Essays on Intergenerational Mobility

NHH
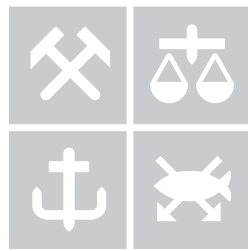
Erling Risa

Acknowledgements

First of all, I would like to thank my main supervisor, Kjell G. Salvanes. You were a big part of the reason why I ended up pursuing a Ph.D. in the first place. You have been generous with your time throughout these years and I have learned a lot from working with you. I also want to thank you for recognizing where my strengths lie and for encouraging me to develop them throughout my Ph.D.

I would further like to extend my gratitude to Aline Bütikofer. You also played an important part in getting me to pursue a Ph.D. Furthermore, you have been very supportive during my Ph.D, both as a co-author and otherwise.

My external supervisor, Mikael Lindahl, also deserves mentioning for providing helpful feedback.

The co-author of the first chapter of this dissertation, Jack Blundell also deserves my gratitude. Writing the chapter together with you was both fun and a great learning experience.

My gratitude also goes to my co-authors in the third chapter, Pedro Carneiro, Sarah Cattan, and Sonya Krutikova.

I further want to thank my colleagues at the Norwegian School of Economics for making my years as a Ph.D candidate both enjoyable and interesting. I especially want to mention the other Ph.D-candidates in my cohort: Félix,

Ingvild, Ole and Sandra. Furthermore, I want to thank Andreas, Evan, Ingar, and Ole for being both outstanding colleagues and friends. I also want to extend a special thanks to my colleagues at the Department of Economics and at FAIR for providing me with an excellent environment during my Ph.D.

I also want to thank my friend Tore for providing me with much-needed distractions from economics through the years. Furthermore, I want to thank my family for their unyielding support both during and before the Ph.D. Finally, I would like to thank my "samboer" Elisa; your support in the completion of this dissertation has been invaluable.

Contents

INTRODUCTION

The focus of this dissertation is intergenerational mobility. One definition of intergenerational mobility is "the relationship between the socioeconomic status of parents and the status their children will attain as adults." In this context, socioeconomic status might refer to several things, but earnings, education, and occupation are typical examples. While there are many reasons why intergenerational mobility is worth studying, perhaps the two most important ones are fairness concerns and efficiency.

Fairness considerations are relevant because some believe that who one's parents are and where one is born should not matter for one's outcomes in life. Consequently, knowledge about the level of intergenerational mobility is necessary to decide whether it is at an acceptable level or whether policy interventions are required.

While people may disagree as to whether or not policy interventions should target intergenerational mobility for fairness reasons, they may still think that it is worth studying for efficiency reasons. The concern in this line of reasoning is that talented individuals are unable to fulfill their potential owing to their socioeconomic background. Such failure to utilize the available human capital in a society is arguably undesirable irrespective of fairness concerns.

This policy relevance has lead intergenerational mobility to be a highly re-

searched topic within economics and other social sciences for a long time. This research has focused on multiple different aspects of social mobility. Parts of the literature have focused on documenting intergenerational persistence along dimensions such as earnings, education, and cognitive ability (e.g. Black, Devereux, and Salvanes 2005; Black, Devereux, and Salvanes 2009; Corak, Lindquist, and Mazumder 2014). Other parts of the literature aim at estimating the causal transmission of socioeconomic status across generations rather than mere correlations.

Another line of research that has gained traction in recent years expands the view beyond the role of parents and looks at the importance of where one grows up (e.g. Chetty, Hendren, Kline, and Saez 2014; Chetty and Hendren 2018a; Chetty and Hendren 2018b; Rothstein 2019). Researchers have also expanded the focus beyond the role of parents by estimating the importance of extended families, including aunts, uncles, and grandparents (e.g. Clark 2014; Güell, Rodríguez Mora, and Telmer 2015; Lindahl, Mårten Palme, Massih, and Sjögren 2015; Solon 2018; Adermon, Lindahl, and Marten Palme 2019).

A final strand of the literature that needs mentioning expands the focus beyond linear relationships and considers non-linear effects and the interplay between the various aspects of individuals' backgrounds in determining their outcomes (e.g. Vosters and Nybom 2016; Pekkarinen, Salvanes, and Sarvimäki 2017; Vosters 2017; Durlauf, Kourtellos, and Tan 2017).

This is by no means an exhaustive summary of the literature, but that is beyond the scope of this introduction. Interested readers can refer to the summaries of the literature given by Black and Devereux (2010) and Björklund and Salvanes (2011). While the intergenerational mobility literature

is vast, there are still a wide array of unanswered questions. This dissertation consists of three chapters, each of which aims to answer some of these unanswered questions:

Chapter 1: Income and family background: Are we using the right models? Social scientists have long been interested in the relationship between parental factors and later child income. Finding the best characterization of this relationship for the question at hand is however fraught with choices. In this paper we use machine learning methods to assess the 'completeness' of one popular modelling approach. Here, completeness refers to how well the model summarizes the total predictive relationship between multiple parental factors and a single child outcome. Machine learning methods enable us to depart from functional form assumptions, allowing flexible interactions between a large set of possible parental factors. Using our most flexible complete model as a benchmark, we assess the popular 'rank-rank' model relating parent and child incomes. Applying our approach to high-quality Norwegian administrative data, we demonstrate that the rank-rank model explains 68% of the total explainable variation in child income rank, based on a broad set of potential parental factors entering a neural network. Parental wealth and parental education explain the majority of the remaining explainable variation. For an extremely tractable model, we consider this to be a relatively high level of completeness. In light of our country-wide estimates, we explore how this measure of completeness varies across regions of Norway, finding broadly similar patterns to those found at the national level. Our results imply that comparisons of regions based on rank-rank mobility measures may indeed reflect differences in broader notions of equality of opportunity.

Chapter 2: Status Traps in Social Mobility and Human Capital Investment

Although intergenerational income mobility is high in Nordic countries, parental education still plays an important role in explaining educational attainment. Using machine learning techniques, we show that, in Norway, obtaining a college degree is not a continuous function of parental years of education and that there are discontinuities and interactions at different parental education levels. Parental earnings and the transmission of cognitive ability are not the only reasons for the status traps in education. Moreover, our findings suggest that parental education can compensate for lower cognitive ability, whereas paternal earnings cannot compensate for low parental education.

Chapter 3: Intergenerational Mobility over time and Across Regions in Norway In this paper we analyze intergenerational mobility in Norway for cohorts of children born from the mid 1950s until the mid 1980s and are grown up today. We focus on regional differences and changes across regions and within regions over time. We use several measures of income mobility, and in addition to relative mobility measures like rank-rank, we use measures to detect changes at different margins, like moving from the bottom to the top quintile and the share of sins have higher earnings than fathers. Next, we focus on the mechanisms behind the differences in mobility across regions and changes over time. We are particularly interested in the role of human capital investments, the role of the labor market and returns to human capital and characteristics of the industrial structure and other labor market characteristics. We use machine learning to identify regional differences and labor market differences. These to parts of the analysis will be analyzed together within a panel regression framework in the next step.

Chapter 1

Income and family background: Are we using the right models?

Jack Blundell and Erling Risa[*]

Abstract

Social scientists have long been interested in the relationship between parental factors and later child income. Finding the best characterization of this relationship for the question at hand is however fraught with choices. In this paper we use machine learning methods to assess the 'completeness' of one popular modelling approach. Here, completeness refers to how well the model summarizes the total predictive relationship between multiple parental factors and a single child outcome. Machine learning methods enable us to depart from functional form assumptions, allowing flexible interactions between a large set of possible parental factors. Using our most flexible complete model as a benchmark, we assess the popular 'rank-rank' model relating parent and child incomes. Applying our approach to high-quality Norwegian adminis-

trative data, we demonstrate that the rank-rank model explains 68% of the total explainable variation in child income rank, based on a broad set of potential parental factors entering a neural network. Parental wealth and parental education explain the majority of the remaining explainable variation. For an extremely tractable model, we consider this to be a relatively high level of completeness. In light of our country-wide estimates, we explore how this measure of completeness varies across regions of Norway, finding broadly similar patterns to those found at the national level. Our results imply that comparisons of regions based on rank-rank mobility measures may indeed reflect differences in broader notions of equality of opportunity.

1.1   Introduction

Intergenerational mobility, long of interest to academics, has become a key topic of policy debate across many countries in recent years. Measures characterizing the statistical relationship between some aspect of individuals' adult outcomes and features of their parents appear to be of normative value, since differences in adult outcomes stemming from circumstances of birth tend to be viewed less favorably than those related to effort or preferences (Roemer and Trannoy 1998). If these differences also reflect inequality of opportunity, there are potential economic efficiency gains to producing such measures. These measures can be used to identify mechanisms through which opportunity is diminished, helping countries better utilize their potential human capital resources. This unique alignment of equity and efficiency concerns make issues surrounding intergenerational mobility a rare point of agreement across all sides of the political spectrum.

Numerous measures capturing mobility have been developed, first by so-

ciologists who tended to focus on transmission of social class, and more recently by economists tending to focus on income mobility. With a broad set of measures to choose from, as researchers we must work to understand the statistical properties of each, and why they may differ to one another. Much academic energy has been spent debating the appropriateness of different measures (Blanden, Greaves, Gregg, Macmillan, and Sibieta 2015), therefore any methodological development in the assessment of different measures is likely to be a valuable contribution.

In this paper, we introduce a notion of 'completeness' relating to an approach outlined in Kleinberg, Liang, and Mullainathan (2018) and apply it to the setting of intergenerational mobility measurement. We define completeness as the extent to which a particular intergenerational model summarizes the full predictive relationship between a broader set of parental resources and later child income.[2] Intuitively, if it is the case that much of the relationship between parental factors and child income is unexplained by an existing model, alternative models and corresponding measures are needed to fully summarize this relationship. This approach requires the estimation of a flexible 'benchmark' model. To generate this model we utilize recently-popularized tools from the field of Machine Learning. Machine Learning (ML) methods enable us to model relationships flexibly while ensuring we do not confound signal for noise.

Using this general approach, we test the completeness of a model used to infer income mobility in many of the most recent papers in the literature, the income rank-rank model. This is a linear regression of the percentile rank of a child's income in the income distribution on the percentile rank of

---

[2]This relates to but does not coincide with the more conventional definition of completeness found in statistical theory (Casella and Berger 2002)

their parents' income. We ask the extent to which this bi-variate predictive relationship captures the full predictive content of a wide set of parental background measures, as provided by a neural network model.

Our approach is demanding in terms of data requirements. Machine learning methods are not well suited to noisy, small survey data traditionally used in many studies of mobility. Therefore we use high-quality population-level administrative data from Norway. This allows us to investigate relationships for a number of variables jointly and abstract from measurement error issues faced by survey data, which would severely complicate the analysis. We envision our approach being applied to multiple countries in future work, in order to understand how our results here transfer to alternative institutional settings.

Our results show a number of clear patterns. Firstly, we find that the simple linear rank-rank model explains 68% of the total explainable variation in child income rank, relative to our benchmark flexible neural network with a large number of predictors. For such a simple, tractable model, we consider this to be a high level of completeness and hence an encouraging result for users of measures based on rank-rank models, such as the recent set of papers using US administrative data (Chetty, J. N. Friedman, Saez, Turner, and Yagan 2017). Our full neural network model includes father and mother income separately, father and mother education separately, household wealth measures, occupation, marital status, family size and region of birth. Secondly, we find that a simple model with length of parental education and wealth rank approaches our flexible benchmark in terms of predicting child income rank. A model containing income, education and wealth predictors explains 90% of the total explainable variation.

As an extension to our main results, we explore how completeness varies across labour market regions in Norway. We find that completeness is relatively homogeneous across regions. For most areas, the income rank-rank model captures a significant share of the total explainable variation, as it did at the national level. One interesting implication of this regional analysis is that at least in this settings, comparisons of regions based on rank-rank income mobility estimates coincide with those based on a broader notion of equality of opportunity discussed in the political philosophy literature.

Our contributions are two-fold. Firstly, our methodological contribution is to introduce a new concept of completeness, opening the door to applications in many empirical settings. Secondly we implement this completeness measure, finding the intergenerational income rank-rank model to be a relatively complete summary of broader parental influence on child income. This has important implications for the growing empirical literature on income mobility.

In Section 1.2, we discuss how our approach here builds on the existing literature. Section 1.3 introduces a conceptual framework for our completeness measure, for which we outline estimation issues including a brief introduction to machine learning in Section 1.4. We discuss the data and empirical setting in Section 1.5, before showing our results in Section 1.6. In Section 1.7 we conclude and indicate directions for future research.

1.2 Literature

Becker and Tomes (1979) and Becker and Tomes (1986) are often cited as the start of the literature analyzing intergenerational mobility in Economics,

building on an earlier Sociology literature Blau and O. D. Duncan (1967). A vast number of papers on the topic have been published since these early innovations, for which a complete survey is beyond the scope of this literature review. Comprehensive surveys of the literature in general can be found in Solon (1999) and Black and Devereux (2010).

The rank-rank model on which we focus stems originally from Dahl and DeLeire (2009), but is now most associated with the work of researchers at the Equality of Opportunity Project (Chetty, Hendren, Kline, Saez, and Turner 2014). These papers demonstrate that the linear rank-rank model is more robust to sample definition and measurement error than the previous benchmark measure, the intergenerational elasticity. For these reasons, we consider the rank-rank model the state-of-the-art approach and hence focus on its properties in this paper. In addition to producing country-wide estimates, large-scale administrative data has allowed researchers to investigate differences across regions within countries (Chetty, Hendren, Kline, and Saez 2014). In this paper, regions are compared using a large number of measures, some of which are based on rank-rank relationships. In our analysis we test the extent to which such regional estimates reflect broader parental influence, and whether this is constant across regions.

As well as the above papers using US data, availability of high-quality linked administrative data in Europe has given rise to a range of empirical papers studying intergenerational mobility. Examples include Bratberg, Nilsen, and Vaage (2005) who look at income mobility over time in Norway and Bratberg, Davis, Mazumder, Nybom, Schnitzlein, and Vaage (2017) who compare income mobility across a number of countries. All studies find relatively high rates of intergenerational mobility in Norway. Given this, we would

expect in our exercise to find that parental income have little explanatory power over child income. The contrast however between mobility estimates in Norway and estimates elsewhere reinforces the need to replicate the exercise adopted here in different settings.

The majority of papers in the literature assume linearity. Examples of papers exploring non-linearities include Durlauf, Kourtellos, and Tan (2017), Pekkarinen, Salvanes, and Sarvimäki (2017) and Bratberg, Nilsen, and Vaage (2007). A consistent result, which we also find, is that child income are approximately linear in parent income throughout the middle of the parental income distribution, but increased persistence at the top and bottom of the distribution leads to non-linearities. As will be expanded on in Section 1.5, rather than income or income at a single point in time, interest is primarily in relationship between lifetime income or income. More recent papers on the US administrative data have found that non-linearities tend to emerge when later ages are used for child income (Chetty, J. N. Friedman, Saez, Turner, and Yagan 2017). In our empirical exercise we are able to quantify the importance of these non-linearities by progressively allowing more flexibility in parental income.

A small number of papers investigate the joint impact of a broader set of factors for both parents and children. In a well-known and expansive historical study of surnames, Clark (2014) argues that analyzing single measures of socioeconomic status leads to estimates of intergenerational mobility with a severe upward bias. This work sparked discussion, and several papers have addressed his claim, for example Vosters and Nybom (2016) and Vosters (2017). These papers estimate a "least-attenuated" measure of persistence on data from Sweden and the United States. They find no evidence of sub-

stantial bias in prior estimates. These measures assume that measures of socioeconomic status are proxies for a single latent variable. An advantage of this approach is that multiple factors can be observed and incorporated on the parent side, however a disadvantage is the assumption of linearity throughout. While we do not explore the importance of allowing for multiple child factors, we are able to flexibly include many parental factors.

