate regarding the interpretation of the regional differences in estimated mobility measures within countries. For example, for the US, [Rothbaum \(2016\)](#) and [Gallagher et al. \(2019\)](#) suggest that a substantial share of the geographical variation in the intergenerational mobility measures reported in [Chetty et al. \(2014\)](#) can be explained by differences in household characteristics across commuting zones. Unfortunately, this can not be directly tested in the administrative tax data used in [Chetty et al. \(2014\)](#), as it contains only limited information on household characteristics.

In contrast, the German census data allows us to test whether regional differences are muted when we account for household characteristics. We do so by computing conditional rank gradients, which we then compare to the corresponding regional parental income gradients. The set of conditioning variables that we use for this exercise are the discrete attributes that we used for our subgroup analysis in Section 2.4.1.<sup>18</sup> Figure 2.12 panel (a) plots the marginal distributions of conditional and unconditional rank gradients. It shows that the CDF of the unconditional gradient first order stochastically dominates the CDF of the conditional gradient, which is expected given the patterns documented in Table 2.5. At the same time, the variance of the distribution of conditional rank gradients is approximately the same as the variance of the unconditional gradient. While this suggests that sorting does not play a major role, the same pattern would emerge if our regional estimates were dominated by noise, in the sense that the between local labor market variation in gradients was negligible relative to the estimation uncertainty. However, as displayed in panel (b), we find that the relative ordering of gradients

---

<sup>18</sup>In contrast to Section 2.4.1, we include four categories of parental education and do not include the East-West indicator.

Figure 2.12: Sorting: Conditional and Unconditional Rank Gradients



*Notes:* Panel (a) plots the Cumulative Distribution Function (CDF) of our estimates of the conditional and unconditional parental income gradients across local labor markets. Panel (b) shows a scatter plot of our estimated conditional and unconditional rank gradients for each local labor market as well as their linear fit.

is unaffected by conditioning, which strongly suggests that regional sorting of households can not explain the regional variation in relative social mobility as we measure it. Conditional and unconditional gradients are strongly correlated, with a Pearson correlation of 0.91 and a Spearman rank correlation 0.89. In Appendix 2.C.5 we show that our regional results on relative mobility are unaffected if we compute rank gradients based on the state or regional income distribution.

**Predictors of Mobility** The local labor market-level estimates of our mobility statistics further allow us to characterize mobile regions by conducting a prediction exercise. To this end, we construct a database of 71 regional indicators with information on labor market participation, economic conditions, infrastructure, demographics, housing and living conditions, as well as indicators on the local educational institutions and social characteristics. Table 2.C.2 lists all regional indicators as well as their respective sources.

In order to obtain an optimal set of regional mobility predictors, we rely on a Random Forest based measure of variable importance.<sup>19</sup> The set of the 15 most informative predictors

<sup>19</sup>There are several ways to compute a Random Forest based measure of variable importance. We choose the implementation proposed by Strobl et al. (2008), which computes a conditional permutation importance measure that accounts for the dependence structure between the predictors. We split our data in a training and test data set (75-25 split) and fit 1000 trees, randomly selecting  $72/3 = 24$  variables for each split. The Random Forest algorithm predicts 38% of the variation in the testing data ( $R^2 = 0.38$ ).

is displayed in Table 2.8.<sup>20</sup> The last column of the table displays the sign of the bivariate correlation between each variable and the parental income gradient. A positive sign implies that the indicator is predictive for low relative mobility (a high gradient). For example, LLMs with a high prevalence of school dropouts are associated with low relative mobility. Overall,

Table 2.8: The 15 Most Informative Predictors of Relative Mobility

Variable	Importance Measure	$\rho$
School Dropout Rate	0.85	+
Share Married	0.60	-
Teenage Pregnancies	0.42	+
Students	0.39	-
Median Income Vocational Qualification	0.18	-
Broadband Availability	0.17	+
Distance to Next College	0.15	-
Unemployment Rate	0.14	+
Gender Wage Gap	0.14	+
Share without Vocational Qualification	0.13	-
Gini Parental Income	0.08	-
Share Marginal Employment	0.07	-
Share Children 0-2 in Childcare	0.07	+
Share Social Assistance	0.07	+
Share on Vocational A-Level Track	0.07	-

*Notes:* This table lists the optimal predictive set of 15 regional indicators for the local labor market parental income gradient estimates, as chosen by a Random Forest based measure of variable importance. The last column shows the sign of the Pearson correlation coefficient between each variable and the estimates of the parental income gradient.

our selection procedure highlights social characteristics, the local organization of the education system and labor market conditions.<sup>21</sup> For example, local labor markets with a high prevalence of school dropouts are associated with low relative mobility. The same applies to the share of teenage pregnancies, the prevalence of child poverty and the share of individuals which are dependent on social assistance. All these indicators point to comparatively disadvantaged social contexts in these labor markets, consistent with social capital based explanations of regional disparities in mobility. Other variables like the access to broadband internet or the distance to the next university are less straightforward to interpret. In our view, the findings reported in Table 2.8 support the view that our measurement approach, despite its descriptive nature, is

<sup>20</sup>The ranking of the selected predictors varies for different implementations of the Random Forest algorithm. We are therefore cautious not to over-interpret the ranking between single predictors.

<sup>21</sup>In Table 2.C.3, we report analogous results obtained by conducting our prediction exercise separately for the 129 largest and smallest local labor markets. Reassuringly, similar sets of predictors are selected.

able to detect meaningful variation in regional mobility patterns and we hope that future work will be able to build on our analysis to shed light on the causal determinants of social mobility.

## 2.6 Conclusion

This paper proposes a measurement strategy for and provides novel empirical evidence on the level, evolution and geography of social mobility in Germany. Our statistics characterize social mobility by the association between children's educational opportunities and their parents' relative position in the national income distribution, allowing for the use of census data.

At the national level, we document that a 10 percentile increase in the parental income distribution is associated with a 5.2 percentage point increase in the probability to obtain an A-Level degree, implying a top-bottom gap of approximately 50 percentage points. This gap remained stable for the 1980-1996 birth cohorts, despite a large-scale expansion of upper secondary education. We document substantial variation in our mobility measures across regions of Germany, including within states, and show that regional differences in household characteristics cannot account for these disparities. As such, our findings are consistent with place-based rather than sorting-type explanations of geographical dispersion in mobility measures. Obtaining an optimal set of mobility predictors based on our disaggregated estimates, we find that social characteristics, the local organization of the education system and labor market conditions are best suited to predict our relative mobility estimates.



# Appendices to Chapter 2

## 2.A Additional Information on the Microcensus

The Microcensus (MZ) is the largest household survey in Europe. Conducted annually with a sampling fraction of 1% of all individuals who have the right of residence in Germany, it yields representative statistics on the German population. The MZ has been conducted in West Germany since 1957 and in the new federal states (East Germany) since 1991. It is planned and prepared by the Federal Statistical Office of Germany and carried out by the statistical offices of the 16 German states. The legal basis of the MZ are the Federal Statistics Law and the Microcensus Law, which make it compulsory for households to provide answers to the core items of the survey. The non-response rate is further minimized by repeated visits of interviewers to irresponsive households and multiple possible ways for the sampled households to submit information.

Since 1972, the MZ uses a single-stage stratified cluster sampling design. The primary sampling units typically consist of neighbouring buildings (larger buildings are divided into smaller partitions). For the survey waves utilized in this paper, the target size for a cluster is 7-15 households. All households and residents in the sampled clusters are interviewed. The database used to assign households to clusters is created based on the most recent full census and updated annually using information on new construction activities. Since 1977, each cluster is assigned to a "rotation quarter" that remains in the survey for four years. Each year, a quarter is replaced by new clusters. The survey does not follow individuals who leave their cluster, but replaces them by the new residents. The MZ survey design results in data best described as a repeated survey with partial overlap of units as sketched in Figure 2.A.1.<sup>22</sup>

Figure 2.A.1: Illustration of the Microcensus Survey Design

Survey Wave	Rotation Quarter							...
	1	2	3	4	5	6	7	
1	✓	✓	✓	✓	X	X	X	...
2	X	✓	✓	✓	✓	X	X	...
3	X	X	✓	✓	✓	✓	X	...
4	X	X	X	✓	✓	✓	✓	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

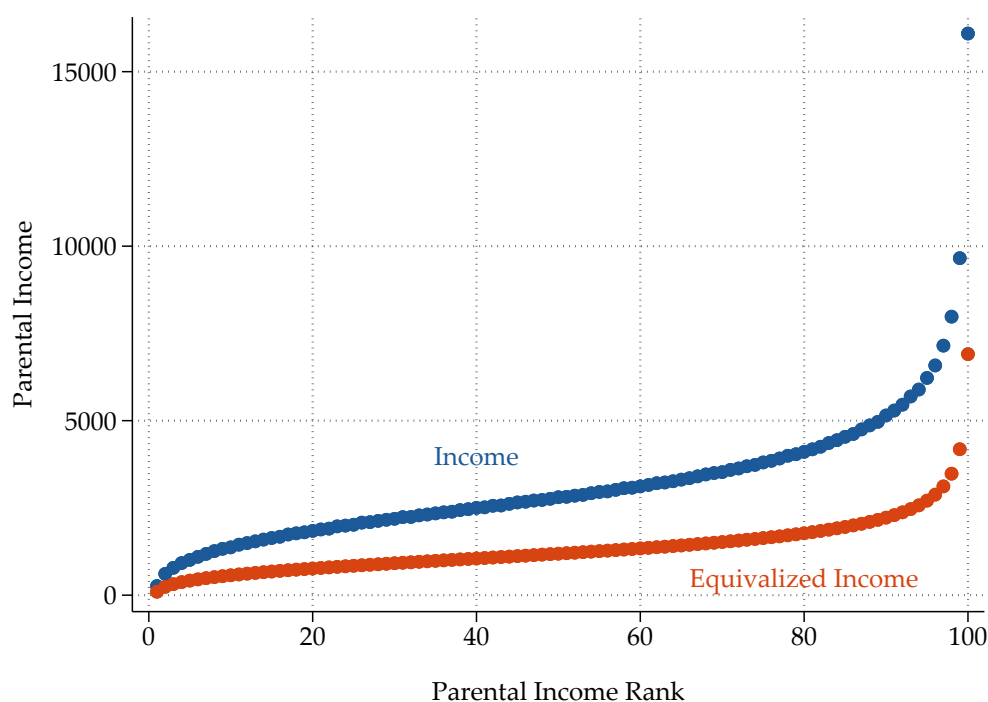
<sup>22</sup>For more information, see <https://www.gesis.org/en/missy/metadata/MZ/> and the references listed there.

## 2.B Additional Material: National Estimates

### 2.B.1 Sample Income Distribution and Ranks

Figure 2.B.1 displays the sample distributions of (equivalized) CPI-adjusted monthly household net income in the 2011-2018 MZ data. We CPI adjust all household incomes in order to allow for meaningful aggregation of survey-years before computing ranks. Ties are broken by allocating households to the lower quantile. Our findings are insensitive to the choice of tie-breakers.