A series of papers related to ours attempt to decompose variation in child income by family effects and neighborhood effects. Papers taking this approach include Solon, Page, and G. J. Duncan (2000), Page and Solon (2003), Raaum, Salvanes, and Sørensen (2006) and Nicoletti and Rabe (2013). This literature typically finds that family effects seem to explain more of the variation than neighborhood effects. While variance decomposition gives an estimate of how predictive (observed and unobserved) family effects are compared to neighborhood effects, it is uninformative regarding which particular features of the family and neighborhood are driving predictive performance. Hence our approach is more informative for understanding mechanisms linking parental background to child income.

Applying machine learning techniques in economics is becoming increasingly commonplace, as discussed in Mullainathan and Spiess (2017) and Athey and Imbens (2017). Some such applications include Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015) and Mullainathan and Obermeyer (2017). As discussed in the introduction, our method here relates to a concept of completeness introduced in Kleinberg, Liang, and Mullainathan (2018). The authors use a similar notion of completeness, estimated via Machine Learning, to assess a variety of behavioral models seeking to explain human perception of randomness. The idea of using machine learning methods to

provide an upper bound on explainable variation is also used in Gathergood, Mahoney, Stewart, and Weber (2018) when assessing different models of debt repayment. Our approach differs in that we are assessing a descriptive relationship rather than competing behavioural models. We also allow our complete model to include predictors outside of the original model, and hence are simultaneously testing both the cost of particular functional form assumptions and the cost of limiting the predictor set. This and our paper fits into a broader set of papers using machine learning methods for model building and evaluation purposes (for example Liang and Fudenberg 2018).

Related to our approach, a small and growing number of papers are drawing on machine learning methods to use predictive performance as an object of interest in itself. As discussed below, one interpretation of our fully flexible model is as an equality of opportunity measure in its own right. The interpretation of predictability as a measure of (in)equality of opportunity stems from theoretical contributions in political philosophy associated with Roemer and Trannoy (1998). This is the approach taken in Brunori, Hufe, and Mahler (2018), who interpret the predictive power of regression tree models as an equality of opportunity measure using cross-country survey data. Other work using predictive power as a measure include Gentzkow, Shapiro, and Taddy (2016), who estimate political polarization by predictive power of congressional speeches over party membership. Bertrand and Kamenica (2018) use a number of sets of measures including consumption and time use to predict group membership over time, investigating whether there has been a change in the 'cultural distance' of difference parts of US society.

1.3 Conceptual Framework

In this section we introduce our measure of completeness and outline why this is a useful approach for understanding the measurement of intergenerational mobility. As demonstrated in the literature review, the dominant approach to measuring mobility is to select a single indicator such as income, and inspect the relationship between parental and child values of that variable. Using identical indicators for both parents and children allows within-family income across generations to be viewed as an $AR(1)$ process.[3] This then allows consideration of long-run patterns. For example, these estimates allow claims such as 'at this level of mobility, it would take X years for group A to catch up with group B'. This approach to mobility has a rich history going back to Galton (1877).

While we acknowledge this benefit, we argue that this long-run interpretation of mobility is not particularly useful, as there is little reason to believe that mobility rates remain fixed across many generations. Additionally, the great level of interest in these measures appears to be primarily due to their link to equality of opportunity. Crucially, we argue that if one is interested in measuring equality of opportunity, beyond ease of interpretation there is no advantage to constraining a model to include only a single parental factor, and no convincing reason for this single factor to be identical to the child outcome. While we consider multiple parental factors in this paper, we consider only a single child outcome, namely the child's rank in the income distribution. We hold this fixed, while varying the number of parental factors and the way in which they are included as predictors. Of course, if one

---

[3]As well as linearity, this is also assuming the impact of previous generations is fully captured by the most recent generation. There is good reason to question this markov assumption, as shown for example by Long and Ferrie (2018).

considers child income rank to be a poor outcome measure, the methods here could be applied to any alternative measure, such as education, social class or accumulated wealth. In practice, there could be many alternative outcomes of interest which society chooses to value, such as life expectancy, health or wealth. We choose child income rank as it is the dominant outcome measure in the current literature, and income is highly correlated with most other outcomes associated with wellbeing.

This point can be made clear by way of a simple example, in which we only consider parental income and parental education. Consider a world in which the observed returns to education among parents are low, so that parental incomes are only loosely related to parental education. Among the next generation however, returns to education are high, meaning that education and income are highly correlated among the children's generation. Assume that in this world one's education is closely linked to the education of one's parents, perhaps through information or through preferences. Clearly, income mobility would be high as the link between parent and child income is weak. However, incomes of children are tightly linked to education of parents. By focusing on income mobility alone, this dependence of child income on parental characteristics is left undetected. The total explainable variation of child income based on parental characteristics is high, but a model including parental income alone would explain very little of the variation in child income. Using our approach, in this setting a model with solely parental income as a predictor would be associated with a low completeness score.

The world described above contrasts with a world in which returns to education are high both for parents and for children. Let us again assume

that education is tightly linked to parental education. In this case, parental income will also be tightly linked to child income. The additional predictive power over child income one could achieve by including parental education would be low. Here then, a model with solely parental income as a predictor would be associated with a high completeness score.

A simple way to test for the above is to include education as a predictor in a linear model and inspect the R^2, or any other measure of fit. The degree to which R^2 (or adjusted R^2) increases with the inclusion of an additional variable is informative of how 'complete' the restricted model is. While a much-simplified version, this captures the spirit of our approach. Abstracting from small-sample estimation issues, this simplified approach would be valid if two conditions are satisfied:

1. Child income is linear in all predictors

2. There are no interaction terms

There is little evidence to suggest that either of these conditions holds in practice. It is well documented, for example that child income rank is non-linear in parent income rank at the tails of the distribution. Not only this, but the linearity assumption will depend on the exact form in which a variable enters. For example, is it years of education that matter, or is it highest qualification achieved? Similarly, ruling out interaction terms seems implausible in this setting. Therefore more complex models are needed, which bring their own estimation problems.

An alternative interpretation of completeness as a measure of whether a linear rank-rank income provides a 'sufficient statistic' for the expected lifetime income rank of children. Though not the focus of this article, this has

practical applications when one thinks of targeting early years interventions. If one considers the targeting of early childhood interventions based on expected future income, it is valuable to know whether this tagging can be done purely using simple models including parental income alone.[4]

Clear from the above exposition is that completeness is defined relative to a set of parental factors. This is a disadvantage of the approach. Ideally, one would like to include as many features of parents as possible. In practice then, our completeness score can be thought of as an upper bound on the completeness score achievable if all variables were included. Which variables one decides to include will inevitably be in part dictated by data availability, but it may also be dictated by normative arguments. The Equality of Opportunity literature in political philosophy contains ample discussion of how the relevant set of family and background 'circumstances' can be determined (Roemer and Trannoy 1998). Throughout our results, our approach is to be transparent on the set of variables used as predictors in each model and to include as wide a set of family predictors as possible. It could however be the case that there exists some unobserved factor, for example parental altruism, which is yields additional predictive power over later child income yet does not enter our model.

While our idea is intuitive, to frame discussion we now give a brief formal outline of the compleness measure. Let y_c denote child lifetime income and y_p denote total parent lifetime income. For the child's year-of-birth cohort, let y_c^r and y_p^r denote the percentile ranks of child and parent lifetime income within the child's cohort. The rank-rank model $f_{rr}(y_p^r)$ is the following linear

---

[4]This is of course abstracting from the practical issue that these parameters can only be estimated after the child's cohort has reached adulthood. In practice some temporal stability of functional form must be assumed.

projection of child rank on parent rank:

$$\hat{y}_c^r = \alpha + \beta y_p^r$$
$$= f_{rr}(y_p^r) \tag{1.1}$$

where α and β are the usual OLS parameters. $\hat{y}_c^r$ is the predicted child rank from this linear projection. Parameter β is the 'rank-rank slope'. Note that one particular feature of the rank rank model is a one-to-one mapping between coefficient β and R^2.

Next define X_p as a full set of parental characteristics, which includes y_p^r. Let us define the expectation of child income rank, conditional on parental factors as $G(X_p)$, so that $\mathbb{E}[y_c^r|X_p] = G(X_p)$. Conditional expectation $G(X_p)$ minimizes the following population mean square error loss function:

$$G(X_p) = \text{argmin}_{\pi(X_p)}\mathbb{E}[(y_c^r - \pi(X_p))^2] \tag{1.2}$$

Letting $\mathcal{L}(\pi(X_p)) = \mathbb{E}[(y_c^r - \pi(X_p))^2]$ for any $\pi()$ function, this implies that:

$$\mathcal{L}(G(X_p)) \leq \mathcal{L}(f_{rr}(y_p^r)) \tag{1.3}$$

The rank-rank model must perform weakly worse than the full conditional expectation $G(X_p)$ in terms of minimizing the MSE loss function. 'Completeness' is defined as the ratio of these two loss functions:

$$Completeness = \frac{\mathcal{L}(f_{rr}(y_p^r))}{\mathcal{L}(G(X_p))} \tag{1.4}$$

Note that due to the above condition and the fact that $\mathcal{L}(f_{rr}(y_p^r)) \geq 0$, completeness is bound between 0 and 1, and can be expressed as a percentage.

Completeness summarizes the cost in terms of explaining variation of the outcome, of restricting the model to $f_{rr}(y_p^r)$ relative to the full conditional expectation $G(X_p)$.[5] Low completeness scores can be driven by both the exclusion of relevant predictors and by poor functional form assumptions. In Section 1.6 we will discuss our attempt to distinguish between these two effects. The completeness statistic can also be written as the ratio of two R^2 values, $\frac{R^2_{f_{rr}}}{R^2_{G(X_p)}}$, where $R^2_{f_{rr}}$ is the population R^2 from the rank-rank model and $R^2_{G(X_p)}$ the equivalent from the full model. Interpreting R^2 in its usual way as *the fraction of variance explained by the predictors*, we then interpret completeness as *the fraction of total explainable variance explained by model f_{rr}*.

1.4 Estimation

1.4.1 Statement of problem

A natural estimator of population mean squared error $\mathcal{L}(G(X_p))$ is the sample MSE. If function $G(X_p)$ is known, this is both unbiased and consistent for $\mathcal{L}(G(X_p))$:

$$\mathbb{E}[\frac{1}{n}\sum_i (y_{c,i}^r - G(X_{p,i}))^2] = \mathcal{L}(G(X_p)) \qquad (1.5)$$

$$\lim_{n \to \infty}[\frac{1}{n}\sum_i (y_{c,i}^r - G(X_{p,i}))^2] = \mathcal{L}(G(X_p)) \qquad (1.6)$$

In practice, $G(X_p)$ is not known and hence must be estimated.

Two immediate issues emerge with the estimation of this conditional expectation. Firstly, $G()$ could be a highly complex function with ample non-linearity

---

[5]What distinguishes this measure of completeness from that which is presented in Kleinberg, Liang, and Mullainathan (2018) is the fact that we allow for a broader set of predictors in our conditional expectation $G(X_p)$. Their notion of completeness differs primarily as it holds the set of predictors fixed.

and interactions between individual components of X_p. As we want to extract the full predictive power of X_p on child income, it is important to allow for these complexities. Non-parametric approaches such as kernel regression could be used to achieve this. Secondly, X_p may be high-dimensional. The number of predictors in X_p may be large relative to the feasible sample size, particularly since we need to allow for a full set of interaction terms to truly extract all explainable variation. With a large number of predictors relative to sample size, conventional methods to estimate $G()$ can be heavily biased. This is the familiar 'overfit' problem in which high dimensionality can lead to noise being mistaken for signal.

In the case of high-dimensional linear regression, one strategy to avoid overfit is to adopt an ad hoc approach to limit the number of variables included in the model. While in practice this may appear to work well, without allowing for the full set of possible predictors we will not know which should be included. Ad hoc methods, such as adding in variables one at a time, or estimating a series of bivariate relationships, will never allow the researcher to be sure we are obtaining a good approximation of $G(X_p)$. If we allow for non-linearities and interactions, the problem becomes vastly more complex and ad hoc model selection methods are not possible. For example, with 10 continuous predictors, allowing for mild non-linearities by the inclusion of third-order polynomials leads to 30 predictors. Allowing for bivariate interactions then leads to 435 predictors. The strategy of adding in predictors individually and simply 'seeing what works' is not possible if one wants to allow for a reasonable degree of flexibility.

1.4.2 Machine Learning Methods

As should be apparent from the above, estimating a conditional expectation perfectly coincides with finding the best out-of-sample prediction, provided 'best' is defined by minimizing mean squared error. This equivalence is extremely convenient, as it allows us to draw on the rapid developments in predictive modeling which come under the umbrella term of Machine Learning.

Broadly speaking, (supervised) machine learning methods are highly optimized for cases such as these where we would like to obtain the 'best' out-of-sample prediction possible, rather than cases where we would like to estimate and inspect a set of parameters.[6] As outlined in Mullainathan and Spiess (2017), the goal of machine learning methods is to provide a predicted outcome $\hat{y}$ rather than an estimated parameter $\hat{\beta}$. Note that in the conceptual framework above there is no discussion of the parameter vectors underlying the conditional expectation, only the expectation (or prediction) itself. The brief overview here is included to help the reader understand our results rather than to give a comprehensive introduction to machine learning methods. For a complete introduction to machine learning methods designed for economists, see Mullainathan and Spiess (2017).

While there exist a vast array of models falling into the machine learning label, there are some common steps which bind many of them:

1. Split the full sample into a 'training' and a 'hold-out' or 'test' set

---

[6]When using the term 'machine learning' in this paper, we refer to 'supervised' machine learning, in which the goal is to predict some outcome y from predictors X, and the researcher has access to a set of example observations.

2. Fit models in the training set

3. Assess out-of-sample performance in the hold-out set

The separation of data into a subsample on which models are fit ('trained') and a mutually exclusive hold-out subsample used to test performance is crucial.[7] This ensures that we obtain an unbiased estimate of out-of-sample performance. When engaged in predictive modeling, we can typically assess how well our models are achieving their goal through sub-sampling. This is distinct to causal analyses, where the objective is to recover a parameter of interest which is fundamentally unobservable. Causal analysis always relies on some untestable prior identifying assumptions, which we do not require here.

Training our machine learning models involves several steps. Each machine learning model has a set of tuning parameters which are to be 'learned' from the training data which typically determine the complexity of the model. A complex model may fit well in-sample, but perform poorly out of simple. On the other hand, a simple parsimonious model may miss key patterns in the data. Therefore choosing complexity parameters involves a trade-off.[8] Cross validation is an extremely-common approach used to choose these tuning parameters, which can be thought of as a grid-search approach in which an estimate of out-of-sample performance is the objective function. For our purposes, cross validation consists of the following steps:

1. The training set is divided into k (usually 5 or 10) equal-sized parts, named 'folds'

---

[7]In some applications, an additional split of data is performed, giving a training, test and 'pure' holdout set. As we are fitting only a small number of models, we do not employ this additional data division.

[8]This is the familiar bias-variance trade-off found in non-parametric econometrics.

2. For a particular set of parameters, the model is fit to data from $k - 1$ of these folds, and performance is recorded on the omitted fold

3. Step 2 is repeated k times, with each iteration seeing a different omitted fold

4. Performance across all k repetitions is averaged

5. Steps 2 to 4 are repeated for a variety of parameter choices

6. The parameter set corresponding to the best average performance is chosen.[9]

Asymptotic results in Vaart, Dudoit, and Laan (2006) suggest that this method of tuning approximates the optimal model complexity for out-of-sample prediction.

The final step of the process is to estimate the model with the cross-validation-chosen parameter set on the test (or hold-out) set. If cross-validation is successful, predictive performance in the test set should be close to performance in cross validation sets.

Table 1.1 briefly outlines the five machine learning methods used in our empirical setting. These are Elastic Net, Regression Trees, Random Forest, Gradient Boosted Trees and Neural Net. The interested reader is referred to Appendix 1.8 for more details on our particular implementation. In our empirical application, we apply these algorithms alongside conventional linear regression for a number of different predictors. These algorithms have been chosen in part for their empirical performance across a wide set of studies, and in part as comparing their performances aids a discussion

---

[9]Alternative ways of selecting parameters also exist. We adopt this approach for simplicity.

of underlying patterns in the data and potential estimation issues. As we will show in the results, the choice of particular machine learning method turns out to be inconsequential. This gives us confidence that we are truly capturing the full extent of predictible variation in our outcome.

test

1.5   Data

Our full dataset consists of all individuals born in Norway between 1970 and 1975 along with information on parents. This high-quality data allows linkages of parental income to child income in tax records, as well as to a variety of other characteristics of parents which will enter our models. We link children only to biological or adopted parents. We do not included predictors based on step-parents. We pool both genders and do not allow predictors to interact with gender.[10]

In total there are 282,770 individuals for whom we observe both child income and the full set of family predictors. For computational reasons, we draw a random sample of 141,385 (50%) and perform the majority of the analysis on that sample.[11] In the remainder of this section we describe the variables used our analysis in detail.

Income

---

[10]While these may indeed be such interactions, this approach is motivated by the observation that gender is not something which is determined by family background.

[11]While it is generally computationally feasible to apply these methods to datasets of over several hundred thousand observations, we are constrained by the computational power of the server on which our administrative data is held.

Table 1.1: Five Machine Learning methods

Model	Short description
Elastic Net	A linear regression model in which coefficients with large absolute values are penalized via the addition of a penalty term in the objective function. The inclusion of this penalty shrinks coefficients towards zero or sets them equal to zero. Can be thought of as an amalgamation of LASSO and ridge regression.
Regression Tree	Regression trees pool observations into mutually-exclusive rectangular subspaces over predictor variables, based on similarity in the outcome variable. The model can be characterized as a sequence of splits over individual variables.
Random forest (Ranger)	Random forests involve the estimation of many regression trees. Randomness across trees is introduced in two ways, firstly by using a different bootstrap sampel for each tree, and secondly by constraining trees to search for splits over a random subset of predictors.
Gradient Boosted Trees (XGBoost)	An ensemble method in which many regression trees are grown and their corresponding predictions are combined. In the first stage, a single shallow regression tree is fit to the data. Each further tree operates on the prediction residuals of previous trees. At each stage, the algorithm attempts to improve poor predictions from previous stages by searching for patterns which explain residuals.
Neural Network (Neural Net)	The most 'black-boxy' of our machine learning models, this type of model is based loosely on the structure of the brain. This family of models contains an interconnected group of nodes, organized into layers. Starting with input nodes corresponding to predictor variables, signals of different strengths are transmitted between nodes, and each node performs some non-linear transformation to the signal. After typically being passed through multiple layers of nodes, the resulting prediction is highly flexible.

Notes: More information on each method is provided in Appendix 1.8, along with full references for further reading.

Our primary income measure is pension-generating income, which includes labor income (employed or self-employed) as well as work-related transfers such as unemployment benefits. Our main child outcome variable is income as average annual pension-generating income between 30 and 35 years old. Children are assigned a percentile rank according to their position in the income distribution. Ranks are calculated within each birth cohort and individuals with the same average income are assigned the same rank. Parental income is the sum of mother and father pension-generating income. In our flexible models we calculate income ranks separately for mothers and fathers in order to extract as much information as possible from the income measures.

For parents, income is averaged between the ages of 40 and 50. This would ideally coincide with the ages of the children (30-35) to truly reflect parental lifetime income ranks, however data constraints make this infeasible. Income densities for children, mothers and fathers are plotted in Figure 1.7 in Appendix 1.9.

Education

We incorporate several aspects of parental education into the analysis, all of which come from the National Education Database. Our main education measures are years of education for the mother and the father. The distributions of years of education for mothers and fathers is given in Figure 1.8 in Appendix 1.9. In addition to years of education, we are able to include indicators for education type in our extended models. Education types in the national education database are classified using the NUS2000 standard. The resulting indicators contain information on both the level of education,

the main field and on subfields. We can for instance distinguish between someone studying physics and chemistry.

Wealth

The administrative tax records also provide measures of parental wealth. The measure is net taxable wealth, i.e. gross taxable financial wealth net of debt, so excludes housing wealth. We measure wealth as average net taxable wealth between the ages of 40 and 50 and calculate wealth ranks separately for mothers and fathers. When calculating the wealth ranks, we assign individuals with the same reported net wealth the same rank. This is particularly relevant for individuals whose debt is larger than their assets. The tax records report these individuals as having zero net wealth, and consequently we assign them the same rank. A large number of individuals hold zero wealth, as demonstrated in the wealth distributions plotted in Figure 1.9 in Appendix 1.9.

Extended predictors

In our broadest models, we include a wide set of predictors in addition to those described above. While our main focus is on parental income, education and wealth, we are able to include a wider set of predictors in extended models. These are occupation type, marital status, whether someone lives in an urban or rural area, whether they are studying, most important source of income, region, total hours worked and the number of individuals in the household.

1.6 Results

In this section we present the main results of our analysis, starting with results from the whole of Norway in Section 1.6.1. Building on our main country-level results, in Section 1.6.2 we perform additional analysis on individual labor market regions of Norway.

1.6.1 Main results

Initial descriptives

Before moving to the calculation of completeness, we first demonstrate several important patterns in the data. In Figure 1.1 we plot the conditional expectation of child income rank on parent income percentile rank for the full dataset. We see substantial regression towards the mean, demonstrated by a slope parameter substantially lower than 1 across the distribution. Here we can also see the approximate linearity, with concavity at the bottom end of the distribution of parent income and convexity at the top. Figures 1.10, 1.11 and 1.12 in the Appendix 1.9 show the equivalent plots for father education, mother education and joint wealth. The education pattern is fairly noisy due to small numbers of individuals with non-standard numbers of years of education, but not clearly non-linear. The pattern for wealth is very similar to that seen here for income.

Figure 1.1: Mean child income rank by parent income rank
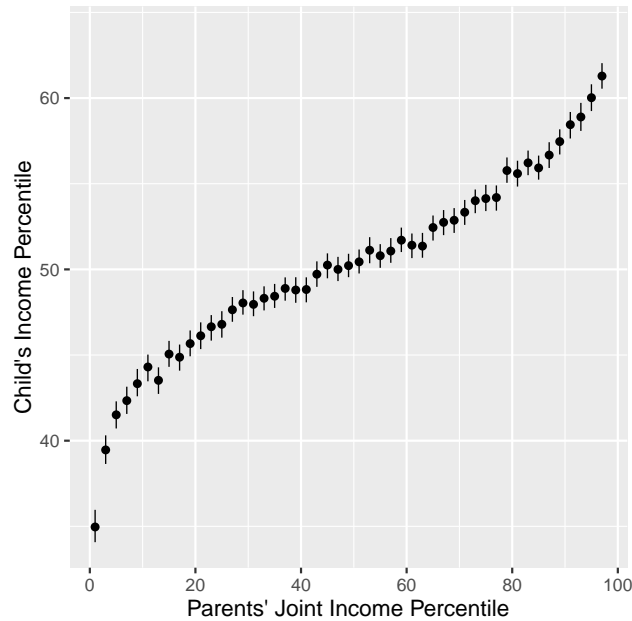
Notes: Conditional expectation of child income percentile aged 30 to 35 by parent child rank aged 40 to 50. Error bars correspond to bootstrapped 95% confidence intervals. Further details on variable construction in Section 1.5.

Rank-rank estimation

We estimate a country-wide rank-rank slope of 0.18. This compares to a slope of 0.34 found for the US in Chetty, Hendren, Kline, and Saez (2014). It is worth noting that our data differs from the US-based research in important ways. For example, our income measure includes several aspects of non-labor income. Our age restrictions also slightly differ, as does the definition of the family unit. Nonetheless, our low estimate of the rank-rank slope is consistent with cross-country comparisons of mobility such as Bratberg, Davis, Mazumder, Nybom, Schnitzlein, and Vaage (2017). In line with the literature, our estimate here suggests Norway to be substantially more mobile in income than the US.

Completeness

Having explored patterns in the data and estimated the rank-rank model, we now turn to the estimation of completeness. This involves choosing a benchmark flexible ("complete") model against which we assess the rank-rank model. Our benchmark model is the neural network applied to the full set of potential predictors listed in section 1.5, including the 'extended' predictors. As will be clear in our results, the precise machine learning model used does not affect our completeness results, as we get similar measures for each. In all but the rank-rank model, parental income enters flexibly with second-order polynomials.[12]

Our first set of results uses our random sample of 141,384 observations. We divide these into a training set of 113,108 observations and a test set of 28,276 observations. Rather than purely show the rank-rank model and the most flexible (complete) model, Figure 1.2 shows the hold-out test set R^2 estimates for all models (where possible) and six different sets of predictors.[13] These test set results are provided also in Table 1.2 in Appendix 1.9. Regression Tree model results are omitted. Consistent with other studies, we find that Regression Tree models overfit the training sample and hence perform poorly in prediction. We maintain the discussion of regression trees in early sections as we believe it useful for understanding more complex tree-based methods.

There are several important results yielded by Figure 1.2. Moving from the left-most (most restricted) models towards the right of the figure, we first see that allowing income to enter flexibly gives moderate gains in predictive

---

[12]The models labelled "Income (flexible)" include third-order polynomials in parent income.

[13]Given the substantial computational time, we only apply the neural net model to the full (extended) set of predictors.

power. This is unsurprising given the conditional expectation plotted in Figure 1.1. There exist non-linearities in the relationship between child income rank and parent income rank, which are reflected in a greater R^2 once income is modeled more flexibly.

Turning to the third set of bars in Figure 1.2, we see a marked increase in predictive power from including wealth variables. Focusing only on the OLS estimates, the R^2 increases from 0.037 to 0.049, a 32% increase. Elastic Net delivers exactly the same predictive performance, and the most flexible method applied to these predictors (Gradient Boosted Trees) delivers a small improvement on the linear models. The random forest model "Ranger", performs poorly here and throughout. Taken together, this then implies that parental wealth indeed carries non-negligible additional predictive power over later child income relative to a prediction based only on parental income. A linear model performs reasonably well relative to a more flexible model. This also suggests that interaction terms between income and wealth may be present, but do not fundamentally affect predictions. This is potentially an important result for future empirical work. Wealth variables are rarely available, but this demonstrates that parental wealth matters for later child income, in a predictive sense, over and above parental income.

Turning now to the fourth set of bars labeled 'income and education length', we that the pattern is similar to that for wealth. Again, the elastic net model does not perform better than OLS. At least in these simpler models, our sample size is sufficiently large that overfit is not an important issue in practice.

The fifth set of models, labeled "Income, wealth and education length" contains all predictors from the previous sets of models. We see that this model

again gives an improvement in predictive power and similarly to previous models, elastic net gives no improvement on OLS. This suggests that overfit is again not a problem here. Interestingly, the predictive performance of models including income, education and wealth together improves substantially on models including only income and wealth, or only income and education.

Observing the performance for the final set of models including the full set of income, wealth and education and extended predictors, we see an improvement for all methods. Elastic net, the regularized linear model, now outperforms OLS. With the extended set of predictors, overfit becomes an issue. The most flexible method offers moderate improvement on the linear model. Non-linearities and interactions are indeed present, but do not lead to substantial gain in predictive power. Without applying our ML methods, we would not have recognized this additional predictive power.

Finally, we move to a discussion of completeness. Our complete model is the very final column in Figure 1.2. This model achieves an R^2 of 0.055. Our rank-rank model obtains an R^2 of 0.037. Dividing these two numbers gives a completeness of 0.68. Therefore we conclude that the rank-rank model is 68% complete relative to the fully flexible model. We consider this to be a relatively high level of completeness for such a simple model with only a single predictor.

To obtain an estimate of the precision with which our R^2 values and hence completeness are estimated, we can use the cross-validation results from the training set. Figure 1.3 shows the mean R^2 values and 95% confidence intervals based on 10 cross-validation sets. Reassuringly, these results are very close to those found in Figure 1.2. We see that our predictions are very precise, due to our large sample size. We can therefore be confident that our
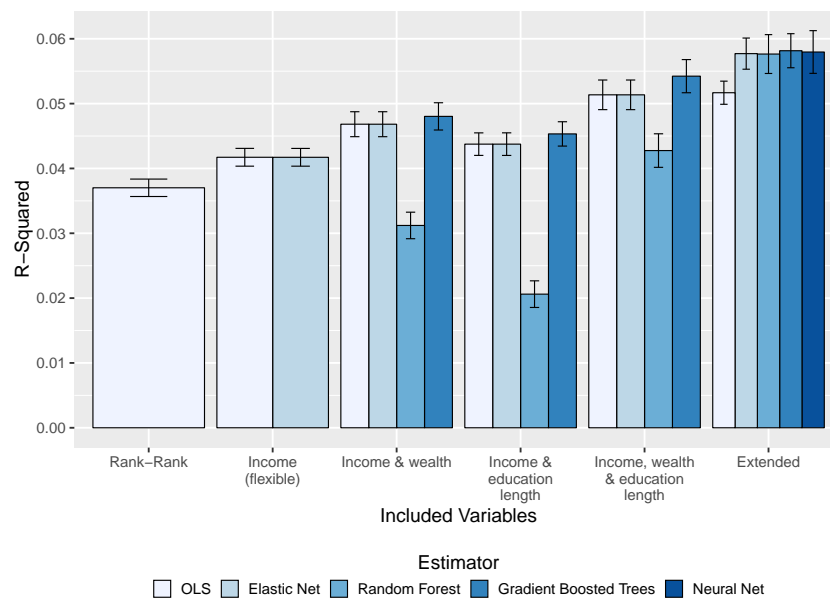
Figure 1.2: Hold-out test set performance
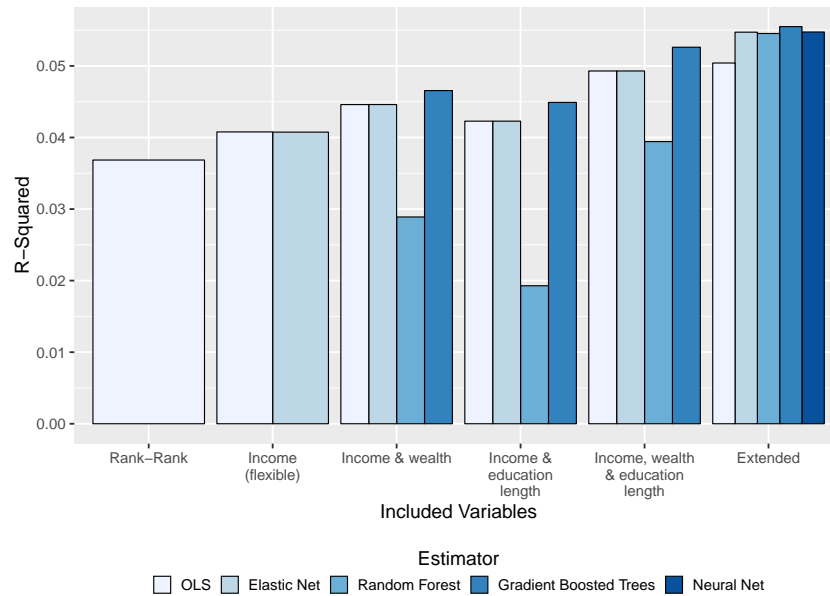
Notes: Rank-rank corresponds to a simple OLS regression of child rank on parent rank. Income (flexible) refers to a model in which income is included separately for each parent, along with third-order polynomials and first-order interactions. For all other models, variables included are as discussed in Section 1.5, along with the income (flexible) variables. The full set of test set estimates are available in Table 1.2 in Appendix 1.9.

estimate of 68% completeness based on the test-set results is not sensitive to our particular choice of sample. The full set of training set estimates are available in Table 1.3 in Appendix 1.9.