Figure 2.B.1: Equivalized Household Income by Percentile Rank



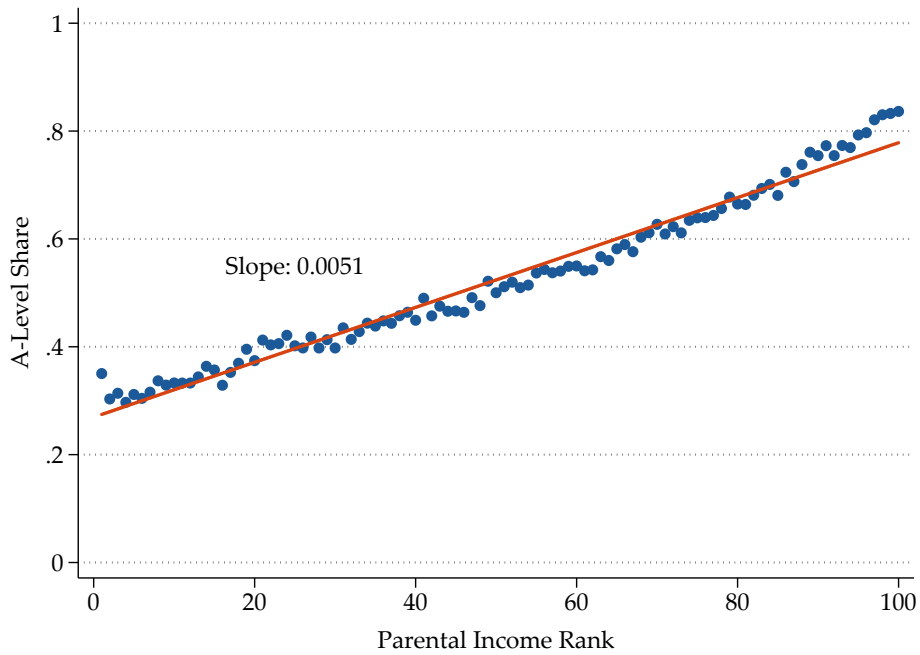
*Notes:* This figure shows the distribution of (equivalized) CPI adjusted household incomes in the 2011-2018 MZ data.

### 2.B.2 Alternative Equivalization Schemes

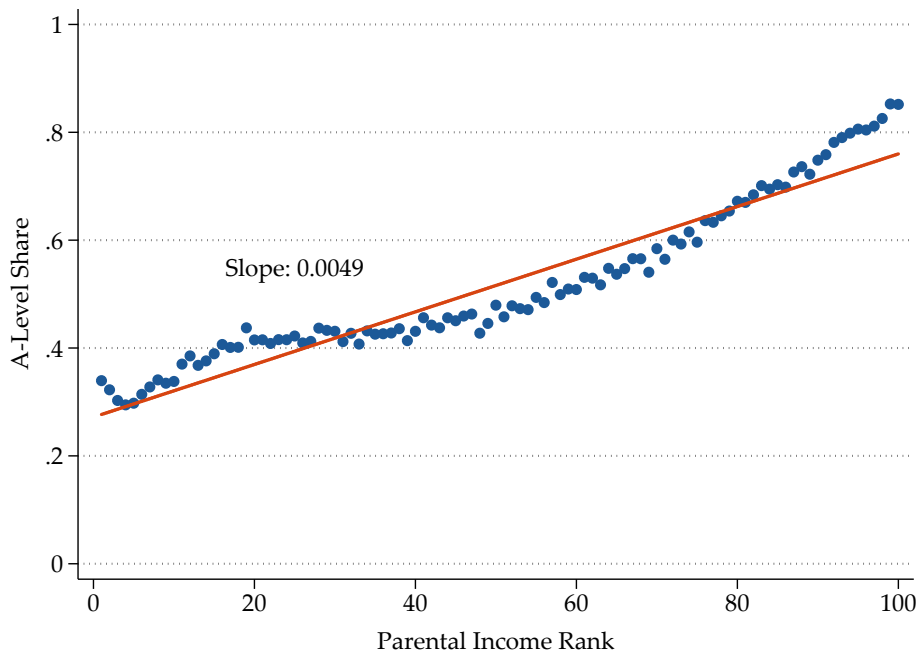
Figure 2.B.2 shows the national empirical CEF for the pooled data, as well as the corresponding best linear approximation for income ranks computed based on the distribution of per-capita (household size) adjusted household incomes (a) and unadjusted household incomes (b). The figure illustrates that the linear approximation to the CEF improves when ranks are computed based on equivalized income and that our estimates are not sensitive to the choice of the equivalization scheme, that is whether we rely on the OECD-modified or a per capita adjustment scheme.



Figure 2.B.2: National Estimates under Different Equivalization Schemes



(a) Per Capita Adjustment



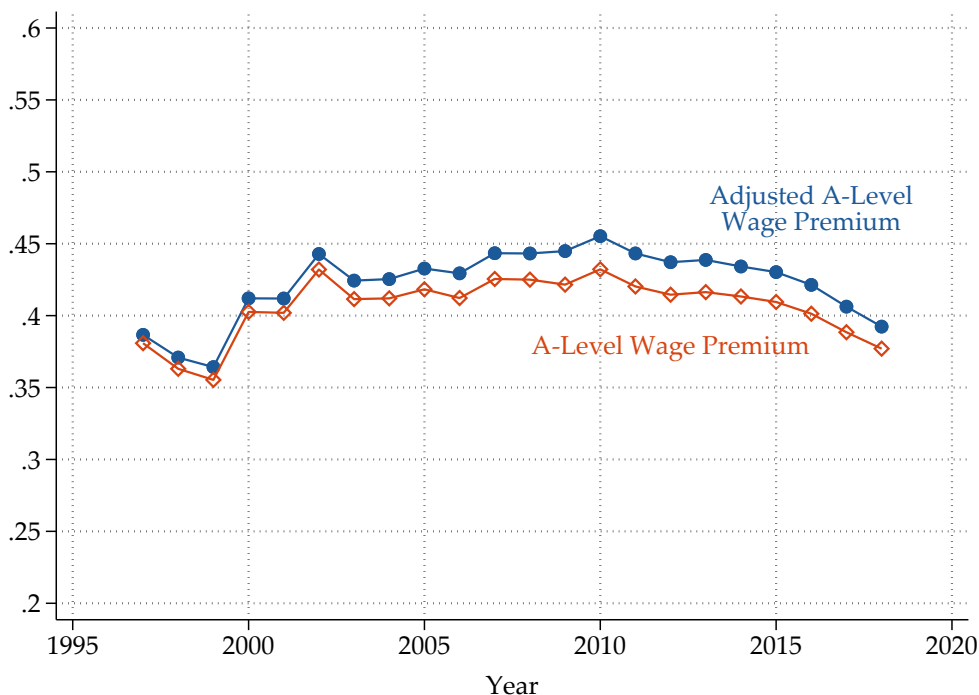
(b) No Adjustment

*Notes:* This figure shows the empirical CEF and the corresponding best linear approximations when income ranks are computed based on per-capita adjusted household income (a) or unadjusted household income (b). The reported slope coefficients are obtained by OLS using the underlying micro data.

### 2.B.3 A-Level Wage Premium 1997-2016

Figure 2.B.3 displays our point estimates for the A-Level wage premium based on the MZ data, obtained by regressing the log of net monthly individual income of full-time workers aged 30-45 on an A-Level indicator in our data, separately for each survey year. The "adjusted" point estimate additionally conditions on age indicators to indirectly account for job experience.

Figure 2.B.3: A-Level Wage Premium, Years 1997-2016

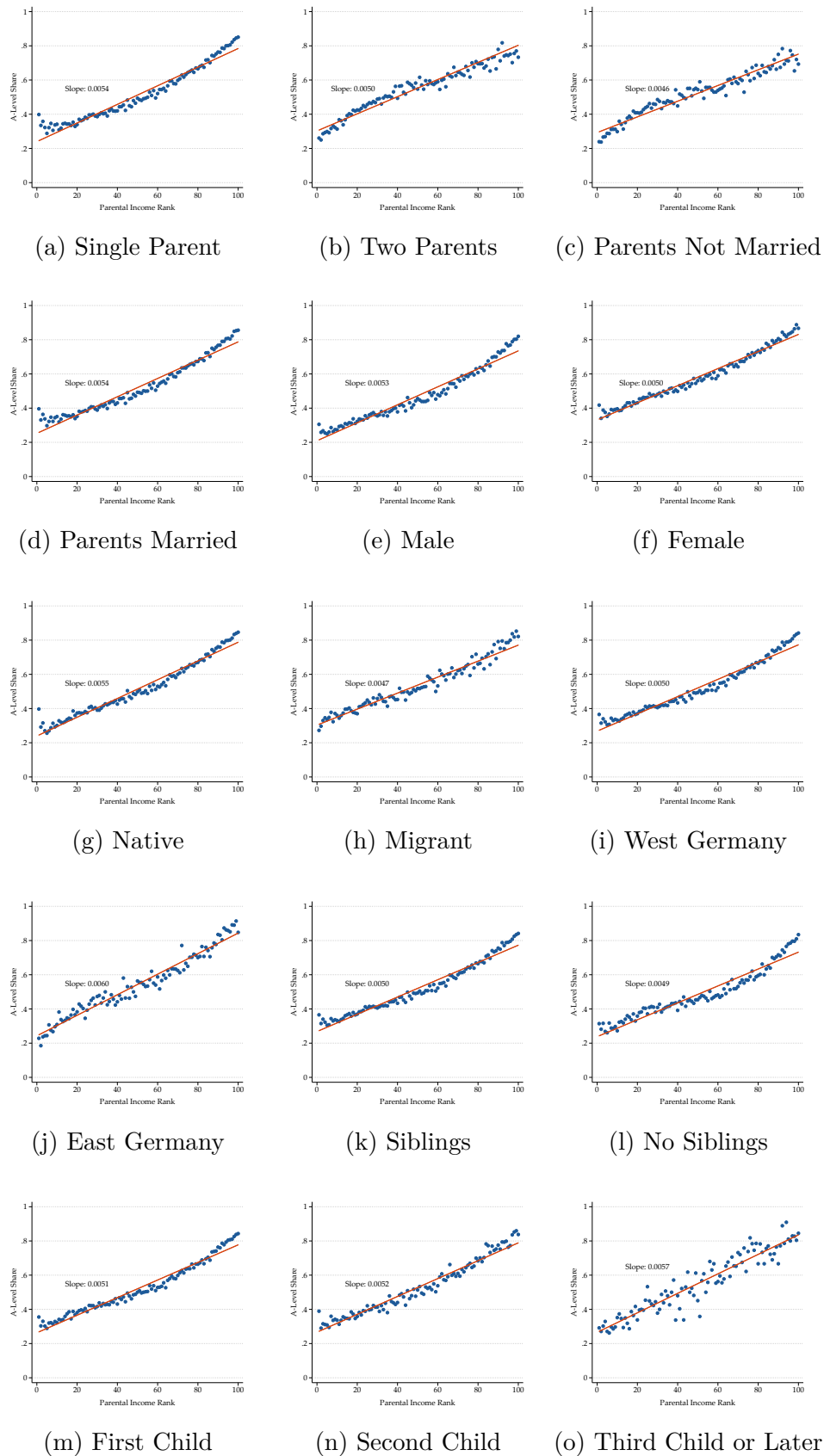


*Notes:* This figure shows the development of the A-Level wage premium for the years 1997-2016 as estimated based on the MZ data. We compute the A-Level wage premium by regressing the log of net monthly personal income of full-time workers aged 30-45 on an A-Level indicator. The adjusted A-Level wage premium is computed by additionally conditioning on a set of age indicators to indirectly account for job experience.