To summarize our main national results, we find that the standard rank-rank model explains two thirds of the explainable variation in child income rank, relative to a flexible neural net model with a wide set of family predictors. A substantial share of this remaining explainable variation is explained by parental wealth and parental education.

Figure 1.3: Training set performance

Notes: Rank-rank corresponds to a simple OLS regression of child rank on parent rank. Income (flexible) refers to a model in which income is included separately for each parent, along with third-order polynomials and first-order interactions. For all other models, variables included are as discussed in Section 1.5. Error bars based on two standard deviations of R^2 estimates drawn from 10 cross validation sets. The full set of training set estimates are available in Table 1.3 in Appendix 1.9.

1.6.2 Region-level analysis

In this section, we explore how our measure of completeness varies across regions.

Rather than estimating our neural network model at the region level, instead we utilise predictions from the national model discussed in the previous section. This means that predictions for each region under this model will not only be based on observations within that region, but can also learn from patterns in the data from outside that region. As region indicators are included, it is possible for the neural network to omit any information from all other regions when making predictions for one region if these other data points provided no useful information. Following in the steps of Chetty, Hendren, Kline, and Saez 2014, we estimate rank-rank relationships individually for each region, where ranks are determined according to the national income distribution. Completeness scores are calculated for each region as described in Section 1.3, by dividing the R^2 of the rank-rank model by that of the fully flexible model.

The first panel of Figure 1.4 shows completeness for each labour market region of Norway. The results are also provided in Appendix 1.9, Table 1.4. The majority of regions fall between 0.4 and 0.6 in completeness. This is lower than the 0.68 found at the national level. This in part reflects the previous discussion, that the neural net model is able to draw on data points from outside each region to deliver predictions. Panels 2 and 3 show the underlying R^2 metrics, first from the rank-rank model and then from the flexible model. Comparing panels 2 and 3, we see that the patterns are broadly similar. There is higher predictability in areas to the Southeast, and

relatively low predictability in the West.

In Figure 1.5 we plot the R^2 results shown in the map in two dimensions, and also show how these relate to the rank-rank slope. While at the national level, there is a one-to-one relationship between the rank-rank slope and the R^2 in the rank-rank model, this need not hold at the regional level. These graphs however show that in practice the two are almost identical, with a correlation of 0.98. As reflected in the bottom panels, the correlation between the R^2 in the rank-rank model and that of the neural net is high (0.87). Areas in which a child's later income is highly predictable are indeed those where the rank-rank slope is higher.

Overall, we interpret our regional results as showing a large degree of homogeneity across regions. While there exists some variation, there are no regions of Norway where parental income captures close to all the predictive power of family background, yet in all regions parental income is certainly an important predictor. An interesting implication of these results relates to the Equality of Opportunity literature spearheaded by Roemer and Trannoy (1998). In this literature, it is argued that predictive power of family background is an object of interest in itself, and is directly reflective of equality of opportunity. This idea is implemented using cross-country survey data in Brunori, Hufe, and Mahler (2018). While this approach to equality of opportunity is not without its critics, our results here suggest that if one does accept predictive power as a measure of equality of opportunity, then the income rank-rank estimates used extensively in recent US papers do indeed broadly correlate with equality of opportunity. This is important for those seeking to understand the normative value of geographical variations in rank-rank income mobility.

1.7 Conclusion

In this paper introduce a new notion of model completeness and apply it to the setting of intergenerational mobility measurement. Our main finding is that rank-rank income mobility models explain approximately two thirds of the total explainable variation in child income rank, relative to a neural net model with a wide set of predictors. We consider this to be a relatively high level of completeness, given the tractability of the rank-rank model. At the national level, parental wealth and education are important predictors in determining future child income, in addition to parent income.

As an extension to our national results we explore how our completeness measure varies across labour market regions. We find generally stable completeness across regions. A potentially important implication of this is that regional mobility measures stemming from the rank-rank model do, in this instance, reflect broader notions of equality of opportunity proposed for example by Roemer and Trannoy (1998).

The research presented here leaves many avenues open for further work on the topic. Firstly, we believe that the general approach of utilising machine learning to provide a flexible benchmark against which more tractable models can be assessed is a valuable addition to the toolkit of economists. As demonstrated by the small recent literature using these types of strategies, this forces economists to consider not only whether their theory is consistent with the data, but also the extent to which their model captures the explainable variation in the data. Secondly, while we approach the problem here from a purely statistical perspective, embedding this statistical approach within a social planner's problem via a social welfare function would be an

important step in linking this exercise with the extensive literature on welfare economics. Understanding the conceptual link between measures of equality of opportunity proposed in the political philosophy literature and measures of intergenerational mobility estimated by economists is a potentially rich area of interdisciplinary research. Finally, we encourage the replication of our approach to other countries, so that we might understand whether the completeness found here is similar in alternative institutional settings.

Figure 1.4: Completeness and R^2 across regions
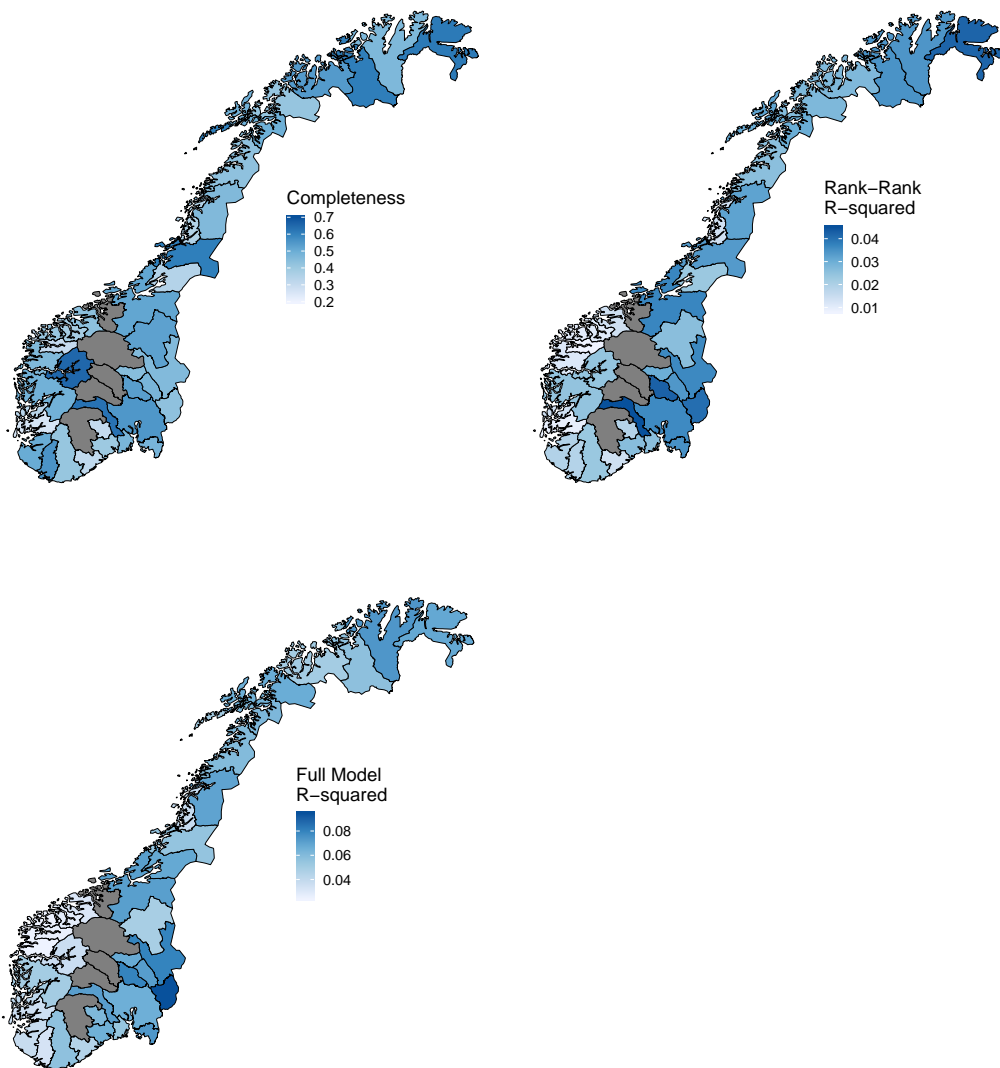
Notes: Maps of labour market regions of Norway. The first panel shows our "completeness" measure calculated across regions. The second and third panel decompose this into first the R^2 of the rank-rank model, then the R^2 of the fully flexible (neural net) model. See Table 1.4 for full set of regional results. Grey areas indicate regions with fewer than 1,000 observations.

39

Figure 1.5: Rank-rank R^2 and full R^2 across regions
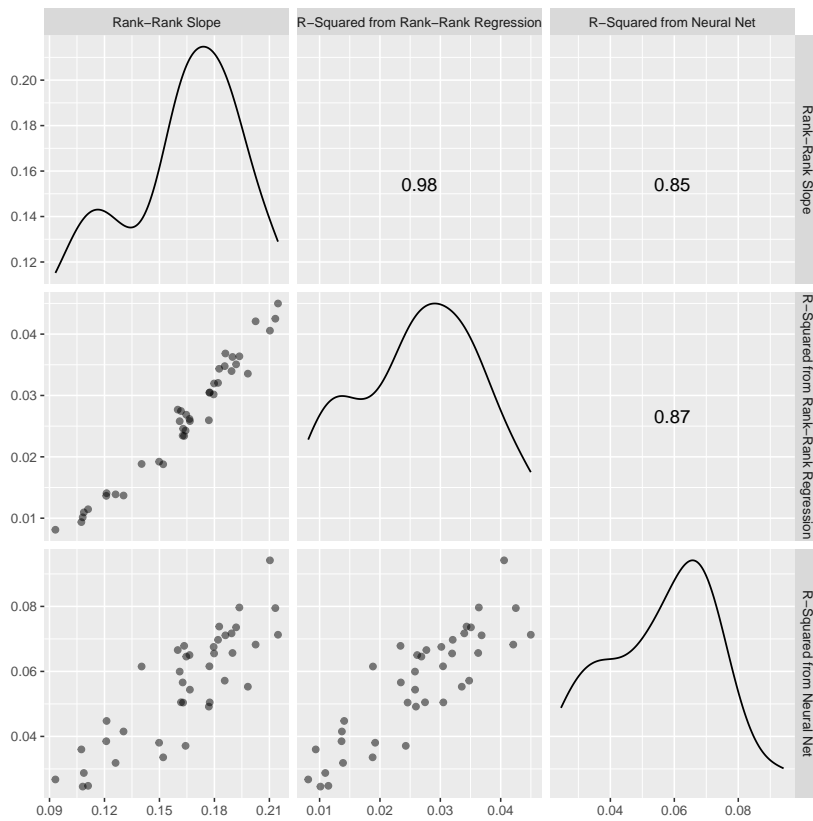
Notes: Relationships between the regional rank-rank estimates, R^2 from rank-rank income regression and R^2 from full model. Each point in the bottom left corner panels (scatter plots) is a single labour market region of Norway. Diagonal panels show distribution of estimates. Top right-hand corner panels show correlations relating to scatter plots diagonally across.

1.8 Appendix: Machine Learning methods

1.8.1 Elastic Net

Elastic Net[14] is a regularized regression which includes a linear combination of the L_1 and L_2 penalties. It can be viewed as an amalgamation of the better-known LASSO and ridge methods. Parameter estimates $\hat{\beta}$ are given by the solution to:

$$\min_{\beta} \left\{ \underbrace{\sum_{i=1}^{n}(y_i - \beta x_i)^2}_{\text{OLS}} + \lambda_1 \underbrace{\sum_{j=1}^{P}|\beta_j|}_{\text{LASSO}} + \lambda_2 \underbrace{\sum_{j=1}^{P}\beta_j^2}_{\text{Ridge}} \right\} \tag{1.7}$$

The addition of this penalty term induces the model to select variables with the most predictive power over the outcome, and shrinks coefficients towards zero. Relative to the standard OLS estimator, Elastic Net is less prone to overfit and hence more appropriate for prediction exercise. This however comes at the expense of introducing bias to coefficient estimates. Tuning parameters λ_1 and λ_2 are chosen via cross-validation, as described in the main text. A thorough introduction to Elastic Net is given in Zou and Hastie (2005).

1.8.2 Regression Tree

To understand tree-based methods, one must first understand regression trees.Breiman 2017 Regression trees are a series of univariate if-then state-

---

[14]We use the R package glmnet to implement elastic net. For details see `https://cran.r-project.org/package=glmnet`

ments, with each statement referring to a single predictor variable. For example, a simple regression tree is:

```
if X $>$ 3.5 then predict Y = 2.4
else predict Y = 5.3
```

These if-then statements can be layered, resulting in a set of mutually-exclusive partitions. Within each partition, observations are assigned the same prediction. The goal of the fitting process is to discover partitions resulting in groups which are similar in the outcome variable. Typically, the predicted value for each group is the average outcome for that group within the training data. In model fitting, a regression tree algorithm attempts many different splits at each stage, choosing that which delivers the most distinct groups based on the outcome variable.

When growing regression trees, the researcher must decide how "deep" to grow the tree. The deepness of the tree refers to the number of different splits involved in the full model. Growing a very deep tree will result in a highly flexible model, but this comes at the risk of overfitting as some of the observed group differences may be spurious. With an unconstrained number of splits, one could perfectly fit any dataset, however the predictive performance of such a model outside of the observed dataset is likely to be poor. On the other hand, growing an extremely shallow tree such as that given above could potentially over-simplify the patterns in the data and also deliver poor predictive performance. Obtaining the correct depth is an important part of tuning, which again is implemented through cross validation. The depth of the tree is controlled by a complexity parameter that controls how much in-sample predictive accuracy has to increase in order for a split to be implemented.

1.8.3 Random forests

While regression trees are attractive in their simplicity, in practice they are often found to be prone to overfitting, exhibiting high variance. Many successful models use aspects of regression trees, often estimating multiple trees and combining results in an ensemble. Random forests[15] improve upon regression trees by estimating many regression trees and introducing randomness in the estimation of each tree and then aggregating over all the resulting trees. This randomness is introduced in two ways. The first way is growing each of the trees on a bootstrapped sample from the training data, meaning that each of the trees uses different data. The second way is by only considering a random subset of the available predictors at each stage of growing the tree. Varying the number of predictors considered at each stage can increase or decrease the randomness, and cross-validation is one way to make this choice. Introducing randomness leads the grown trees to be somewhat uncorrelated, and the intuition is that aggregating over these trees should reduce the variance of the prediction and improve predictive performance over a single regression tree.

1.8.4 Gradient boosted trees

The general approach of boosting[16] refers to the process of fitting multiple models, with each model acting on the residuals of previous models (Freund and Schapire 1997). As the name suggests, in this method each of these sub-models is a regression tree. The boosting algorithm first grows a shallow

---

[15]We use the R package Ranger to implement boosting. For details see `https://cran.r-project.org/package=ranger`

[16]We use the R package XGBoost to implement boosting. For details see `https://cran.r-project.org/package=xgboost`

tree. Predictions are calculated using this simple tree, and compared to the observed values to generate residuals. A second shallow tree is then grown using the residuals from the first tree as the outcome. The model's predictions are then updated by including some function of the predictions the second tree. This iterative process of prediction on residuals and calculating new residuals is then repeated until a given number of trees have been grown. By recognizing the equivalence between residuals and negative gradients (J. H. Friedman 2001), it is possible to extend this method beyond mean-squared error loss functions. This process results in highly flexible predictions which are not prone to overfit.

Tuning parameters are set via cross validation, as discussed in the main text. These parameters include the total number of trees to grow, the extent to which each additional prediction is incorporated into the overall model (the learning rate) and the depth of each individual tree.

1.8.5 Neural networks

Neural networks[17] are a form of non-linear statistical model that has become increasingly popular with the growing availability of data and computation power. Conceptually, a neural network creates derived "features" from linear combinations of predictors, and then produce a prediction as a nonlinear function of those derived features (Hastie, Tibshirani, and J. Friedman 2009). An illustration of how this works can be seen in Figure 1.6.