### 2.B.4 Subgroup CEFs and Trends

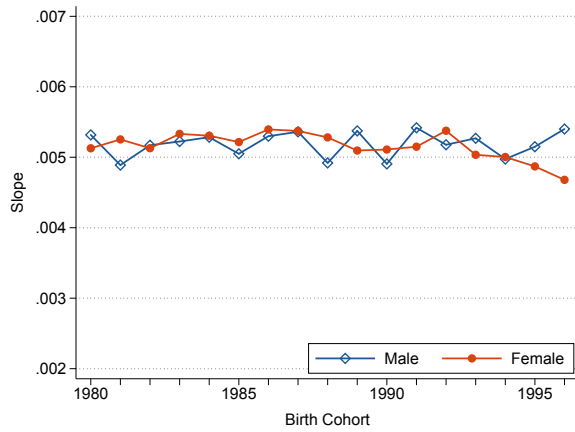
Figure 2.B.4 displays the empirical CEFs as well as the best linear approximation to the empirical CEFs for each subgroup discussed in Section 2.4.1. Figures 2.B.6 and 2.B.5 display our estimates of the parental income gradients and A-Level shares by birth cohort for the same subgroups. The figures illustrate that the linearity assumption underlying our parametric mobility statistics provide a reasonable approximation to the data and that the stability of the parental income gradient for the birth cohorts 1980-1996 holds for most subgroup with one notable exception: the parental income gradient for children with A-Level educated parents became more muted over time, although with high statistical uncertainty.

Figure 2.B.4: Social Mobility for Subgroups

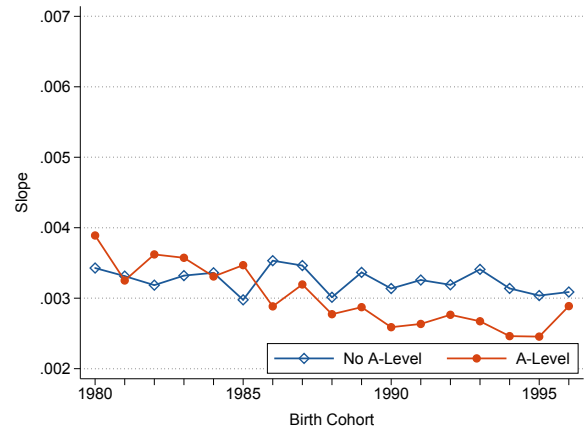


*Notes:* This figure shows the empirical CEFs and the corresponding best linear approximations for each of the subgroups discussed in Section 2.4.1 based on the MZ waves 2011-2018.

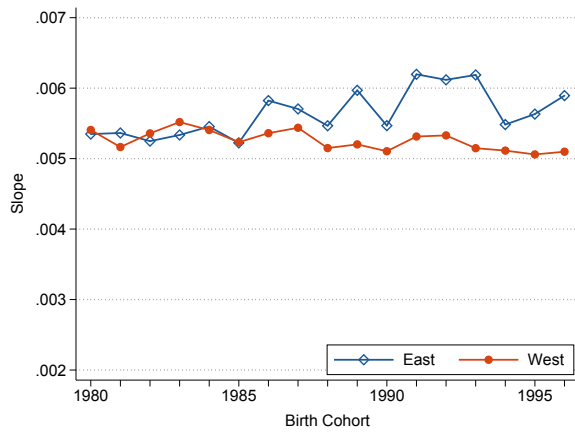
Figure 2.B.5: Trends in Parental Income Gradients by Subgroups



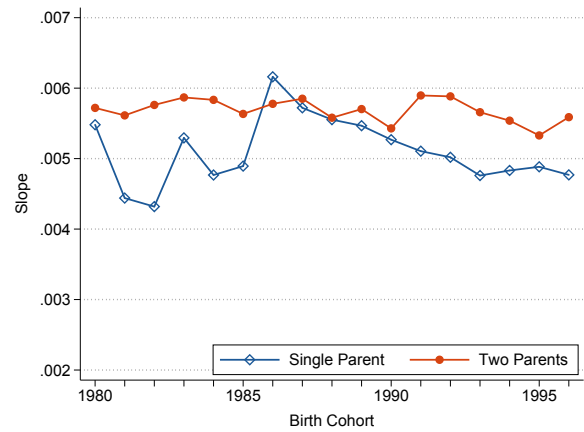
(a) Gender



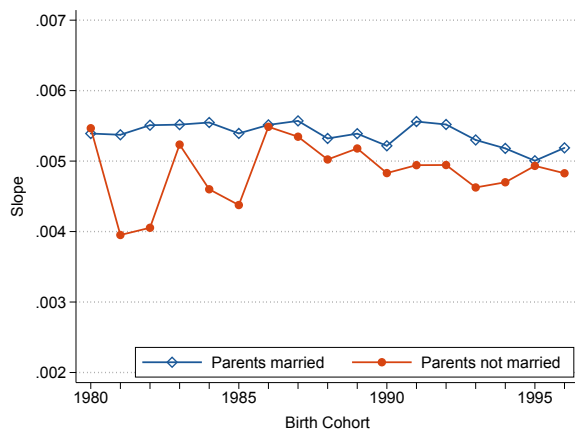
(b) Parental Education



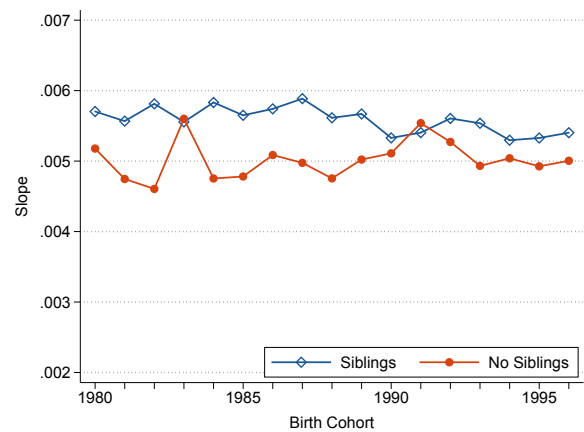
(c) Region



(d) Parenting Status



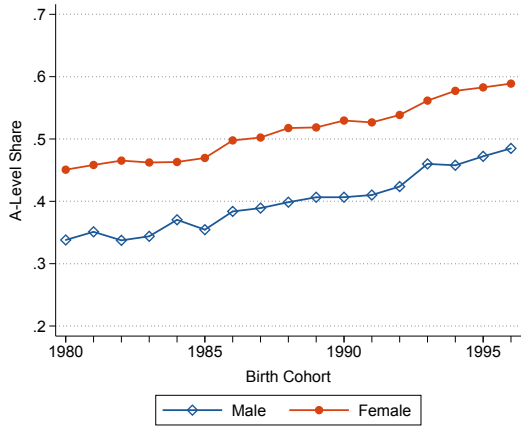
(e) Parental Marital Status



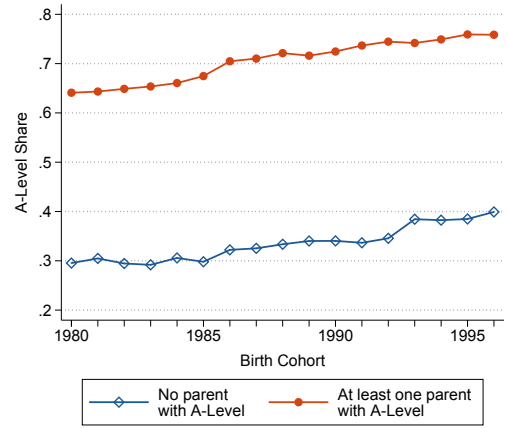
(f) Siblings

*Notes:* This figure displays our point estimates for the parental income gradient by birth cohort for the population subgroups discussed in Section 2.4.1 and the birth cohorts 1980-1996. The estimates are based on all children belonging to the respective birth cohorts observed in the MZ data.

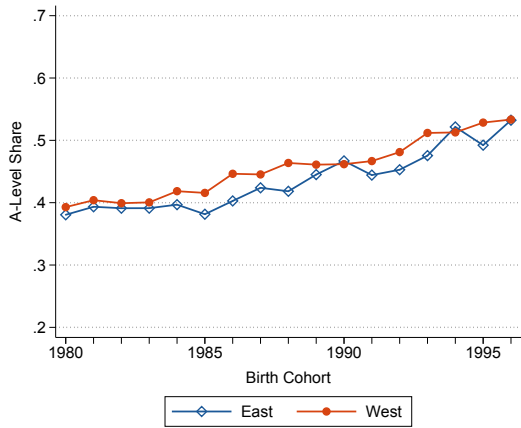
Figure 2.B.6: Trends in A-Level Shares by Subgroups