The input layer consists of one neuron for each of the model's predictors (variables). Then there are connections between the neurons in the input layer

---

[17]We use the R package neuralnet to implement boosting. For details see `https://cran.r-project.org/web/packages/neuralnet/index.html`

Figure 1.6: Illustration of neural network
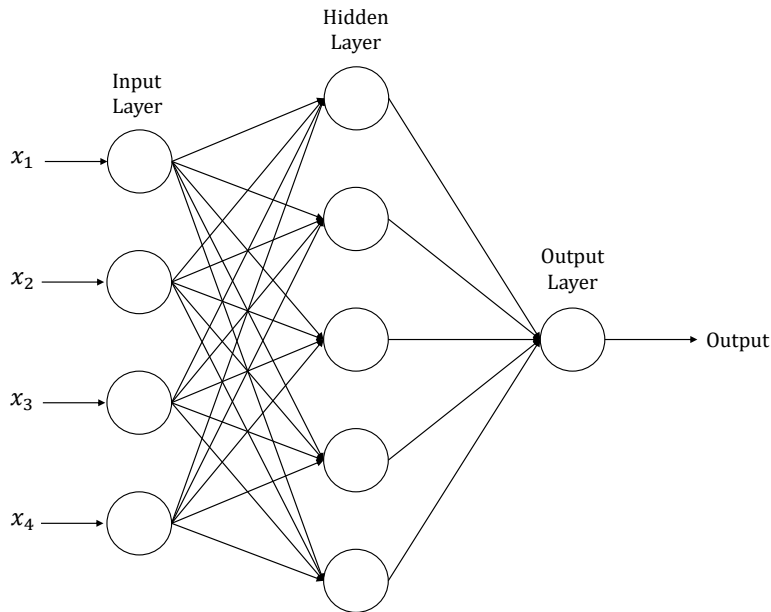
Notes: Neural network with a single hidden layer of 5 nodes. Each circle represents a neuron, each line a connection.

and the neurons in the next layer. All intermediate layers are labeled 'hidden' layers. The estimated strength of these connections between neurons, often referred to as weights, determines how the linear combinations of the inputs form derived features. The neurons in the hidden layer then transforms the inputs to outputs through a non-linear activation function (e.g. logistic function).

Finally, the output layer combines the outputs from the hidden layer using estimated weights to form predictions. As long as the network contains sufficiently many nodes, this approach allows neural networks to approximate highly complex functions. Based on this discussion, it is clear that there are a number of choices to make when estimating a neural network. These choices include the number of hidden layers, the number of nodes in each of the hidden layers and the activation functions. The flexibility of neural networks makes them prone to overfit, so often additional regularization is added to

45

the framework described above.

1.9 Additional Tables and Figures

Figure 1.7: Income distributions
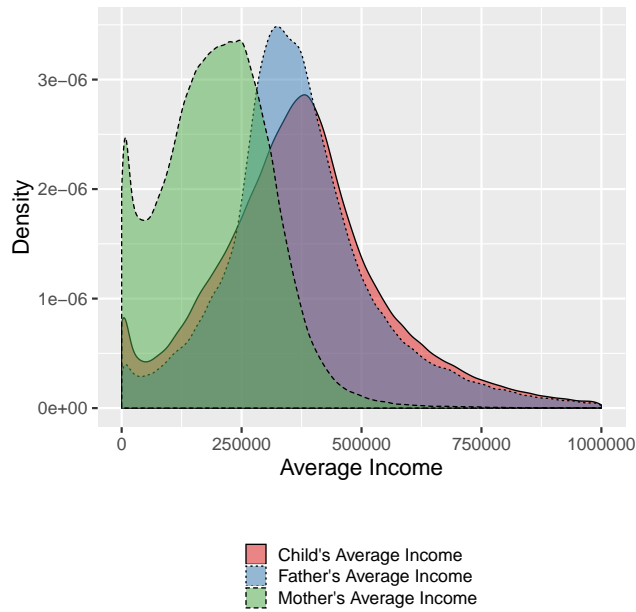
Notes: Income distributions of children, mothers and fathers in full sample. See Section 1.5 for a description of data construction.

Figure 1.8: Parent years of education
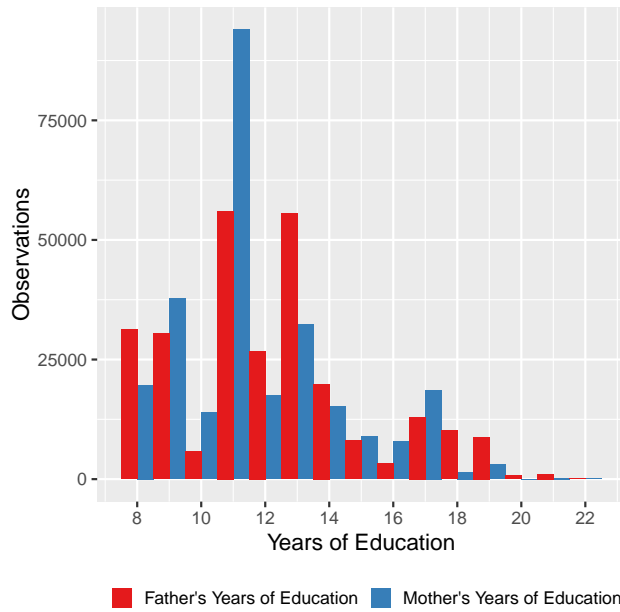
Notes: Years of education of mothers and fathers in full sample. See Section 1.5 for a description of data construction.

Figure 1.9: Parent wealth distribution
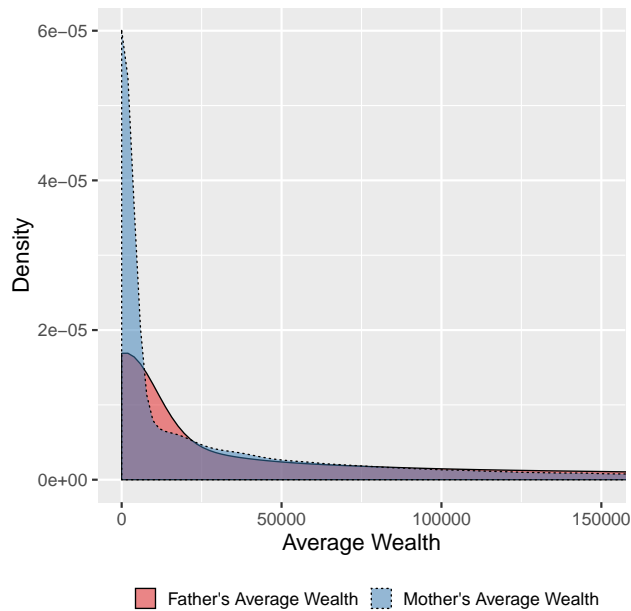
Notes: Wealth distributions of mothers and fathers in full sample. See Section 1.5 for a description of data construction.

Figure 1.10: Mean child income rank by father education
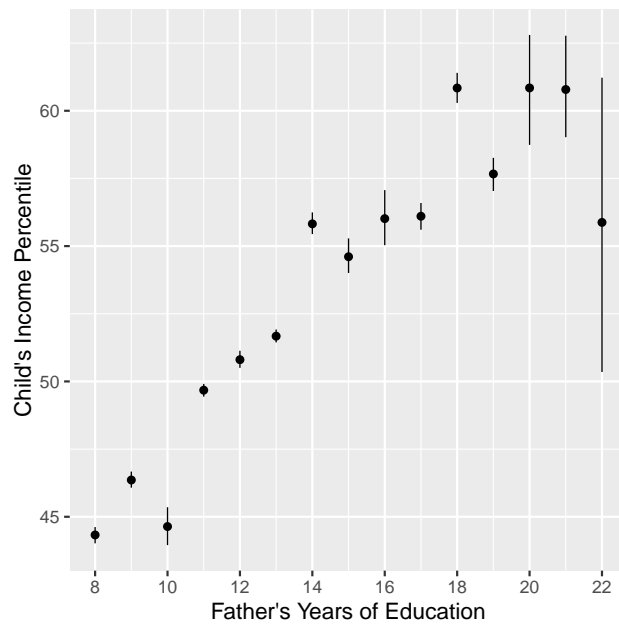
Notes: Child income averaged across ages 30 to 35 before ranks generated. Error bars correspond to bootstrapped 95% confidence intervals. Further information on variable construction in Section 1.5.

Figure 1.11: Mean child income rank by mother education
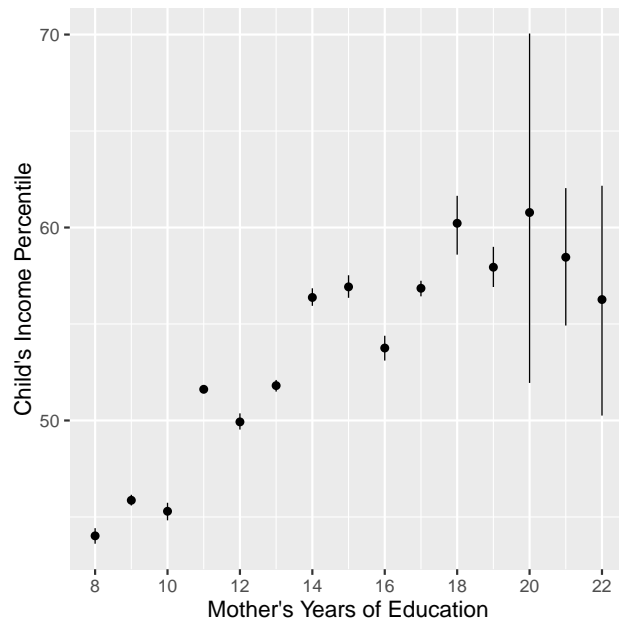
Notes: Child income averaged across ages 30 to 35 before ranks generated. Error bars correspond to bootstrapped 95% confidence intervals. Further information on variable construction in Section 1.5.

Table 1.2: Hold-out Test set results

	variables	model	Test R2	completeness
1	Extended	ElasticNet	0.055	1
2	Extended	NeuralNet	0.055	1
3	Extended	OLS	0.05	0.91
4	Extended	Ranger	0.055	1
5	Extended	XGBoost	0.055	1
6	Income (flexible)	ElasticNet	0.041	0.75
7	Income (flexible)	OLS	0.041	0.75
8	Income and education length	ElasticNet	0.042	0.77
9	Income and education length	OLS	0.042	0.77
10	Income and education length	Ranger	0.019	0.35
11	Income and education length	XGBoost	0.045	0.82
12	Income and wealth	ElasticNet	0.045	0.82
13	Income and wealth	OLS	0.045	0.82
14	Income and wealth	Ranger	0.029	0.53
15	Income and wealth	XGBoost	0.047	0.86
16	Income, wealth and education length	ElasticNet	0.049	0.9
17	Income, wealth and education length	OLS	0.049	0.9
18	Income, wealth and education length	Ranger	0.039	0.71
19	Income, wealth and education length	XGBoost	0.053	0.97
20	Rank-Rank	OLS	0.037	0.68

Notes: R^2 performance in test set. Completeness is given by dividing each model's R^2 by the R^2 in Model 2.

Figure 1.12: Mean child income rank by parent wealth
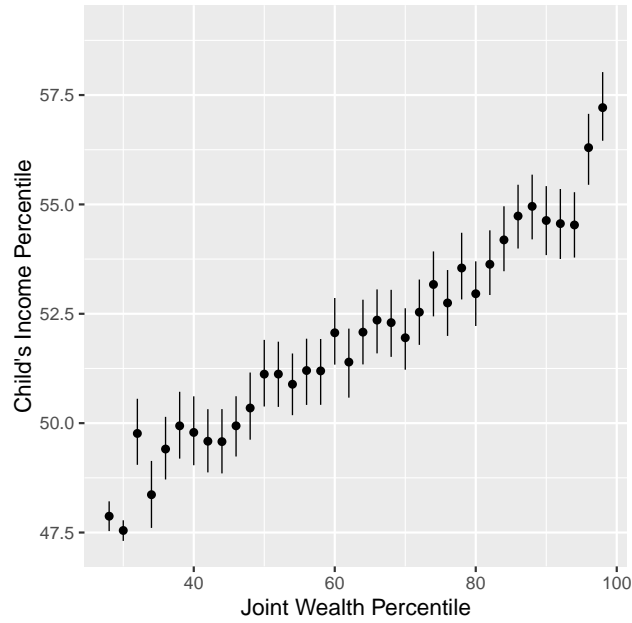
Notes: Child income averaged across ages 30 to 35 before ranks generated. Error bars correspond to bootstrapped 95% confidence intervals. See Section 1.5 for description of wealth measure construction.

Table 1.3: Training set results

	variables	model	Training R2 (mean)	Training R2 (sd)
1	Extended	ElasticNet	0.058	0.0038
2	Extended	NeuralNet	0.058	0.0052
3	Extended	OLS	0.052	0.0028
4	Extended	Ranger	0.058	0.0047
5	Extended	XGBoost	0.058	0.0042
6	Income (flexible)	ElasticNet	0.042	0.0022
7	Income (flexible)	OLS	0.042	0.0022
8	Income and education length	ElasticNet	0.044	0.0028
9	Income and education length	OLS	0.044	0.0027
10	Income and education length	Ranger	0.021	0.0032
11	Income and education length	XGBoost	0.045	0.003
12	Income and wealth	ElasticNet	0.047	0.003
13	Income and wealth	OLS	0.047	0.003
14	Income and wealth	Ranger	0.031	0.0032
15	Income and wealth	XGBoost	0.048	0.0033
16	Income, wealth and education length	ElasticNet	0.051	0.0036
17	Income, wealth and education length	OLS	0.051	0.0036
18	Income, wealth and education length	Ranger	0.043	0.0041
19	Income, wealth and education length	XGBoost	0.054	0.004
20	Rank-Rank	OLS	0.037	0.0021

Notes: Mean and standard deviation of R^2 performance in training set.

Table 1.4: Regional estimates

	obs	completeness	R2 (rank)	R2 (neural net)
1	7481	0.26	0.009	0.036
2	1828	0.303	0.008	0.027
3	1218	0.306	0.019	0.062
4	2819	0.314	0.014	0.045
5	2777	0.33	0.014	0.042
6	5415	0.345	0.023	0.068
7	1686	0.354	0.014	0.039
8	7343	0.381	0.011	0.029
9	8887	0.402	0.026	0.065
10	3226	0.413	0.01	0.025
11	8508	0.414	0.023	0.057
12	2699	0.416	0.027	0.065
13	1513	0.416	0.028	0.067
14	5545	0.43	0.026	0.06
15	1867	0.431	0.041	0.094
16	4034	0.436	0.014	0.032
17	4609	0.447	0.03	0.068
18	3592	0.457	0.036	0.08
19	3985	0.46	0.032	0.07
20	2626	0.462	0.011	0.025
21	2640	0.465	0.034	0.074
22	3697	0.474	0.034	0.072
23	10281	0.475	0.026	0.054
24	7740	0.477	0.035	0.074
25	27729	0.488	0.025	0.05
26	2361	0.488	0.032	0.066
27	2639	0.494	0.03	0.062
28	17048	0.505	0.019	0.038
29	21359	0.518	0.037	0.071
30	1437	0.528	0.026	0.049
31	4840	0.535	0.043	0.079
32	7510	0.544	0.027	0.05
33	72922	0.552	0.036	0.066
34	2743	0.56	0.019	0.034
35	1547	0.604	0.03	0.05
36	3274	0.607	0.034	0.055
37	1299	0.608	0.035	0.057
38	2585	0.617	0.042	0.068
39	2079	0.631	0.045	0.071
40	1774	0.654	0.024	0.037

Notes: Full estimates from the 46 labour market regions of Norway, omitting 6 areas with fewer than 1000 observations. Sorted by completeness.