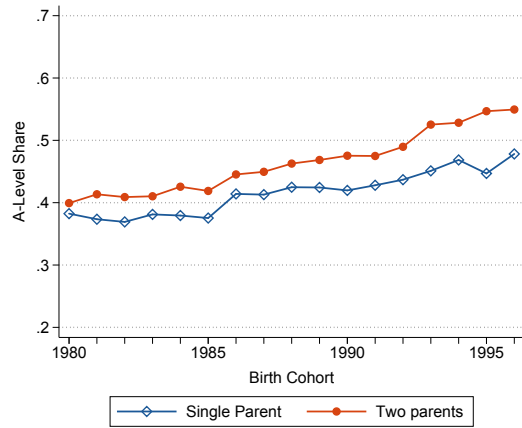
(a) Gender



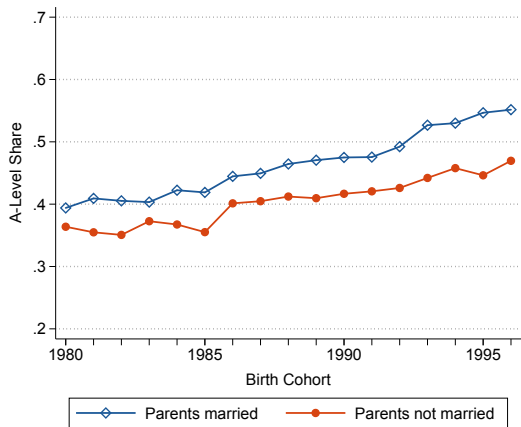
(b) Parental Education



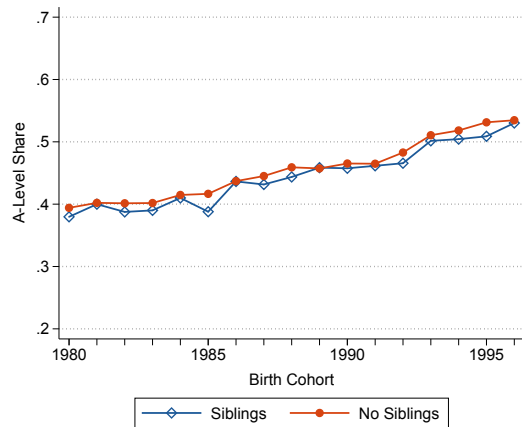
(c) Region



(d) Parenting Status



(e) Parental Marital Status



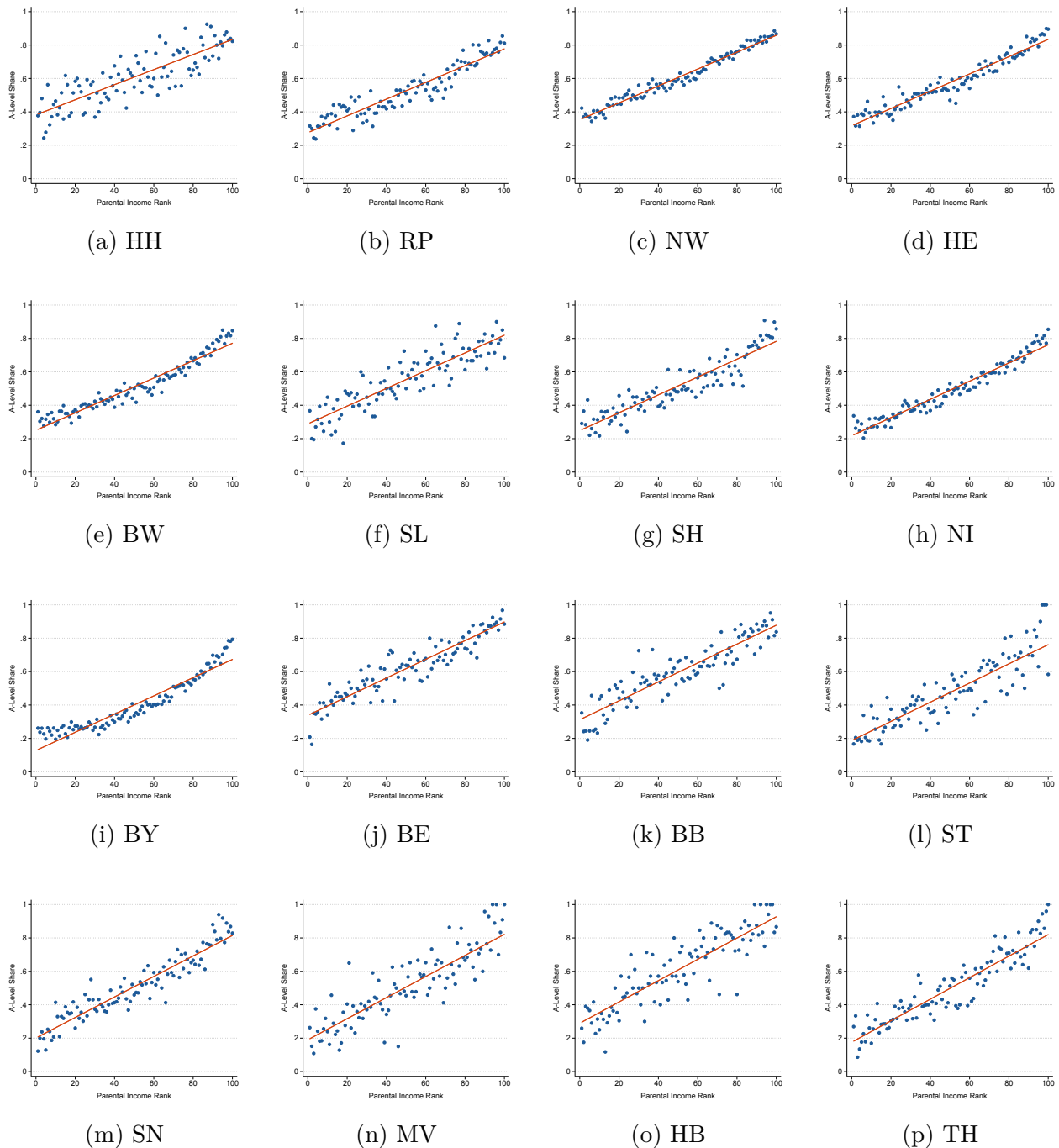
(f) Siblings

*Notes:* This figure shows the development of the A-Level share for different population subgroups for birth cohorts 1980-1996 in the MZ. The A-Level share is given as the fraction of children aged 17-21 that are either enrolled in the last two/three years of the A-Level track or already completed an A-Level degree.

## 2.C Additional Material: Regional Estimates

### 2.C.1 State Level CEFs

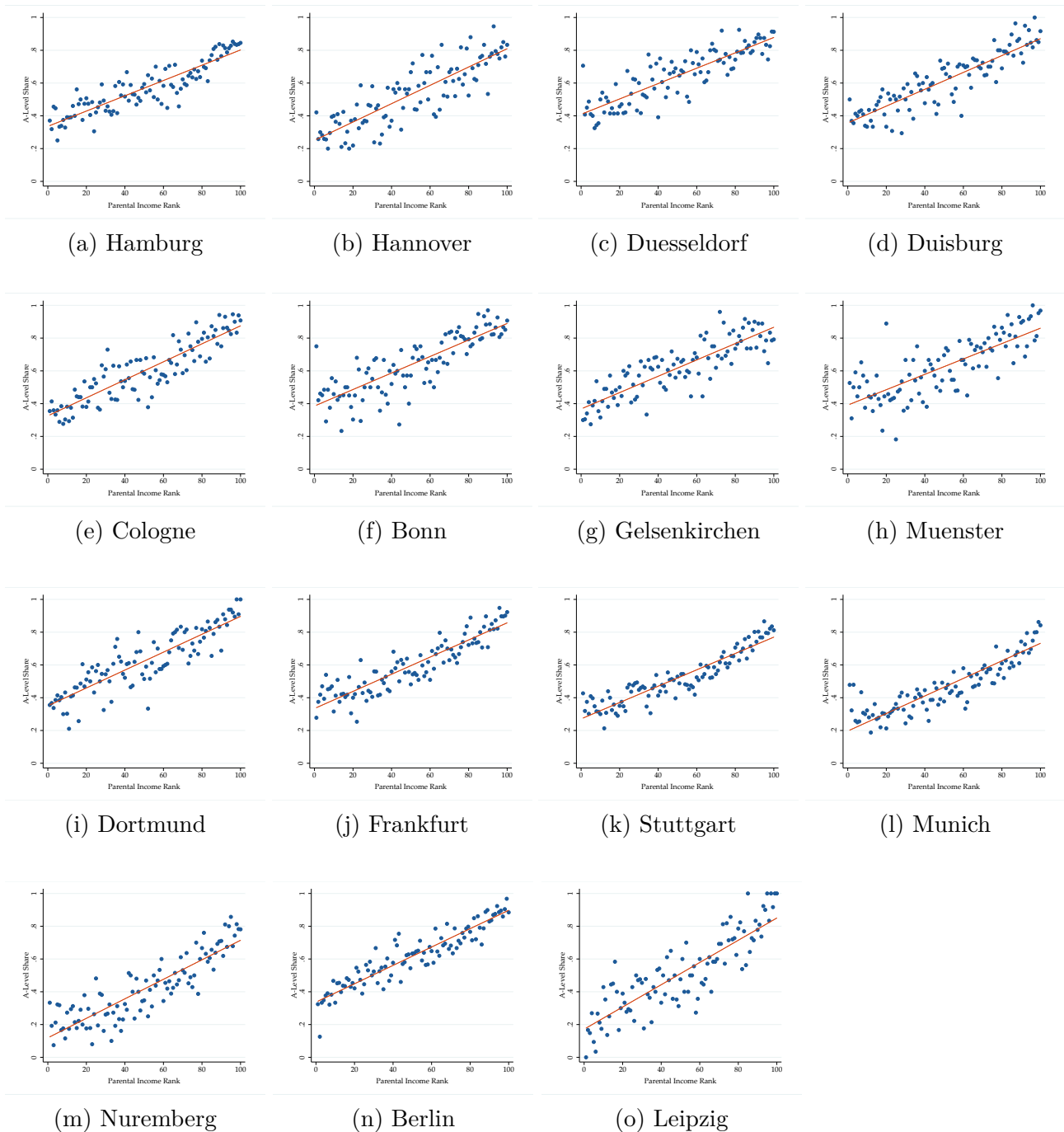
Figure 2.C.1: Social Mobility at the State Level



*Notes:* This figure shows the empirical CEFs and corresponding best linear approximations for each of the 16 German states based on the MZ waves 2011-2018. The point estimates and standard errors are reported in Table 2.6. Children are assigned to a state based on the location of their household as reported in our data.

## 2.C.2 City Level CEFs

Figure 2.C.2: Social Mobility in the 15 Largest Local Labor Markets

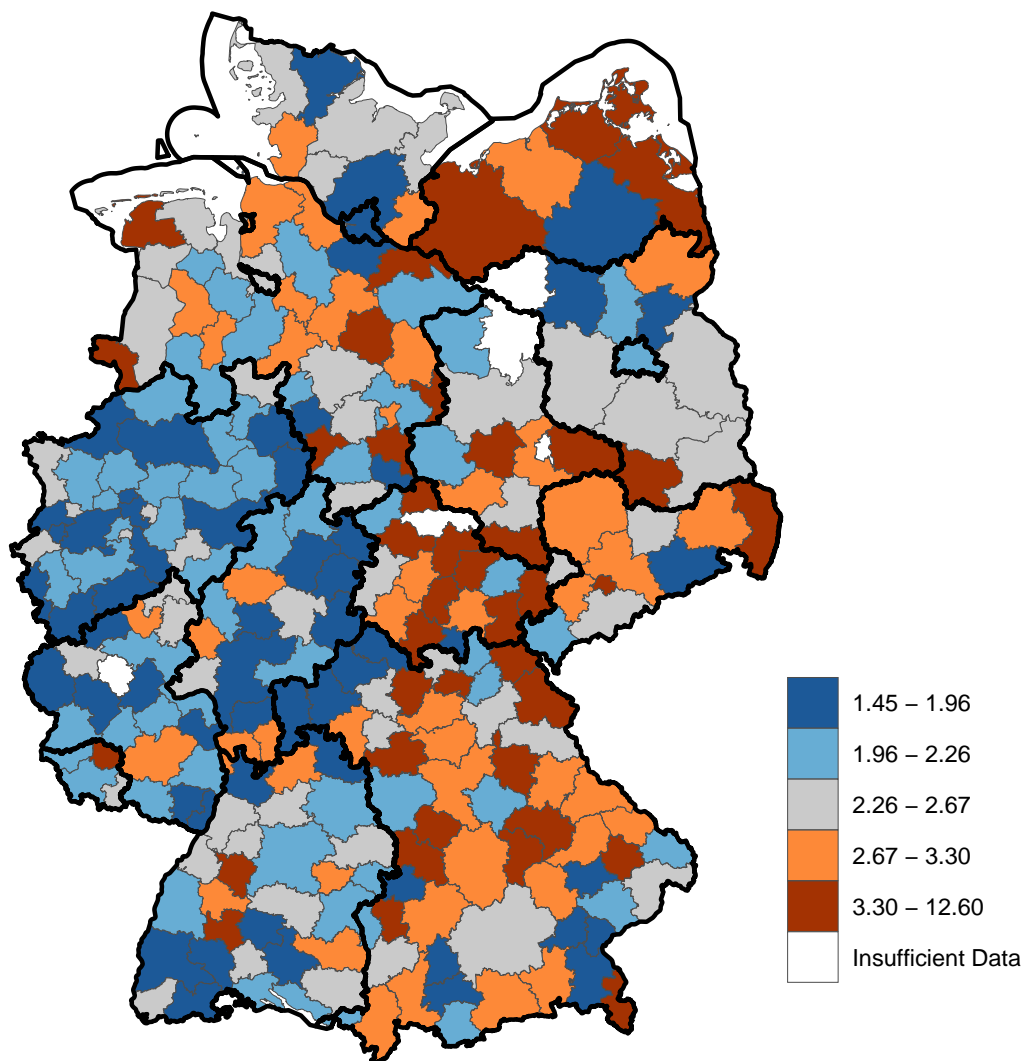


*Notes:* This figure shows the empirical CEFs and corresponding best linear approximations for the 15 largest local labor markets in Germany based on the MZ waves 2011-2018. The point estimates and standard errors are reported in Table 2.7. Children are assigned to a local labor market based on the location of their household as reported in our data.

### 2.C.3 Q5/Q1 Ratio by Local Labor Market

Figure 2.C.3 displays the Q5/Q1 ratio by local labor market. Due to data privacy restrictions we are unable to report the ratio for 6 local labor markets with insufficient number of children observed in one of the two relevant quintiles of the income distribution. Note that, in contrast to the maps reported in the main body of the paper, the colors indicate the quintile of the respective LLM-level estimate in the distribution of estimates. The schema is chosen to account for outliers with extreme values of the Q5/Q1 ratio induced by small denominators.

Figure 2.C.3: Q5/Q1 Ratio by Local Labor Market



*Notes:* This figure presents a heat map of the Q5/Q1 ratio by local labor market. Children are assigned to labor markets according to their place of residence at observation. The estimates are based on the MZ waves 2011-2018. The Q5/Q1 ratio is computed by dividing the share of children with an A-Level degree in the top 20% through the share of children with an A-Level degree in the bottom 20% of the parental income distribution.

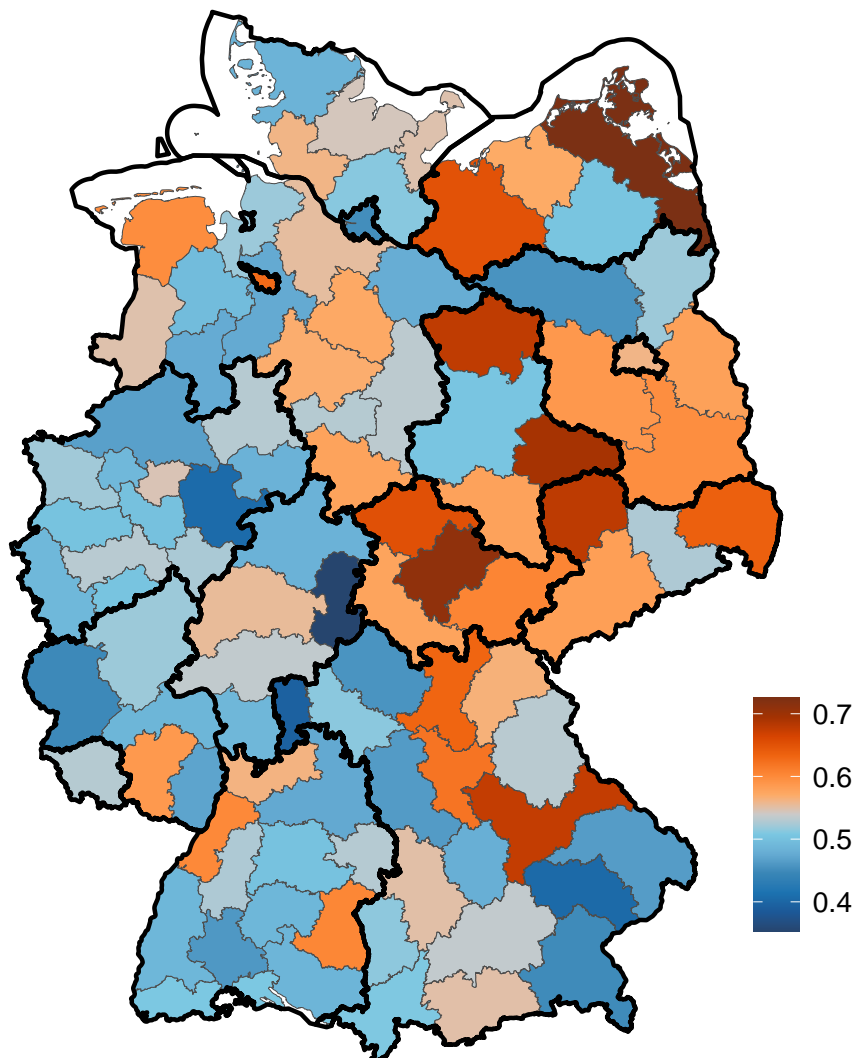


## 2.C.4 Mobility Estimates by Planning Region

The figures presented below display our mobility estimates for the 96 spatial planning regions of Germany. Spatial planning regions (SPRs) are aggregations of commuting zones. The median number of observations per spatial planning region is 1741 (mean: 2406). The maps demonstrate that the geographic patterns documented and described in Section 2.5.3 can also be found when estimation uncertainty is less of a concern.

**Gradient by SPR** Figure 2.C.4 presents SPR-level estimates of the parental income gradient.

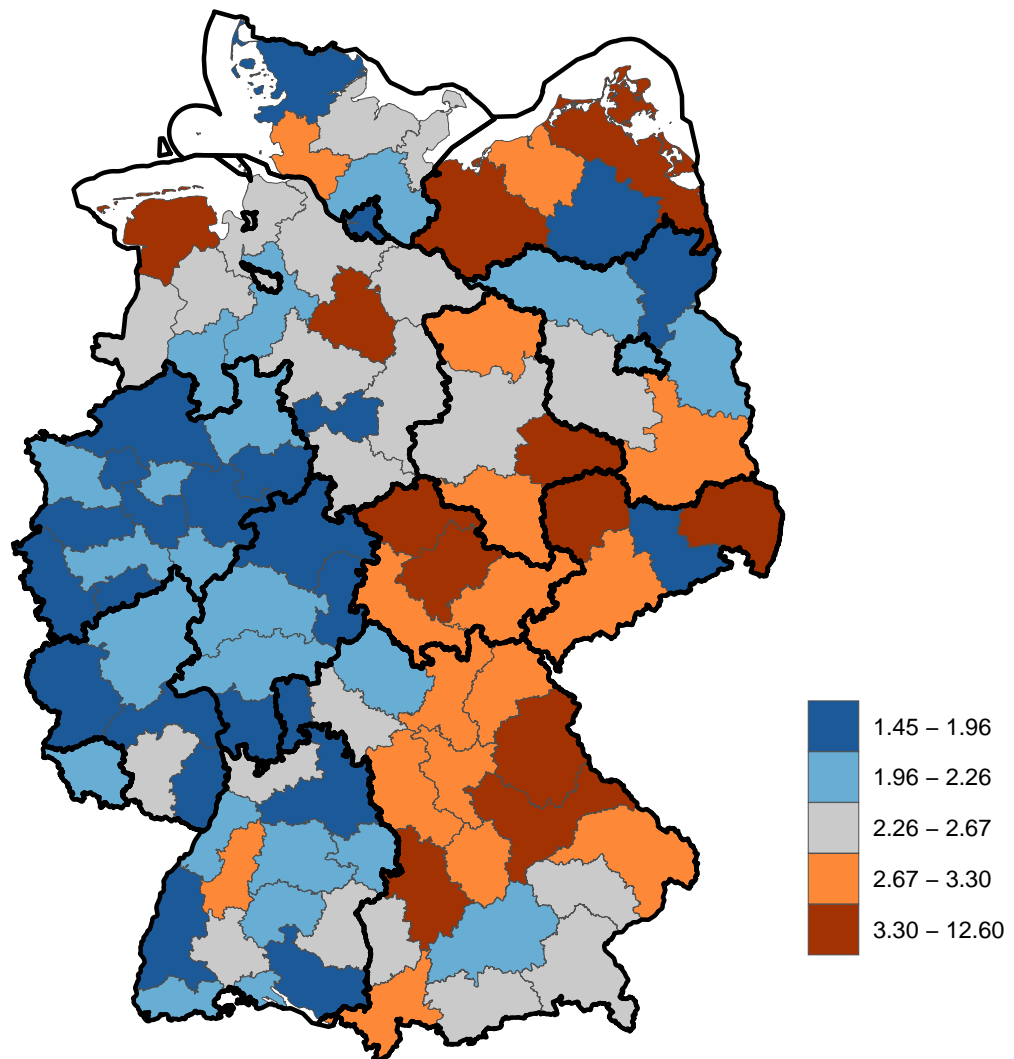
Figure 2.C.4: Parental Income Gradient by Spatial Planning Region



*Notes:* This figure presents a heat map of the parental income gradient for the 96 spatial planning regions of Germany. Children are assigned to spatial planning regions according to their current place of residence. The estimates are based on children aged 17-21 in the MZ waves 2011-2018.

**Q5/Q1 Ratio by SPR** 2.C.5 presents SPR-level estimates of the Q5/Q1 Ratio. Note that, in contrast to all other SPR-level maps, the colors indicate the quintile of the respective SPR-level estimate in the distribution of estimates. The schema is chosen to account for outliers with extreme values of the Q5/Q1 ratio induced by small denominators.

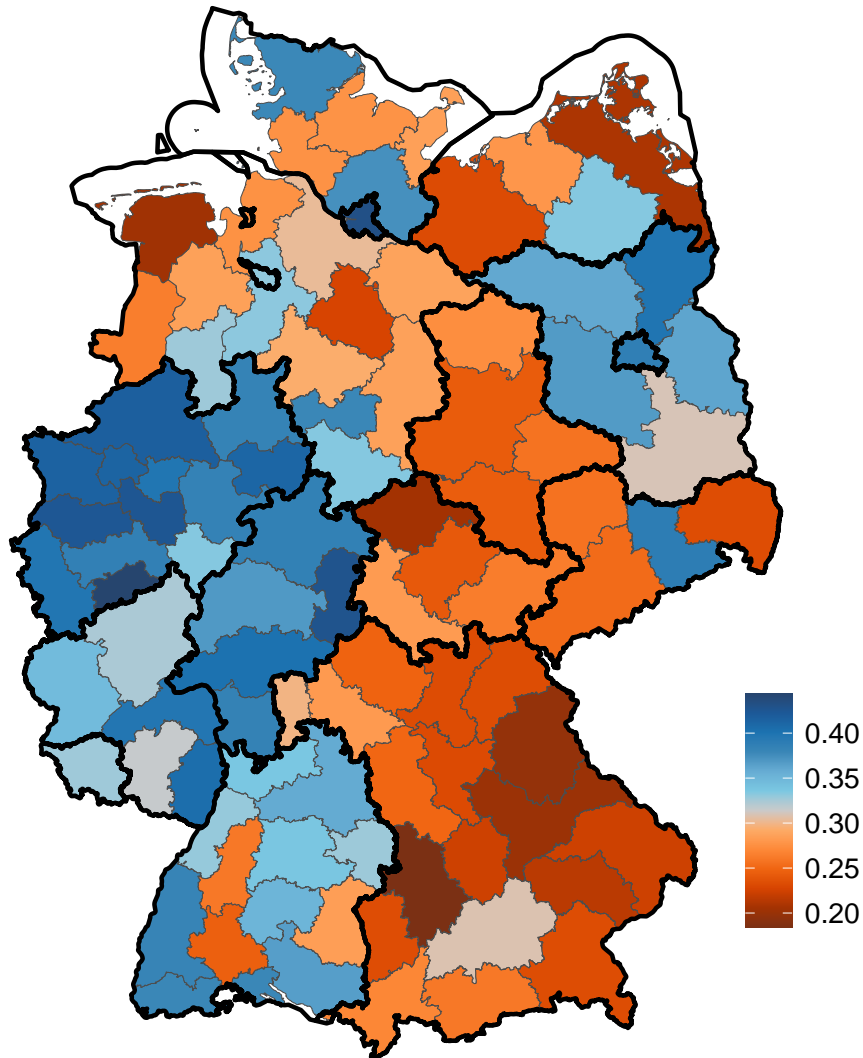
Figure 2.C.5: Q5/Q1 Ratio by Spatial Planning Region



*Notes:* This figure presents a heat map of the Q5/Q1 Ratio for the 96 spatial planning regions of Germany. Children are assigned to spatial planning regions according to their current place of residence. The estimates are based on children aged 17-21 in the MZ waves 2011-2018.