Chapter 2

Status Traps and Human Capital Investment

Aline Bütikofer, Erling Risa and Kjell G. Salvanes[*]

Abstract

Although intergenerational income mobility is high in Nordic countries, parental education still plays an important role in explaining educational attainment. Using machine learning techniques, we show that, in Norway, obtaining a college degree is not a continuous function of parental years of education and that there are discontinuities and interactions at different parental education levels. Parental earnings and the transmission of cognitive ability are not the only reasons for the status traps in education. Moreover, our findings suggest that parental education can compensate for lower cognitive ability, whereas paternal earnings cannot compensate for low parental education.

2.1 Introduction

Access to higher education is widely seen as a pathway to economic success, better health and well-being. Whereas income inequality is low and intergenerational income mobility is high in Nordic countries compared with other OECD countries (Black and Devereux 2011; Corak, Lindquist, and Mazumder 2014), the intergenerational persistence in education is at a similar level in the Nordic countries as in the US (Björklund and Salvanes 2011; Hertz 2007; Landersø and Heckman 2017). That is, parental education still plays an important role in explaining educational attainment, particularly in, university education. The strong role of parental background in educational attainment is rather surprising in the context of the social welfare state and an environment in which education is freely available and a generous scholarship system exists. This puzzling phenomenon raises the question at which margins along the educational distribution this strong persistence arises and whether there are specific dimensions of the family background or the interplay of various dimensions that are the most important drivers for the intergenerational persistence in education.

In this paper, we use machine learning methods to analyze these margins and the interplay of different dimensions of family background in the context of one Nordic country—Norway—to get a better understanding of the mechanisms behind this persistence. Although the level of education in Norway has been increasing steadily since the 1930s, there is a strong intergenerational persistence in education. Figure 2.1 illustrates that there is a clear gradient in father's years of education and the share of children who completed a bachelor or a master degree. Whereas less than 20 percent of chil-

dren with fathers who have only eight years of education obtain a three-year bachelor degree, more than 50 percent of the children of fathers with a Ph.D. obtain a five-year master degree. The mechanisms through which social background influences educational and labor market outcomes are likely highly complex and include both nonlinear effects and interactions between various dimensions of the social background. Recent literature discusses, for example, the role of neighborhoods or regions in shaping social mobility (e.g. Chetty, Hendren, Kline, and Saez 2014; Chetty and Hendren 2018a; Chetty and Hendren 2018b). Hence, the social background is not limited to an individual's parents but includes an individual's neighbors. Moreover, several papers show that the traditional parent–child model does not sufficiently describe social mobility and will underestimate the long-term persistence of social status across generations. In particular, this literature documents a much higher persistence across multiple generations (Braun and Stuhler 2018; Lindahl, Mårten Palme, Massih, and Sjögren 2015; Long and Ferrie 2018; Clark 2014; Mare 2011) or family dynasties (Adermon, Lindahl, and Marten Palme 2019; Jæger 2012) both in income and education than a simple model would capture.[2] That is, grandparents, aunts, and uncles form another dimension of social background. In addition, there is empirical evidence that large economic shocks or educational reforms break intergenerational persistence (Nybom and Stuhler 2013; Butikofer, Dalla Zuanna, and Salvanes 2018). Hence, the various dimensions of social background might have a different importance at different times.

To study these status traps in human capital investments across generations

---

[2]Other recent approaches using a latent variable approach such as the role of surnames across generations, for instance, Güell, Rodríguez Mora, and Telmer (2015) and Clark (2014), also find higher long-term intergenerational social persistence than in simpler models. These results are debated and it can be argued that these relationships may be spurious (Solon 2018). Moreover, group-level estimates, as in Clark (2014), are distinct from estimates of the traditional parent–child parameter (Chetty, Hendren, Kline, and Saez 2014; Solon 2018).

and to understand differences in human capital investment across families, we use a model-based recursive partitioning approach developed by Zeileis, Hothorn, and Hornik (2008). By leveraging rich administrative data and machine learning techniques that allow for highly flexible functional forms, we explore how the interplay of various background dimensions determine educational attainment. Hence, this approach is well-suited in this setting because it provides a way of flexibly uncovering nonlinear and interaction effects while preserving a reasonable amount of interpretability. This paper, therefore, contributes to the recent literature on intergenerational mobility in two distinct ways: first, we do not need to assume linearity and, second, we are not bound merely to analyze a single aspect of social background such as parental earnings, parental education or neighborhoods but are able to capture more complex relationships. Because university education is often described as a pathway to greater job market opportunities, we focus on inequality in access to universities and elite educations that offer the best chances of labor market success. Hence, we study—indirectly—mobility into high-income jobs. In particular, we study whether individuals complete bachelor degrees, master degrees, and master degrees from elite programs. Whereas the percent of individuals with bachelor degrees has doubled from birth cohort 1955 to birth cohort 1980, the number of individuals with master degrees only increased by 50 percent. Hence, the different dimensions of the social background might be important drivers of intergenerational persistence for each educational margin. Advancing the understanding of the interplay of different dimensions of family background in intergenerational education persistence for these three educational margins will inform the role of public policy in reducing these educational status traps (Durlauf, Kourtellos, and Tan 2017). Because the patterns in intergenerational mobil-

ity in education are similar across OECD countries, the findings might be generalizable to other countries (OECD 2014).

Using this machine learning framework and registry data of individuals born from 1955 to 1980 and their parents, we first uncover strong nonlinearities in intergenerational persistence in educational attainment. That is, we show that the educational attainment of children is not a continuous function of parental years of education. There are clear discontinuities at different parental education levels. For example, there is a large jump in the likelihood that girls obtain a master degree (about 8–10 percentage points) if the girl's father has 17 years of education (corresponding to a completed master degree) compared with 16 years of education (corresponding to started but not completed a master degree). We further explore borrowing constraints as an explanation for educational status traps and show that parental income is a much less important predictor of children's education once mother's and father's education is controlled for. In particular, there are some nonlinearities in parental earnings toward the bottom of the earnings distribution for obtaining bachelor degrees and toward the very top of the earnings distribution among highly educated parents to explain elite education attendance of daughters. For master degrees, we do not find any non-linearalities or interactions for parental earnings. As a second channel, we study cognitive ability (for men) that is both highly correlated with college attendance rates (e.g., Belley and Lochner 2007) and with parental cognitive ability (Black, Devereux, and Salvanes 2009). Whereas we present evidence that cognitive ability is a good and non-linear predictor of educational attainment, we show that our main findings are not solely driven by the transmission of cognitive ability between generations and that mothers' and fathers' education level is still very important. There are also interesting interactions between

parental education and cognitive ability and our findings suggest that father's education can compensate for son's lower cognitive ability in terms of obtaining educational degrees. Moreover, we show that grandfathers' education, aunts' and uncles' education, as well as interactions of all these dimensions, are much less important than mothers' and fathers' education. Although there is regional variation in educational attainment and although there is an expansion in the number of colleges and in access to college loans in Norway during the time we analyze, we find that neither geographic areas nor specific years are important predictors once we control for mothers' and fathers' education. Last, we show that the patterns for boys and girls are relatively similar.

The paper unfolds as follows: Section 2 outlines the institutional context. In Section 3, we present the data used in the analysis. Section 4 discusses the methodological approach used in the analysis and Section 5 presents our findings. Section 6 provides concluding remarks.

2.2   Institutional Context

The Norwegian education system consists of four levels: primary school (grades 1–7), middle school (grades 8–10), high school (grades 11–13), and college and university education.[3] Currently, Norwegian compulsory education starts at age six, lasts for 10 years and consists of primary and middle school. Compulsory schooling is organized by Norwegian municipalities and the vast majority (98%) of pupils attend public, local schools. All pupils are allocated to schools based on fixed school catchment areas within mu-

---

[3]Note that middle school is sometimes called lower secondary education, high school upper secondary education, and college or university higher education.

nicipalities. Except for some religious schools and schools using specialized pedagogic principles such as Montessori schools, there is no school choice among publicly funded schools. There exist no grades in primary school. Nevertheless, the school system gets more competitive from middle school onward, where exit exams and grades from teachers are crucial for entry to the best high schools. Note that, for individuals born before 1947, the number of compulsory school years was seven years. Hence, attending middle school was voluntary, but not uncommon. The mandatory school years were then increased to nine years (Black, Devereux, and Salvanes 2005) and in 1997, the mandatory school years were increased to ten years. Moreover, the school starting age was seven years prior to 1997.

High schools have two main tracks: vocational and academic. The academic track prepares students to attend college and there are different specializations within the academic track (e.g., science, languages, music). Within the vocational track, there are currently nine distinct programs such as "health and child development", "restaurant and food", and "construction" teaching students professional qualifications. High schools are administered at the county level (above the level of municipalities). High school is not mandatory in Norway, however, since the early 1990s all students graduating from middle schools are guaranteed a slot in high school. However, many high schools, tracks and programs are oversubscribed. Admission procedures differ across counties and over time. Whereas students are allocated to schools based on catchment areas in some counties, pupils in most counties can currently freely choose schools within their county based on their grades and test results from middle school. In counties with high school choice, a county's central school authority matches students to schools, tracks, and programs based on their middle school score and their preference listing in

a deferred acceptance procedure. Typically, a student's preference listing would contain six schools and different tracks and study programs at each school. Although grade cutoffs vary by year, parents and students are aware of what programs a student might approximately qualify for based on his or her grades when submitting the preference list as grade cutoffs for popular schools are reported in newspapers. In counties with catchment areas, students would typically list three tracks or programs available at their local high school and would be allocated to the highest listed choice they qualify for based on their middle school grades. Note that although, currently, most students are in a system with free high school choice, admission based on catchment area was the most common for early cohorts in our sample. About 95 percent of students enroll in high school the same year they finish middle school. About 45% of students enroll in the academic track, the rest of the students enroll in the vocational track. For students in the academic track, the high school exit exam together with grades from the teachers form a grade average point (GPA) that determines who is being accepted to different higher education institutions.

Higher education consists of universities, scientific schools, and university colleges. Until the late 1990s, when the majority of our sample attended higher education, Norway had three universities (Oslo, Bergen, and Tromsø) offering a very wide range of study subjects including medicine, law, humanities, and social science, and six scientific schools offering specific subjects at university level such as business schools, engineering schools, architect and design schools, and arts schools. In 2019, there were ten universities (most of them established after 2000) and nine scientific schools. Since the early 2000s, Norwegian universities and scientific schools offer three-year bachelor and five-year master degrees. Students would usually receive a

bachelor degree after three years and then continue for two additional years to receive a master degree. Before 2000, a degree from these institutions would normally last 3–6 years. In addition, there are many regional university colleges offering mostly professional degrees in business, health, and teaching. Today, these institutions mostly offer bachelor degrees. Before 2000, university colleges offered two- to four-year degrees. Most students attend a public institution, and even private institutions are funded and regulated by the Ministry of Education and Research. There are generally no tuition fees for attending public higher education in Norway, and most students are eligible for financial support (part loan/part grant) from the Norwegian State Educational Loan Fund. Admission to a combination of detailed field and institution (e.g., law at the University of Oslo) is based on high school GPA (a combination of teachers' grades and centralized exit tests). Since the late 1990s, the admission process to higher education is centralized (see Kirkeboen, Leuven, and Mogstad 2016). Before this reform, the admission process was organized by each educational institution. Nevertheless, during the whole period we study, admission to degrees such as medicine and admission to the science schools was very competitive.

Whereas obtaining an elite education in the US is often defined by attending highly competitive, private institutions with high tuition fees such as Ivy League colleges (see, e.g., Chetty, J. N. Friedman, Saez, Turner, and Yagan 2019), elite education in the context of Norway is often described as specific degrees at specific institutions with the best earnings outcomes. Like the US, access to these degrees is highly competitive. Important differences, however, are that there are no tuition fees for these degrees and no easier access for legacy students because admission is solely based on high school GPA. Among the elite programs are, for example, master degrees in law,

medicine, and economics at the University of Oslo, economics and business administration at the Norwegian School of Economics, or engineering at the Norwegian University of Science and Technology.

Since the 1930s, the level of education in Norway has been increasing steadily with an accelerated increase after World War II (see Figure 2.2). Among recent cohorts, 40 percent have at least a three-year bachelor degree and Norway ranks among the countries with the highest education levels within the OECD. However, the increase in education varies by type of degree. Whereas the share of individuals with at least a bachelor degree doubled from birth cohort 1955 to 1980, the share of master degree graduates increased by 50 percent (see Figure 2.2). Figure 2.3 documents that this increase in children's education levels is, to a large degree, driven by individuals with less-educated fathers. Whereas, for example, the share of individuals who obtain a bachelor degree hardly changes between birth cohort 1955 and 1980 if they have very well-educated fathers, the share doubles for individuals whose fathers have a very low level of education. These clear variations by educational margin suggest that the interplay of different background dimensions might vary for each of the three educational outcomes we study.

2.3   Data

The primary data source is Norwegian administrative data from different administrative units such as tax authorities and educational institutions and census data from 1970 and 1980. In this paper, we exploit two features of these data: the long panel structure and the parent–child link. The panel structure allows us to identify where and when people are born and follow them for several decades in the data. Hence, we are, for example, able

to construct characteristics of the regions they grew up and link those to medium- and long-term educational outcomes. The fact that the data contain unique personal identifiers enables us to link children and their parents, grandparents, and siblings, as well as aunts and uncles.

We focus on the children born from 1955 to 1980 to ensure that we have enough education and income data for their parents as well as adult outcomes of the children. A detailed description of the different variables in the data and descriptive statistics are provided below.

2.3.1 Education

Our primary outcome is whether a child has education at the bachelor, master, or elite education level. These measures come from the national education database, which contains codes for the highest completed level of education. These codes are in the NUS2000 format, which is a six-digit code containing highly detailed information on both the level and field of a person's education. The level of a degree is given in the first digit. Individuals whose highest completed level of education has a NUS2000 starting with a six, which corresponds to short tertiary education (Barrabés and Østli 2016), are defined as having education at the bachelor level or higher. We define someone as having education at the master level if the NUS2000 code of their highest completed level of education starts with a seven or an eight, corresponding to long tertiary education such as a master degree and doctoral education, respectively (Barrabés and Østli 2016). Moreover, we study whether a child obtains a degree from a highly competitive elite program. The elite programs are defined as degrees at the master level or above in law, medicine, economics, economics and business administration

at the Norwegian School of Economics or engineering at the Norwegian University of Science and Technology. Note that college and university education at public institutions is free and that completing a master degree at an elite institution is not more expensive than at any other public university or university college. Because the youngest cohort in our sample was born in 1980, the highest completed level of education is measured when the children are at least 35 years of age, meaning that the vast majority have completed their education.

Figure 2.4 documents why all these three educational margins are interesting. In particular, the figure visualizes that the likelihood of achieving top earnings percentiles is substantially higher for individuals with a master degree than for individuals with a bachelor degree and that there is a significant financial return to a degree from an elite institution. Moreover, Table 2.1 lists the three most common occupations for each education type and the mean earnings percentile that graduates achieve. At each margin, there are clear differences in occupations and in earnings. Hence, different parental characteristics might be important determinants of prestigious degrees that have traditionally been pathways to a range of high-profile positions than of less-prestigious degrees.

We measure parental education in years of education using the same national education database as for children. We chose to look at years of education rather than educational level because it allows us to identify nonlinearities at a more detailed level. The measure of grandfather's education is years of education and is based on reported educational attainment in the 1960 census.

2.3.2 Earnings

Our earnings measure comes from administrative tax records beginning in 1967 and includes mean yearly pension-generating earnings. Pension-generating earnings consist of labor market income from wages and self-employment as well as some transfers from the social security system such as unemployment- and sickness benefits. Parental earnings are measured when the parents are between 40 and 50 years of age. We chose this age range to ensure that we observe earnings for most of the parents in all our child cohorts. Children's earnings are measured when individuals are between 30 and 35 years of age. Because of the advantages emphasized in several papers (e.g. Chetty, Hendren, Kline, Saez, and Turner 2014; Pekkarinen, Salvanes, and Sarvimäki 2017), such as lower sensitivity to the age of measurement and handling of zero income observations, we chose to include parents earnings as percentile ranks rather than levels.

2.3.3 Cognitive Ability

Our data include a measure of cognitive ability for men that was obtained when they were evaluated for military service at the age of 18—19. The test used to measure cognitive ability consists of three parts, one focusing on arithmetic, one on word similarities and one on figures. The scores from the three parts of the test are aggregated and reported on the standard nine scale, which ranges from one to nine, with a mean of five and a standard deviation of two (Thrane 1977; Sundet, Barlaug, and Torjussen 2004; Sundet, Tambs, Harris, Magnus, and Torjussen 2005). Because military service was compulsory for all males in the cohorts we are analyzing, we have the

measure of cognitive ability for more than 98% of all males in the sample. Military service was not compulsory for women in this time period. Hence, most women were never evaluated for military service and we do not include this measure of cognitive ability for the women.

2.3.4 Geographical Characteristics

The data include information about the municipality of birth. We aggregate the more than 400 municipalities (the smallest political entity in Norway) to 46 labor market areas based on commuting patterns (M. S. Bhuller 2009). We use labor market regions as the unit of aggregation because we believe they provide sufficient control for geographical differences with fewer geographical units than municipalities. It is advantageous to use fewer geographical units when estimating the logistic regressions in the analysis, especially once we start partitioning the sample into subgroups. In the main specification, we control for geographical characteristics by including local labor market fixed effects.

2.3.5 Summary Statistics

We focus on children born from 1955 to 1980. Because we are interested in the effect of multiple parental background characteristics, the final sample contains the children for whom we have data on parents' earnings and education. After dropping observations with missing background characteristics, we obtain a sample of roughly 950,000 individuals. Summary statistics for this sample are shown in Table 2.2. The three panels represent three different samples. The first panel provides summary statistics for our main sample.

In our main sample, about 34% of the children have a bachelor degree, 10% a master degree, and 3% a degree from an elite program. Fathers' average years of education are 11.6 and mothers' average years of education are 11 years. The second panel shows the means and standard deviations for the sample where we observe a child's cognitive ability. Because we only have this information for Norwegian men born after 1950, the sample does not include women, is substantially smaller, and individuals are, on average, somewhat younger. The sample averages are, however, like the full sample. The third panel shows summary statistics for the sample of individuals where we observe both education of paternal grandfathers and education of at least one aunt or uncle. Hence, all individuals who do not have an aunt or uncle are excluded and individuals are, on average, slightly younger because we rarely observe information on three generations for the oldest cohorts in our main sample. Importantly, the mean values of the variables of interest are not substantially different from those for our main sample.

2.4 Methodology

This paper aims to improve our understanding of how human capital transfers from parents to children by identifying whether there are important nonlinearities and interactions in the transmission process. Because this goal is fundamentally about describing patterns in the data, it is well-suited for applying machine learning techniques. Among the many machine learning techniques available, we have chosen the model-based recursive partitioning technique developed by Zeileis, Hothorn, and Hornik (2008). This approach resembles other classification and regression tree approaches in the sense that it partitions the data into groups based on an objective function. However,

unlike most other tree-based methods that aim to split the data into groups that are similar in terms of an outcome, this algorithm seeks to create groups where the estimated parameters in a statistical model are similar. Hothorn and Zeileis (2016) explain the algorithm used to accomplish this as follows:

1. Fit the desired model on the entire sample.

2. Test the estimated parameters in the model for instability with respect to one or more partitioning variables. If the hypothesis tests reject parameter stability, split the sample using the partitioning variable with the lowest p-value. If parameter stability is not rejected the algorithm stops. We correct the p-values for multiple hypothesis testing using the Bonferroni correction. We also require the p-values to be below 0.01 to qualify for a partition. Lastly, we require the partition to lead to an improvement in the Akaike Information Criterion (AIC).

3. In cases where parameter stability is rejected, the algorithm then searches for the optimal split in the partitioning variable chosen in step 2 using the objective function of the estimated model.

4. The model is then refitted in both subsamples and the algorithm keeps iterating until no more partitions are found.

The approach of partitioning the data into groups with similar estimated effects of various background characteristics is attractive in this setting because it provides a way of flexibly uncovering non-linear and interaction effects while preserving a reasonable amount of interpretability.

The baseline model we estimate is:

$$
\begin{aligned}
\text{Education\_Child}_i = {}& \beta_0 + \beta_1 \text{Eduy\_Father}_i + \beta_2 \text{Eduy\_Mother}_i \\
& + \beta_3 \text{Earnperc\_Father}_i + \beta_4 \text{Earnperc\_Mother}_i \qquad \text{(2.1)} \\
& + \xi \text{Yob\_Child}_i + \gamma \text{Labor\_Market}_i + \varepsilon_i,
\end{aligned}
$$

where $Education\_Child_i$ signifies one of the three educational outcomes we analyze for child i, $Eduy\_Father_i$ and $Eduy\_Mother_i$ are the years of education of the father and mother of child i, $Earnperc\_Father_i$ and $Earnperc\_Mother_i$ are the earnings percentile of the father and mother of child i, Yob_Child_i is the year of birth of child i and $Labor\_Market_i$ the labor market of birth of child i. Because we are working with binary outcomes, we have chosen to estimate the equation using logistic regression.

We test for parameter instability with respect to earnings and education of both parents, as well as the child's year of birth, gender, and macro-region of birth. Note that the model allows us to test for instability with respect to gender although we do not control for gender in the empirical model. Whereas the model contains fixed effects for the 46 labor markets of birth, we test for instability across five macro-regions (Eastern, Southern, Western, Middle, and Northern Norway) to limit the number of possible interactions with all the other variables. Testing for nonlinearities and interactions among these characteristics while controlling for differences between birth cohorts and across geographical regions allows us to provide a good indication of some important determinants of children's educational outcomes. Because we are flexibly looking for interactions and nonlinearities, there is also interesting information on what dimension we do not find any instabilities. For instance, if we do not detect nonlinear or interaction effects toward the bottom of the earnings distribution, this would suggest that there is no threshold value at

which credit constraints become an important factor.

The final methodological choice we have made relates to overfitting, which is a common concern in machine learning approaches such as this one. Overfitting in this context would imply identifying spurious nonlinearities or interactions in the data. Zeileis, Hothorn, and Hornik (2008) argue that the hypothesis testing approach implemented in the model-based recursive partitioning algorithm makes it far less prone to overfitting than several other machine learning techniques. Despite this, we have chosen to be even more conservative than the algorithm is by default by limiting the depth of the number of splits permissible after one another to reach the final subsamples to four.[4] The primary reason for this is preserving interpretability, but it also has the added benefit of making overfitting less likely. This can also be seen in Figure 2.5, which shows the results from 10-fold cross-validation of predictions from Equation 2.1 with obtaining a master degree as the outcome. The figure documents two important points. First, we see that a maximum depth of five results in a mean squared error close to the minimum of the graph, indicating that we are not overfitting. Moreover, we see that the mean squared error with a maximum depth of four is noticeably higher, which implies that restricting the model further would lead to worse predictive performance. It is also interesting to note that the mean squared error remains about the same when increasing the maximum depth beyond five. This is because the algorithm only provides a few more partitions beyond the ones found with a maximum depth of five if left unrestricted, supporting the claim made by Zeileis, Hothorn, and Hornik (2008) that the algorithm is quite robust to overfitting.

---

[4]This implies that we restrict the maximum depth of the tree produced by the algorithm to five.

2.5 Results

Below, we discuss our findings for three educational degrees of interest: bachelor degree, master degree, and degrees from an elite program. Because master degrees are a pathway to high earnings and prestigious occupations (see Figure 2.4 and Table 2.1), we describe these results first and then compare them with the findings for the two other educational margins. For each outcome, we discuss whether parental earnings or cognitive ability are potential drivers of the strong intergenerational persistence in education.

2.5.1 Master Degree

Longer university educations such as master degrees have significant financial returns and arguably give access to jobs with additional nonpecuniary benefits (see Figure 2.4). Figure 2.6 shows the split points from running the partitioning algorithm with Master's degrees or higher education as the outcome. The algorithm splits the sample first based on whether the father has more or less than 13 years of education. Having more than 13 years of education corresponds to having at least some college education. The hierarchy of the splits indicates the importance of the nonlinearities. Hence, the likelihood of obtaining a master degree changes the most when going from fathers with 13 to fathers with 14 years of education. Independent of the education of the father, the next split is based on the gender of the child. This suggests that, depending on social background, men and women are not equally likely to complete master degrees. Although gender differences are small, women are more likely to obtain a master degree. This is particularly true if their mothers are well educated. Among the lower-educated fathers,

70

mother's years of education is the next crucial variable. That is, for both boys and girls, the sample is first split at whether the mother has 10 years or less of education or more than 10 years of education and then at 13 years of education. Having more than 10 years of education corresponds to having at least some high school education. For boys with highly educated fathers, the sample is split based on mothers having more than 15 years of education (one year more than a bachelor degree). For girls with highly educated fathers, having a father with education above the master level (more than 17 years) is the next instability that the model detects. Interestingly, the nonlinearities and interactions are not at completed high school, bachelor degrees, or master degrees but at one or two years above. This might indicate that parents who started, but did not completed the next level of education, are more encouraging toward long college educations such as master degrees. Overall, in a highly flexible specification, allowing for nonlinearities and interactions between parental education, parental income, the region of birth, year of birth and gender, significant nonlinearities and interactions exist only for parental education and gender. There are no split points in parental earnings percentiles indicating that credit constraints are not the main obstacle to obtaining a long university education. Moreover, there are no split points in the region of birth or year of birth. Even though there are macroeconomic regions without a university until the 1990s, and our data span significant changes in access to scholarships for students from all social backgrounds, there are no parameter instabilities in regions or birth years.

How important are these background variables in determining educational outcomes? Figure 2.7 plots the predicted likelihood of obtaining a master degree by father's and mother's education for boys and girls separately. In this and all other figures plotting predictions, we fix all other background

71

characteristics at their sample medians to only focus on the estimated effects of the variables featured in the plot. The horizontal and vertical lines represent the split points discussed in Figure 2.6. Note that because there were only splits with respect to gender and parental education, we can adequately describe the extent of nonlinearities and interactions using two dimensions. Darker colors indicates a higher predicted likelihood of obtaining a master degree. For both genders, the estimated association of father's and mother's education and the educational outcome are relatively symmetric. For example, girls whose mothers have only compulsory education and whose fathers have a Ph.D. have a predicted likelihood of obtaining a master degree of 25%; in the opposite case, the predicted likelihood of obtaining a master degree is 27%. Hence, a very highly educated mother compensates somewhat for having a lowly educated father. Moreover, there is a strong estimated effect of having two highly educated parents, because the predicted probability of obtaining master degrees among girls where both parents have a Ph.D. is 65%. Keeping the mother's education at 13 years, the likelihood that a boy or a girl obtains a master degree changes from 10 percent when the father has nine years of education to 37–39 percent when the father has a Ph.D. degree. Hence, the likelihood increases by three to four times. In summary, the figure demonstrates that both mother's and father's education are highly predictive on their own, but there also appears to be a strong interaction between the two. The largest gender differences in the marginal effects are among individuals with well-educated mothers and low-educated fathers.

Although Figure 2.6 does not reveal any split point by earnings, we plot the predicted likelihood of obtaining a master degree as a function of father's education and father's earnings in Figure 2.8 for completeness. The plot documents that higher levels of both earnings and education lead to higher

predicted likelihoods of finishing a master degree. However, the estimated effect of increasing father's earnings is considerably smaller than the effect of increasing father's education. Importantly, the gradient in earnings is much steeper for children with highly educated fathers: whereas the likelihood of obtaining a master degree for boys whose fathers have 21 years of education increases by 13 percentage points from the lowest to the highest earnings percentile, this increase is only two percentage points among boys whose fathers have nine years of education. This suggests that paternal earnings cannot compensate for low parental education.[5] There are only small gender differences in the marginal effects of father's earnings percentile and father's years of education.

To summarize our findings and visualize how the various parental characteristics interact in determining children's educational outcomes, we present alluvial diagrams in Figure 2.9. The figure shows the shares of boys and girls obtaining master degrees in the leftmost column. The second and third columns document the shares of master graduates with different levels of paternal and maternal education, respectively. The different colors of the second and third columns indicate the split points in Figure 2.6. The last column indicates the share of graduates coming from different paternal earnings quintiles. Hence, following the colored flows through the figure reveals the background composition of the children who do obtain master degrees. For instance, following the top flow from left to right in Figure 2.9 documents that about 10 percent of our sample obtains a master degree and about half of the men who obtained a master degree have fathers with more than 13 years of education and about two-thirds of these men with highly educated fathers also have mothers with more than 13 years of education. This indicates,

---

[5]The results are similar for maternal earnings and education and available on request.

on the one hand, that there is assortative mating among the parents and, on the other hand, that both parents' education is predictive of educational outcomes of children. A substantial share of men obtaining master degrees with highly educated parents do also have fathers in the highest income quintile. This implies that there is a strong correlation between paternal education and earnings and that father's earnings percentile is, in the absence of father's education, also a predictor of children's educational achievements. It is, however, worth noting that about one-third of the men with a master degree have fathers with earnings in the lowest two earnings quintiles and that the link between paternal earnings and boys' education is weaker than between paternal education and boys' education. These findings are consistent with the previous results showing that earnings play a significantly smaller role than education. The plot for females shows a similar picture. Overall, these findings imply that having a mother and a father who at least started a master degree is a good predictor for the likelihood that the child obtains a master degree. Hence, these results document status traps in terms of long university education.

2.5.2 Elite Education

Access to elite programs requires a very high GPA and these degrees have traditionally been pathways to a range of high-profile positions in the public administration and the private sector (see Table 2.1). We define master or higher degrees in medicine, law, economics, engineering at the Norwegian University of Science and Technology, and economics and business administration degrees at the Norwegian School of Economics as elite educations because they are generally perceived as high status and have a very high

GPA cutoff (Kirkeboen, Leuven, and Mogstad 2016). Figure 2.10 documents the results of the partitioning algorithm with elite master degrees or higher education as the outcome. As for master degrees, the first split point is for fathers having more than 13 years of education and the second split point for gender.[6] For boys and girls with fathers with more than 13 years of education, there are no nonlinearities in mothers' education. Interestingly, for females with highly educated fathers, we detect parameter instability at the 93rd percentile of the paternal earnings distribution. Because education programs at these elite institutions are not more expensive than degrees at other public institutions, the split point in earnings of fathers is unlikely to suggest that financial constraints in terms of tuition fees hinder children of less well-off families to undertake elite education. However, because having an elite education has a strong financial return (see Table 2.1), this split could be a proxy for fathers who themselves have such an elite education and, thereby, encourage their daughters to proceed in their footsteps. For both males and females with lower-educated fathers, there are additional split points in mothers' years of education which are at the same levels as for master degrees. Note that there are no split points in the region of birth or year of birth.

Figure 2.11 plots our model predictions for different levels of parental education. The educations of both fathers and mothers have a substantial impact on the predicted educational attainment of children and there is an added benefit from having two highly educated parents compared with one. There is also a very strong gradient in parental education: keeping the mother's education at 13 years, the likelihood that a boy obtains an elite degree changes from two percent when the father has nine years of education to 12 percent

---

[6]Table 2.3 summarizes the split points for different outcome variables.

when the father has a Ph.D. degree. Hence, the likelihood is six times as high when comparing male offspring of fathers with either very low or very high education levels. This gradient is also substantially larger than for master degrees where the likelihood for degree completion increased by three to four times. Moreover, the figure shows that men are more likely to obtain elite degrees. There are however, small gender differences in the marginal effects. Figure 2.12 plots our model predictions as a function of father's earnings and education. The overall picture is similar for men and women: the predicted effect of increasing father's earnings is modest at low education levels. However, at high paternal education levels, the effect of increasing fathers' earnings increases. As previously discussed, this might indicate that parents with prestigious, high-paying educations pass these educational choices on to their children. This interaction between parental earnings and education is also apparent in Figure 2.13, where there are strong flows from high-earning fathers through high education levels for both parents to the attainment of elite educations.