**Q1 Measure by SPR** 2.C.6 presents SPR-level estimates of the Q1 Measure.

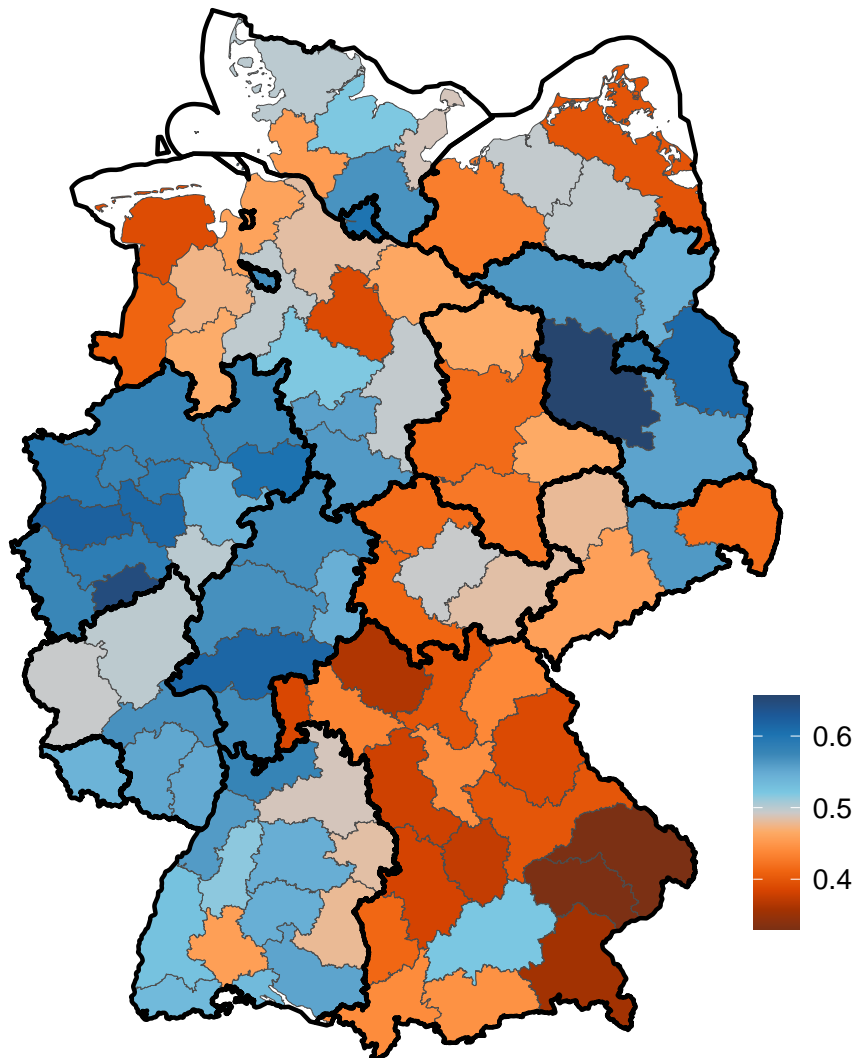
Figure 2.C.6: Q1 Measure by Spatial Planning Region



*Notes:* This figure presents a heat map of the Q1 Measure for the 96 spatial planning regions of Germany. Children are assigned to spatial planning regions according to their current place of residence. The estimates are based on children aged 17-21 in the MZ waves 2011-2018.

A-Level Share by SPR 2.C.7 presents SPR-level estimates of the A-Level share.

Figure 2.C.7: A-Level Share by Spatial Planning Region

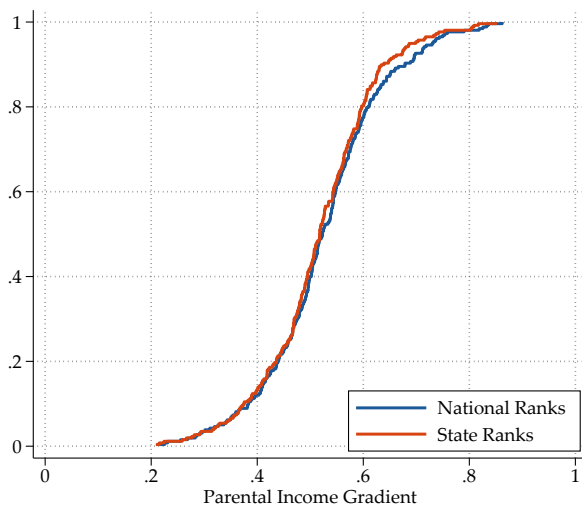


*Notes:* This figure presents a heat map of the A-Level share for the 96 spatial planning regions of Germany. Children are assigned to spatial planning regions according to their current place of residence. The estimates are based on children aged 17-21 in the MZ waves 2011-2018.

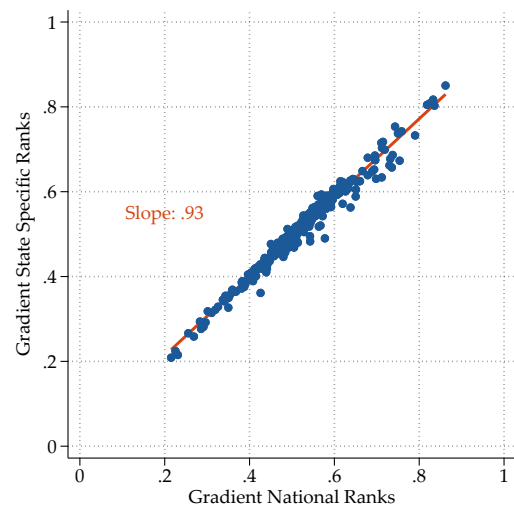
## 2.C.5 Alternative Reference Distributions

Figure 2.C.8 displays the CDFs and scatter plots of our LLM-level point estimates for the parental income gradient based on different reference distributions. The figure shows that changing the reference distribution from the national to the state or regional distribution has negligible impact on our point estimates, implying that regional differences are unlikely to be driven by differences in the shape of marginal income distributions across local labor markets.

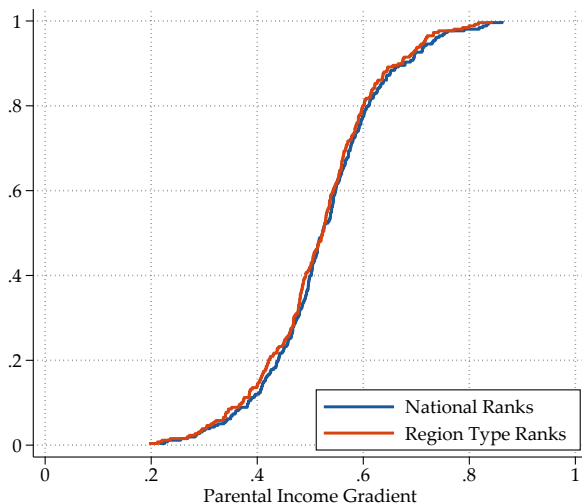
Figure 2.C.8: State and Region Specific Parental Income Ranks



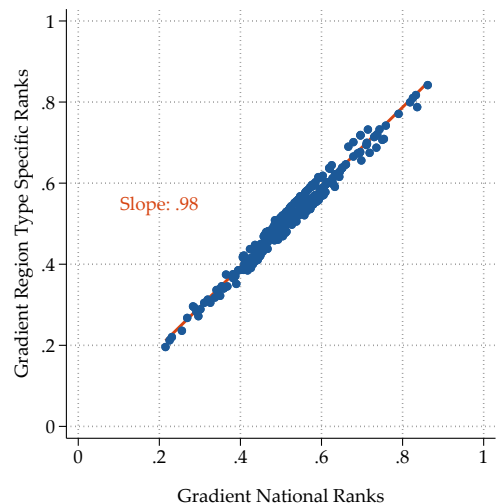
(a) CDFs



(b) Scatter Plot National/State



(c) CDFs



(d) Scatter Plot National/Region

*Notes:* This figure displays the sensitivity of our LLM-level gradient point estimates with respect to the reference income distribution. While the upper panels (a) and (b) compare gradients computed based on the national and state income distributions, the bottom panels (c) and (d) compare the gradients obtained by computing income ranks based on the national and regional income distribution. The reported slope parameters of 0.93 and 0.98 correspond to the OLS slope estimates obtained by regressing the gradients obtained using the respective local ranks on the gradients obtained by using national income ranks.

## 2.C.6 Correlation between Mobility Estimates

Table 2.C.1 reports the pairwise correlations between our point estimates of the local labor market-level mobility measures, as well as the local A-Level share.

Table 2.C.1: Correlation between Mobility Statistics

Measure	Corr.	A-Level	Q1	Q5/Q1	Gradient
A-Level	$\rho$	1	-	-	-
	$r$	1	-	-	-
Q1	$\rho$	0.75	1	-	-
	$r$	0.77	1	-	-
Q5/Q1	$\rho$	-0.39	-0.73	1	-
	$r$	-0.45	-0.84	1	-
Gradient	$\rho$	-0.03	-0.47	0.66	1
	$r$	-0.09	-0.49	0.77	1

*Notes:* This table reports the correlations between our estimates of the different measures of social mobility across local labor markets.  $\rho$  denotes the Pearson correlation coefficient of two measures across local labor markets and  $r$  denotes the Spearman rank correlation coefficient.

As discussed in the main body of the paper, the A-Level share is strongly correlated with the Q1 measure. As a consequence of the approximate proportionality between the A-Level share and the quintile shares, the Q5/Q1 ratio is negatively correlated with the A-Level share. In contrast, there is little linear dependence between the A-Level share and the parental income gradient, as the A-Level shares are sufficiently far from 0 and 1, so that there is little mechanical correlation. Altogether, these correlations mirror our considerations regarding the time trends in the previous section. While the A-Level share is strongly linked to the quantile measures, its correlation with the parental income gradient is negligible.

## 2.C.7 Regional Indicators

Table 2.C.2 displays the 71 regional indicators that were used as predictors in the Random Forest algorithm discussed in Section 2.5.3, as well as the respective data source. The data was obtained as follows: In a first step, we retrieved data from the Federal Institute for Building, Urban Affairs and Spatial Research (BBSR), which maintains the INKAR database of regional indicators (<https://www.inkar.de/>). These data are collected from various government bodies in Germany, including the German Statistical Office Destatis and the Institute for Employment Research (IAB). We select all indicators which we think may potentially be relevant for mobility and are not too collinear to each other (for example, we do not include the general

unemployment rate and the unemployment rates among males and females at the same time). In a second step, we add data from Destatis publications with information on the share of Gymnasium students among all secondary school students and compute the distance of the geographical center of each LLM to the next college based on data from the website of the

Table 2.C.2: List of Regional Indicators

Category	Variable	Source
Labor Market	Unemployment Rate	INKAR
	Share Long Term Unemployed	INKAR
	Share Female Employees	INKAR
	Share Part Time Employees	INKAR
	Share without Vocational qualification	INKAR
	Share Marginal Employment	INKAR
	Share Employed in Manufacturing Sector	INKAR
	Apprenticeship Positions	INKAR
	Apprentices	INKAR
	Vocational School Students	INKAR
	Employees with Academic Degree	INKAR
	Commuting Balance	INKAR
	Hours Worked	INKAR
	A-Level Wage Premium	MZ
Education	Students (before Tertiary Education)	INKAR
	Students (Tertiary Education)	INKAR
	Students (Universities of Applied Sciences)	INKAR
	School Dropout Rate	INKAR
	Highly Qualified Persons	INKAR
	Share Children 0-2 in Childcare	INKAR
	Share Children 3-5 in Childcare	INKAR
	Share of all Students Enrolled in Gymnasium	INKAR
	Share of all Secondary School Students Enrolled in Gymnasium	Destatis
	Distance to Next College	HRK
	Distance to Next Elementary School	INKAR
Share on Vocational A-Level Track	MZ	
Economy	GDP per Capita	INKAR
	Municipal Tax Revenues per Capita	INKAR
	Municipal Debt per Capita	INKAR
	Business Creation	INKAR
Housing	Construction Land Prices	INKAR
	New Apartments	INKAR
	Building Permits	INKAR
	Living Area	INKAR
	Share of Apartment Buildings	INKAR
	Rent Prices	INKAR

	Median Household Income	INKAR
	Median Household Income with Vocational Qualification	INKAR
	Gender Wage Gap	INKAR
	Child Poverty	INKAR
Income	Mean Household Income	INKAR
	Gini Household Income	MZ
	Expected Rank Difference Parental Income	MZ
	Mean Parental Income	MZ
	Gini Parental Income	MZ
	Ratio p85/p50 (Household Income)	MZ
	Ratio p50/p15 (Household Income)	MZ
	Physician Density	INKAR
Infrastructure	Broad Band Availability	INKAR
	Passenger Car Density	INKAR
	Hospital Beds	INKAR
	Average Age	INKAR
	Share Female	INKAR
	Share Foreigners	INKAR
	Share Asylum Seekers	INKAR
	Total Net Migration	INKAR
	Births Net of Deaths	INKAR
Demographics	Fertility Rate	INKAR
	Teenage Pregnancies	INKAR
	Life Expectancy	INKAR
	Child Mortality	INKAR
	Population Density	INKAR
	Share Single Parents	MZ
	Share Married	MZ
	Share Divorced	MZ
	Voter Turnout	INKAR
	Vote Share CDU	INKAR
	Vote Share SPD	INKAR
Social	Share Social Assistance	INKAR
	Mean ISEI	MZ
	Gini ISEI	MZ

*Notes:* This table displays all regional indicators considered for our analysis. The third column reports the data source, which is either the INKAR database, the Statistical Office of Germany (Destatis), the Hochschulrektorenkonferenz (HRK) or the Mikrozensus (MZ).

Hochschulrektorenkonferenz (HRK; <https://www.hochschulkompass.de/hochschulen/downloads.html>). In a third step, we compute additional statistics on the LLM level in the MZ data like the Gini coefficient in household income, the local A-Level wage premium or the ISEI (an international index of social status). We construct our final indicators as the time averages of variables over the years 2011 to 2018 at the LLM level.



## 2.C.8 Split Sample Predictors

Table 2.C.3 displays the set of predictors selected by the Random Forest algorithm discussed in Section 2.5.3 when employed separately to our estimates for the 129 largest and smallest local labor markets.

Table 2.C.3: Optimal Predictors of Relative Mobility by Size Category

Variable	Importance Measure	$\rho$
<i>Panel (A): 129 Largest Local Labor Markets</i>		
School Dropout Rate	0.42	+
Gini Parental Income	0.23	–
Share Married	0.16	–
Share without Vocational Qualification	0.10	–
Students	0.09	–
Physician Density	0.09	+
Teenage Pregnancies	0.06	+
Mean Parental Income	0.06	–
Share Marginal Employment	0.06	–
Students (Universities of Applied Sciences)	0.05	–
Median Income Vocational Qualification	0.05	–
Distance to next Elementary School	0.03	–
Share Children 0-2 in Childcare	0.03	+
Child Mortality	0.03	–
Ratio p50/p15	0.03	–
<i>Panel (B): 129 Smallest Local Labor Markets</i>		
School Dropout Rate	0.75	+
Unemployment Rate	0.45	+
Child Poverty	0.40	+
Students	0.40	–
Share Married	0.33	–
Teenage Pregnancies	0.33	+
Median Income Vocational Qualification	0.19	–
Gender Wage Gap	0.19	+
Share Social Assistance	0.18	+
Total Net Migration	0.12	–
Highly Qualified Persons	0.10	+
Broadband Availability	0.10	+
Share on Vocational A-Level Track	0.08	–
Building Permits	0.08	–
Share Apartment Buildings	0.07	+

*Notes:* This table lists the set of the 15 most predictive indicators for explaining between LLM variation in our estimates of the parental income gradient, separately for the 129 largest (Panel A) and the 129 smallest (Panel B) local labor markets. The last column shows the sign of the Pearson correlation coefficient between each variable and our estimates of the parental income gradients.



# References

- Abadie, Alberto and Guido W Imbens (2006). “Large Sample Properties of Matching Estimators for Average Treatment Effects”. In: *Econometrica* 74.1, pp. 235–267.
- Abadie, Alberto, Guido W Imbens, and Fanyin Zheng (2014). “Inference for Misspecified Models With Fixed Regressors”. In: *Journal of the American Statistical Association* 109.508, pp. 1601–1614.
- Acciari, Paolo, Alberto Polo, and Giovanni L. Violante (2019). “”And Yet, It Moves”: Intergenerational Mobility in Italy”. In: *NBER Working Paper No. 25732*.
- Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou (2021). “Intergenerational Mobility in Africa”. In: *Econometrica* 89.1, pp. 1–35.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). “Intergenerational Mobility and Preferences for Redistribution”. In: *American Economic Review* 108.2, pp. 521–54.
- Altonji, Joseph G and Rosa L Matzkin (2005). “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors”. In: *Econometrica* 73.4, pp. 1053–1102.
- Anderson, Theodore W and Herman Rubin (1949). “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations”. In: *The Annals of Mathematical Statistics* 20.1, pp. 46–63.
- Armstrong, Timothy B and Michal Kolesár (2018). “Optimal Inference in a Class of Regression Models”. In: *Econometrica* 86.2, pp. 655–683.
- (2020). “Simple and honest confidence intervals in nonparametric regression”. In: *Quantitative Economics* 11.1, pp. 1–39.

- Asher, Sam, Paul Novosad, and Charlie Rafkin (2020). “Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities”. In: *Unpublished Manuscript*. URL: <http://paulnovosad.com/pdf/anr-india-mobility.pdf>.
- Becker, Gary S and Nigel Tomes (1986). “Human Capital and the Rise and Fall of Families”. In: *Journal of Labor Economics* 4.3, Part 2, pp. S1–S39.
- Bertrand, Marianne, Magne Mogstad, and Jack Mountjoy (2021). “Improving Educational Pathways to Social Mobility: Evidence from Norway’s Reform 94”. In: *Journal of Labor Economics* 39.4, pp. 965–1010.
- Biewen, Martin and Madalina Tapalaga (2017). “Life-cycle educational choices in a system with early tracking and ‘second chance’ options”. In: *Economics of Education Review* 56, pp. 80–94.
- Boar, Corina and Danial Lashkari (2021). “Occupational Choice and the Intergenerational Mobility of Welfare”. In: *NBER Working Paper No. 29381*.
- Boyd, Stephen and Lieven Vandenberghe (2004). *Convex optimization*. Vol. 1. Cambridge University Press Cambridge.
- Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel Schnitzlein, and Kjell Vaage (2017). “A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US”. In: *The Scandinavian Journal of Economics* 119.1, pp. 72–101.
- Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell (2018). “Coverage Error Optimal Confidence Intervals for Local Polynomial Regression”. In: *arXiv preprint arXiv:1808.01398*.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs”. In: *Econometrica* 82.6, pp. 2295–2326.
- Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei (2015a). “The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013”. In: *American Economic Review* 105.5, pp. 126–30.

- Card, David, David S Lee, Zhuan Pei, and Andrea Weber (2015b). “Inference on Causal Effects in a Generalized Regression Kink Design”. In: *Econometrica* 83.6, pp. 2453–2483.
- (2017). *Regression kink design: Theory and practice*. Emerald Publishing Limited.
- Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.
- Cheng, Ming-Yen, Jianqing Fan, and James S Marron (1997). “On automatic boundary corrections”. In: *The Annals of Statistics* 25.4, pp. 1691–1708.
- Chetty, Raj (2009). “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods”. In: *Annual Review of Economics* 1.1, pp. 451–488.
- Chetty, Raj and Nathaniel Hendren (2018a). “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects”. In: *Quarterly Journal of Economics* 133.3, pp. 1107–1162.
- (2018b). “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates”. In: *Quarterly Journal of Economics* 133.3, pp. 1163–1228.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. In: *Quarterly Journal of Economics* 129.4, pp. 1553–1623.
- Chuard, Patrick and Veronica Grassi (2020). “Switzer-Land of Opportunity: Intergenerational Income Mobility in the Land of Vocational Education”. In: *University of St. Gallen Discussion Paper* 2020-11.
- Corak, Miles (2020). “The Canadian Geography of Intergenerational Income Mobility”. In: *The Economic Journal* 130.631, pp. 2134–2174.
- Dahl, Molly and Thomas DeLeire (2008). *The Association between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data*. University of Wisconsin-Madison, Institute for Research on Poverty.
- Dahrendorf, Ralf (1965). *Bildung ist Bürgerrecht. Plädoyer für eine aktive Bildungspolitik*. Hamburg: Nannen-Verlag.

- De Brabanter, Kris, Jos De Brabanter, and Bart De Moor (2013). “On Asymptotic Properties of Linear Smoothers”. In: *unpublished manuscript*.
- Deutscher, Nathan and Bhashkar Mazumder (2020). “Intergenerational mobility across Australia and the stability of regional estimates”. In: *Labour Economics* 66, p. 101861.
- Donoho, David L (1994). “Statistical Estimation and Optimal Recovery”. In: *The Annals of Statistics* 22.1, pp. 238–270.
- Dustmann, Christian (2004). “Parental background, secondary school track choice, and wages”. In: *Oxford Economic Papers* 56.2, pp. 209–230.
- Dustmann, Christian, Patrick Puhani, and Uta Schönberg (2017). “The Long-Term Effects of Early Track Choice”. In: *The Economic Journal* 127.603, pp. 1348–1380.
- Eberharter, Veronika V (2013). “The Intergenerational Dynamics of Social Inequality – Empirical Evidence from Europe and the United States”. In: *SOEPpapers* 588-2013.
- Eisenhauer, Philipp and Friedhelm Pfeiffer (2008). “Assessing Intergenerational Earnings Persistence Among German Workers”. In: *Journal of Labour Market Research* 2/3, pp. 119–137.
- Fan, Jianqing (1993). “Local Linear Regression Smoothers and Their Minimax Efficiencies”. In: *The Annals of Statistics*, pp. 196–216.
- Fan, Jianqing, Theo Gasser, Irène Gijbels, Michael Brockmann, and Joachim Engel (1997). “Local Polynomial Regression: Optimal Kernels and Asymptotic Minimax Efficiency”. In: *Annals of the Institute of Statistical Mathematics* 49.1, pp. 79–99.
- Fan, Jianqing and Irene Gijbels (1996). *Local Polynomial Modelling and Its Applications*. 1st ed. Routledge.
- Florens, Jean-Pierre, James J Heckman, Costas Meghir, and Edward Vytlacil (2008). “Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects”. In: *Econometrica* 76.5, pp. 1191–1206.
- Freyberger, Joachim and Brandon Reeves (2018). “Inference under Shape Restrictions”. In: *Available at SSRN 3011474*.

- Gallagher, Ryan, Robert Kaestner, and Joseph Persky (2019). “The geography of family differences and intergenerational mobility”. In: *Journal of Economic Geography* 19.3, pp. 589–618.
- Ganong, Peter and Simon Jäger (2018). “A Permutation Test for the Regression Kink Design”. In: *Journal of the American Statistical Association* 113.522, pp. 494–504.
- Gärtner, Karla (2002). “Differentielle Sterblichkeit. Ergebnisse des Lebenserwartungssurveys”. In: *Zeitschrift für Bevölkerungswissenschaft* 27.2, pp. 185–211.
- Hadjar, Andreas and Rolf Becker (2006). “Bildungsexpansion – erwartete und unerwartete Folgen”. In: *Die Bildungsexpansion*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 11–24.
- Hall, Peter (1992). “Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density”. In: *The Annals of Statistics*, pp. 675–694.
- Hausner, Karl Heinz, Doris Söhnlein, Brigitte Weber, and Enzo Weber (2015). “Qualifikation und Arbeitsmarkt: Bessere Chancen mit mehr Bildung”. In: *IAB Kurzbericht No. 11/2015*.
- Helbig, Marcel and Rita Nikolai (2015). *Die Unvergleichbaren: der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Verlag Julius Klinkhardt.
- Hilger, Nathaniel G. (2015). “The Great Escape: Intergenerational Mobility in the United States Since 1940”. In: *NBER Working Paper No. 21217*.
- Ignatiadis, Nikolaos and Stefan Wager (2019). “Bias-Aware Confidence Intervals for Empirical Bayes Analysis”. In: *arXiv e-prints*, arXiv–1902.
- Imbens, Guido and Karthik Kalyanaraman (2012). “Optimal Bandwidth Choice for the Regression Discontinuity Estimator”. In: *The Review of Economic Studies* 79.3, pp. 933–959.
- Imbens, Guido and Stefan Wager (2019). “Optimized Regression Discontinuity Designs”. In: *Review of Economics and Statistics* 101.2, pp. 264–278.
- Imbens, Guido W and Thomas Lemieux (2008). “Regression discontinuity designs: A guide to practice”. In: *Journal of Econometrics* 142.2, pp. 615–635.

- Imbens, Guido W and Charles F Manski (2004). “Confidence Intervals for Partially Identified Parameters”. In: *Econometrica* 72.6, pp. 1845–1857.
- Kalleberg, Arne L (1977). “Work Values and Job Rewards: A Theory of Job Satisfaction”. In: *American Sociological Review* 42.1, pp. 124–143.
- (2011). *Good Jobs, Bad Jobs: The Rise of Polarized and Precarious Employment Systems in the United States, 1970s-2000s*. Russell Sage Foundation.
- Klein, Markus, Katherin Barg, and Michael Kühhirt (2019). “Inequality of Educational Opportunity in East and West Germany: Convergence or Continued Differences?” In: *Sociological Science* 6, pp. 1–26.
- Kline, Patrick and Enrico Moretti (2014). “People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs”. In: *Annual Review of Economics* 6.1, pp. 629–662.
- Kolesár, Michal and Christoph Rothe (2018). “Inference in Regression Discontinuity Designs with a Discrete Running Variable”. In: *American Economic Review* 108.8, pp. 2277–2304.
- Kolsrud, Jonas, Camille Landais, Peter Nilsson, and Johannes Spinnewijn (2018). “The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden”. In: *American Economic Review* 108.4-5, pp. 985–1033.
- Landais, Camille (2015). “Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design”. In: *American Economic Journal: Economic Policy* 7.4, pp. 243–78.
- Landersø, Rasmus and James J Heckman (2017). “The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US”. In: *The Scandinavian Journal of Economics* 119.1, pp. 178–230.
- Lee, Chul-In and Gary Solon (2009). “Trends in Intergenerational Income Mobility”. In: *The Review of Economics and Statistics* 91.4, pp. 766–772.
- Lee, David S and Thomas Lemieux (2010). “Regression Discontinuity Designs in Economics”. In: *Journal of Economic Literature* 48.2, pp. 281–355.



- Li, Ker-Chau (1989). “Honest Confidence Regions for Nonparametric Regression”. In: *The Annals of Statistics* 17.3, pp. 1001–1008.
- Liang, Kung-Yee and Scott L Zeger (1986). “Longitudinal data analysis using generalized linear models”. In: *Biometrika* 73.1, pp. 13–22.
- Low, Mark G (1997). “On nonparametric confidence intervals”. In: *The Annals of Statistics* 25.6, pp. 2547–2554.
- Mazumder, Bhashkar (2005). “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data”. In: *The Review of Economics and Statistics* 87.2, pp. 235–255.
- (2018). “Intergenerational Mobility in the United States: What We Have Learned from the PSID”. In: *The Annals of the American Academy of Political and Social Science* 680.1, pp. 213–234.
- Mottaz, Clifford J (1985). “The Relative Importance of Intrinsic and Extrinsic Rewards as Determinants of Work Satisfaction”. In: *The Sociological Quarterly* 26.3, pp. 365–385.
- Muñoz, Ercio (2021). “The Geography of Intergenerational Mobility in Latin America and the Caribbean”. In: *Stone Center Working Paper Series* 29.
- Neumark, David and Helen Simpson (2015). “Place-Based Policies”. In: *Handbook of Regional and Urban Economics*. Vol. 5. Elsevier, pp. 1197–1287.
- Noack, Claudia and Christoph Rothe (2019). “Bias-Aware Inference in Fuzzy Regression Discontinuity Designs”. In: *arXiv preprint arXiv:1906.04631*.
- Nybom, Martin and Jan Stuhler (2017). “Biases in Standard Measures of Intergenerational Income Dependence”. In: *Journal of Human Resources* 52.3, pp. 800–825.
- Picht, Georg (1964). *Die deutsche Bildungskatastrophe: Analyse und Dokumentation*. Olten Freiburg im Breisgau: Walter.
- Rambachan, Ashesh and Jonathan Roth (2019). “An Honest Approach to Parallel Trends”. In: *Unpublished manuscript, Harvard University*. [99].

- Riphahn, Regina and Guido Heineck (2009). “Intergenerational Transmission of Educational Attainment in Germany – The Last Five Decades”. In: *Jahrbücher für Nationalökonomie und Statistik* 229.1, pp. 36–60.
- Riphahn, Regina and Parvati Trübswetter (2013). “The intergenerational transmission of education and equality of educational opportunity in East and West Germany”. In: *Applied Economics* 45.22, pp. 3183–3196.
- Rothbaum, Jonathan (2016). “Sorting and Geographic Variation in Intergenerational Mobility”. In: *Unpublished Manuscript*.
- Ruppert, David and Matthew P Wand (1994). “Multivariate Locally Weighted Least Squares Regression”. In: *The Annals of Statistics* 22.3, pp. 1346–1370.
- Sacks, Jerome and Donald Ylvisaker (1978). “Linear Estimation for Approximately Linear Models”. In: *The Annals of Statistics* 6.5, pp. 1122–1137.
- Schennach, Susanne M (2020). “A Bias Bound Approach to Non-parametric Inference”. In: *The Review of Economic Studies* 87.5, pp. 2439–2472.
- Schimpl-Neimanns, Bernhard (2006). *Zur Datenqualität der Bildungsangaben im Mikrozensus*. Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. URL: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200565>.
- Schmillen, Achim and Heiko Stüber (2014). “Lebensverdienste nach Qualifikation - Bildung Lohnt sich ein Leben Lang”. In: *IAB Kurzbericht No. 1/2014*.
- Schnitzlein, Daniel (2016). “A New Look at Intergenerational Mobility in Germany Compared to the US”. In: *Review of Income and Wealth* 62.4, pp. 650–667.
- Sion, Maurice (1958). “On General Minimax Theorems”. In: *Pacific Journal of Mathematics* 8.1, pp. 171–176.
- Solon, Gary (1992). “Intergenerational Income Mobility in the United States”. In: *American Economic Review* 82.3, pp. 393–408.
- Statistisches Bundesamt (2018). *Mikrozensus 2018. Qualitätsbericht*. Wiesbaden. URL: [https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2018.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2018.pdf?__blob=publicationFile).

- (2019). *Einkommensanalysen mit dem Mikrozensus*. Wiesbaden. URL: [https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2019/03/einkommensanalysen-mikrozensus-032019.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2019/03/einkommensanalysen-mikrozensus-032019.pdf?__blob=publicationFile).
- (2021). *Konsumausgaben von Familien für Kinder: Berechnungen auf der Grundlage der Einkommens- und Verbrauchsstichprobe 2018*. Wiesbaden. URL: [https://www.statistischebibliothek.de/mir/receive/DESerie\\_mods\\_00003263](https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00003263).

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis (2008). “Conditional variable importance for random forests”. In: *BMC Bioinformatics* 9.307, pp. 1–11.

Zimmerman, David J. (1992). “Regression Toward Mediocrity in Economic Stature”. In: *American Economic Review* 82.3, pp. 409–429.



# Curriculum Vitae

---

Majed DODIN

## Education

<b>2015 – 2021</b>	<b>Ph.D.</b> Economics, University of Mannheim
<b>2016 – 2017</b>	Visiting Assistant in Research at Yale University
<b>2013 – 2015</b>	<b>M.Sc.</b> Economics, University of Mannheim
<b>2014 – 2015</b>	Visiting graduate student at UC Berkeley
<b>2009 – 2012</b>	<b>B.Sc.</b> Economics, University of Mannheim
<b>2000 – 2009</b>	Abitur, Marie-Curie-Gymnasium Berlin



# Eidesstattliche Erklärung

---

Ich versichere hiermit, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mannheim, 17. November 2021

Majed Dodin