Whereas Chetty, J. N. Friedman, Saez, Turner, and Yagan (2019) show that more students at 12 extremely prestigious colleges in the US come from families in the top 1% of the parental income distribution than from the bottom half, the income differences are less extreme in Norway. About 47% of male elite program students have fathers in the top quintile of their earnings distribution and about 18% in the two lowest quintiles. Nevertheless, 43% have at least one parent with a master degree or higher education. These differences might indicate that the mechanisms behind these status traps are different across countries. On the other hand, part of the differences might also reflect that Norway has less inequality in earnings and that many highly educated parents work in the public sector (e.g., public hospitals or

public universities) and are paid salaries based on government pay scales. Hence, income differences among parents might be less apparent in Norway compared with the US and parental education is a more accurate measure for how encouraging Norwegian parents are toward educational attainment of their children.

Elite Education Controlling for Characteristics of the Extended Family

Previous literature has emphasized the potential role for transmission of human capital from family members beyond the parents, such as aunts, uncles and grandparents (Braun and Stuhler 2018; Lindahl, Mårten Palme, Massih, and Sjögren 2015; Long and Ferrie 2018; Clark 2014; Mare 2011; Adermon, Lindahl, and Marten Palme 2019; Jæger 2012). These dynasty effects are particularly of interest in terms of elite education. Although there are no benefits for legacy students in Norway, there might be an implicit value of having a well-educated relative who is able to guide a student to a suitable elite program. Hence, the scope for extended family to play a significant role is arguably large in the context of elite education because of their perceived status and high earnings potential. We, therefore, investigate the role of grandparents, aunts, and uncles in predicting an individual's likelihood of attending an elite program. Analyzing the role of the extended family is interesting for several reasons: first, comparing the estimated effect sizes of different types of relatives indicates their relative importance in determining children's educational outcomes. Second, our model allows us to analyze whether the characteristics of extended family members interact with parental characteristics. One such conceivable interaction could be that having a highly educated aunt as a role model compensates to some extent

for having parents with lower education levels. We, therefore, proceed by running the partitioning algorithm on Equation 2.2.

$$
\begin{aligned}
\text{Elite\_Child}_i = {} & \beta_0 + \beta_1 \text{Eduy\_Father}_i + \beta_2 \text{Eduy\_Mother}_i \\
& + \beta_3 \text{Earnperc\_Father}_i + \beta_4 \text{Earnperc\_Mother}_i \\
& + \beta_5 \text{Eduy\_Grandfather}_i + \beta_6 \text{Maxeduy\_Extended}_i \\
& + \xi \text{Yob\_Child}_i + \gamma \text{Labor\_Market}_i + \varepsilon_i,
\end{aligned}
\tag{2.2}
$$

where $Eduy\_Grandfather$ refers to the years of education attained by the child's paternal grandfather; and $Maxeduy\_Extended$ refers to the maximum years of education obtained among the aunts and uncles to which the child is related by blood, i.e., the parents' siblings. In addition to controlling for these characteristics of the extended family, we also test for parameter instability with respect to these variables to uncover potential interaction effects. Note that this analysis requires us to observe the education of the child's paternal grandfather and the child needs to have at least one aunt or uncle. Hence, the sample size is significantly reduced in this part of the analysis. Nevertheless, Table 2.2 shows that whereas the smaller sample is, on average, born slightly later, it is very similar in terms of other observables and, most importantly, has the same level of elite education attainment as the full sample. Hence, the analysis of the extended family should yield interesting conclusions.

Figure 2.14 shows the split points from running the partitioning algorithm on Equation 2.2. When controlling for the characteristics of the extended family, the first two split points are the same as without these controls (see Figure 2.10): fathers with 13 years of education and gender. As before, there is also a split point for paternal education above and below 17 years of education. There are, however, no longer parameter instabilities at 10 or 13 years of

maternal education or paternal earnings. A plausible explanation for finding fewer split points when including extended family characteristics is the lower sample size and therefore the loss in statistical precision necessary to satisfy the rather strict requirements we have imposed to identify additional split points.

Figures 2.15, 2.16, 2.17 and 2.18 illustrate how the model predictions depend on the various background characteristics when controlling for the characteristics of the extended family. In particular, Figure 2.15 documents that the years of education of both parents are important predictors of elite education. Figure 2.16 displays that very high paternal earnings are only crucial for men with highly educated fathers. Moreover, Figures 2.17 and 2.18 reveal that there is only a very small positive estimated effect of increasing the education of the extended family. Hence, parental education is a far better predictor of elite education than the education of the grandfather, the aunts, or the uncles.

2.5.3 Bachelor Degree

For offspring of less affluent families, obtaining a bachelor degree might be a path to higher status professions. Figure 2.19 shows the resulting splits from running the partitioning algorithm with bachelor degrees or higher education as the outcome. The figure shows that the first split point selected by the algorithm is gender. The importance of this split results from the fact that girls are more likely to obtain bachelor degrees independent of the social background. Among males, the following split points are primarily in the years of education of both parents. For mothers, there are split points for whether she has more than 10 years of education (more than middle school),

more than 13 years of education and more than 15 years of education (some bachelor-level education). In addition, there are split points for whether fathers have more than 12 years of education (completed high school) and more than 13 years of education (one year of college). One possible reason for having these two split points close together is that the group with some college education is a heterogeneous one, encompassing a wide range of education lengths and types for which different margins of parental education might be relevant. Moreover, there is a split point at the 13th percentile of the father's earnings. A possible interpretation of this split point is that this split might capture whether fathers participate in the labor market. This result indicates that there might be some credit constraint at the very bottom of the paternal earnings distribution. There is, however, no parameter instability at higher levels of paternal earnings percentiles suggesting that, for most men, there are no substantial credit constraints hindering them from attending college. Among females, there are split points for whether the mother has more than nine years of education (completed middle school), more than 10 years (some high school) and more than 13 years of education and whether the father has more than 13 years of education. Moreover, we find split points at the 60th and 72nd earnings percentiles among fathers. These earnings splits may proxy differences in education fields among the fathers and suggest that there are earnings thresholds above which fathers are more encouraging toward education for their daughters. Because the splits are at high levels of paternal earnings, these split points are less likely to indicate credit constraints. Similarly, we find a split point at the 25th earnings percentile among mothers. A possible interpretation of this split point is that it is either capturing whether mothers participate in the labor market or the number of hours they work. This might indicate some credit constraints for educational

investment for girls or suggest that mothers who are not participating in the labor market are not the encouraging role model needed for girls to invest in a college education. Overall, gender and mother's education above 10 years create the most important nonlinearities whereas there are no split points in the region of birth or year of birth. Even though there are large geographic disparities in Norway and our data span changes in access to scholarship and a rollout of university colleges (mostly offering bachelor degrees) to all counties (Carneiro, Liu, and Salvanes 2018), there are no parameter instabilities in regions or birth years. Hence, all these policies aiming at increasing access to college in rural areas do not significantly alter the role of parental education and earnings in predicting attainment of bachelor degrees. Overall, gender and mother's education above 10 years create the most important nonlinearities.

Note that most of the split points computed by the algorithm in Figure 2.19 were also present when analyzing the likelihood of obtaining master degrees (Figure 2.6). For boys, this applies to the split points 13 years of paternal and 10, 13, and 15 years of maternal education. Table 2.3 summarizes the split points for different outcome variables. Among girls, the split points at 13 years of paternal and at 10 and 13 years of maternal education are present in both Figures 2.6 and 2.19. Interestingly, the split point at 17 years of paternal education among girls obtaining a master degree is not present when studying for a bachelor degrees as an educational outcome. In addition, the hierarchy of split points is different between bachelor and master education. This indicates that different margins of parental background may be of varying importance depending on children's educational outcome of interest.

Figure 2.20 plots the estimated probability of obtaining a bachelor degree for

different levels of father's and mother's education separately for boys and girls. There are several noteworthy things in Figure 2.20. First, we see that increasing either parent's years of education leads to significant increases in the predicted probability of completing a bachelor degree. To illustrate this, we see that girls for whom both parents have nine years of education have a predicted probability of obtaining a bachelor degree of 24%. In contrast, girls who have fathers with nine years of education, but mothers with a Ph.D. have a predicted probability of 69%, a significant increase. Moreover, we note that a similar pattern is present among boys, indicating that the education of both mothers and fathers have large impacts on the likelihood of obtaining a bachelor degree irrespective of the child's gender. Although the overall patterns are the same for boys and girls, there are some gender differences. In particular, we see that girls are more likely to obtain bachelor degrees than boys irrespective of their parents' education. For example, boys whose parents both have 13 years of education have a likelihood of 48% of obtaining a bachelor degree; girls with the same educational background have a likelihood of 62%. The gradient in parental education is less strong when analyzing bachelor degrees as an outcome. Keeping the mother's education at 13 years, the likelihood that a boy obtains a bachelor degree changes from 34 percent when the father has none years of education to 75 percent when the father has a Ph.D. degree. Hence, the likelihood is about double when comparing male offspring of fathers with either very low or very high education levels. This gradient is smaller than for master degrees (three to four times) or elite degrees (six times). These differences might result from the large educational expansion campaigns that aimed to increase college attendance mostly at the bachelor level. Hence, more individuals—particularly women—from lower socioeconomic backgrounds

are attending college and becoming teachers, registered nurses, midwives, or office clerks. For obtaining master degrees, a high social background is, however, more crucial.

Figure 2.21 plots the predicted probability of obtaining a bachelor degree depending on the father's earnings percentile and father's years of education. These marginal effects allow us to compare the effect of paternal education and paternal earnings percentiles. The figure documents a significant and positive effect of increasing both father's years of education and father's earnings percentile. Importantly, moving toward higher paternal education leads to a more substantial increase in predicted educational attainment than moving toward higher paternal earnings. Hence, paternal education has a substantially larger estimated effect on the likelihood of obtaining bachelor degrees than paternal earnings. The pattern is similar for boys and girls. Moreover, the figure shows that for high levels of paternal education, the predicted likelihood of obtaining a bachelor degree is high also for the lowest paternal earnings levels. Hence, provided the father has a certain amount of human capital in the form of formal education, credit constraints do not appear to be an important factor. The ability to uncover such patterns is one of the clear advantages from our approach of using machine learning techniques to analyze how multiple background aspects interact in determining children's educational attainment.

To summarize how the various background characteristics interact in determining the attainment of bachelor degrees, we conclude this section with alluvial plots in Figure 2.22. Note that more than 30 percent of our sample obtains a bachelor degree. More than one-third of men who obtained a bachelor degree have fathers with more than 13 years of education and about half

of these men with highly educated fathers also have mothers with more than 13 years of education. Hence, marital sorting is very high among parents of graduates. Although most graduates have fathers who have above-median earnings, about one-third of the men with a bachelor degree have fathers with earnings in the lowest two earnings quintiles. Importantly, about two-thirds of the girls who obtain a bachelor degree have fathers who have 13 years or less of education and about half of the girls who obtain a bachelor degree have parents who both have 13 years or less of education. This is a striking difference to Figures 2.9, and 2.13 where a substantially smaller fraction of elite or master graduates have parents who have 13 years or less of education.

2.5.4 Cognitive Ability

Previous literature documents a significant transmission of cognitive ability from parents to children (Black, Devereux, and Salvanes 2009) and shows that cognitive ability is correlated with college attendance rates (Belley and Lochner 2007). These two facts imply that the results discussed above may be driven, at least partly, by the transmission of cognitive ability from parents to children. That is, if high parental education is a function of high cognitive ability and cognitive abilities are transmitted from parents to children, the high intergenerational persistence in education might simply reflect high intergenerational persistence in cognitive ability. We investigate this by including cognitive ability measures from the military enlistment for sons in

the analysis and use the partitioning algorithm to estimate Equation 2.3:

$$\text{Education\_Child}_i = \beta_0 + \beta_1 \text{Ability\_Child}_i + \beta_2 \text{Eduy\_Father}_i + \beta_3 \text{Eduy\_Mother}_i$$
$$+ \beta_4 \text{Earnperc\_Father}_i + \beta_5 \text{Earnperc\_Mother}_i$$
$$+ \xi \text{Yob\_Child}_i + \gamma \text{Labor\_Market}_i + \varepsilon_i,$$

$$(2.3)$$

where $Ability\_Child_i$ indicates the cognitive abilities of child i. Including the son's cognitive ability in the empirical model and testing for parameter instability with respect to it enables us to analyze the estimated effect of the son's cognitive ability and how it interacts with the other background characteristics in the model. These estimates give an indication of how much of our previous results are driven by intergenerational transmission of cognitive ability between generations. Note that we do not have cognitive ability measures for women, and we are, therefore, not able to test for parameter instability with respect to gender in these models.

How much of this strong association between parental education and children's educational attainment is related to cognitive ability? Figure 2.23 shows the results from the partitioning algorithm on Equation 2.3 for men obtaining master degrees. The first split is for having a cognitive ability score above five, which is the mean value of the test scores. Hence, above-average cognitive ability is an important determinant of completing a master degree. For both men with below- and above-average cognitive ability, the subsequent split points are for fathers having more than 13 years of education. There is an additional split at the ability level of seven for children whose fathers have 13 years of education or less. When controlling for cognitive ability, there is no parameter instability in mother's education nor parental

85

earnings, year of birth, or region of birth.

Given the high admission requirements for the elite educations in terms of high school GPA, it is likely that cognitive ability is an important determinant for who obtains such an education. As with master degrees, the first split point is again in cognitive ability when studying elite educations (Figure 2.24). Nevertheless, the parameter instability is at a higher level of cognitive ability (level six instead of five). This finding is reasonable given the high GPA thresholds for admission to elite educations. For men with the cognitive ability of six and below, we detect further instability at 13 and 17 years of paternal education.

As for master degrees, the first two splits for men with bachelor education (Figure 2.25) divide men into individuals with a cognitive ability score below and above the mean (five) and individuals whose fathers have more than 13 years of education. The additional split points are cognitive ability above three among sons with well-educated fathers and for mother's education above 10 and above 13 years among sons with less-educated fathers.

At all three educational margins, the largest nonlinearities are in cognitive ability levels. This suggests that having a cognitive ability level above the mean is an important indicator for whether an individual completes a college degree. Nevertheless, there are also interaction effects and nonlinearities in father's education for all three educational margins—particularly among the individuals with lower cognitive ability. We only find split points in mother's education when analyzing bachelor degrees as outcomes.

To visualize how the estimated effect of the boys' cognitive ability compares with the effect of background characteristics, we plot how the predictions

from the model change when varying the cognitive ability score and father's years of education in Figures 2.26, 2.28, and 2.27. Both son's cognitive ability and father's years of education have significant positive estimated effects on the likelihood of obtaining a master degree, an elite degree, or a bachelor degree. For example, among men whose fathers have 13 years of education, those with a cognitive ability score of two have an estimated probability of obtaining a bachelor degree of 19% and those with a score of nine have a probability of 72%. Despite this sizable effect of cognitive ability, the figures also reveal a steep socioeconomic gradient even when conditioning on cognitive ability. For master and elite degrees, the nonlinearities and interaction effects are substantial. In particular, there is a steep gradient in the predicted likelihood of obtaining a bachelor degree by father's education. For example, among children with an average cognitive ability score (level five), 6% of men whose fathers have nine years of education obtain a master degree, whereas 28% of men whose fathers have 21 years of education get such a degree. Hence, these results indicate that parental education can partly compensate for low ability. Moreover, the likelihood of obtaining a master degree or an elite degree increases more by father's years of education for men with a cognitive ability of five or more. These disparities clearly demonstrate that children's cognitive ability is important, but not the only reason for the high intergenerational persistence in education.

The alluvial plots in Figures 2.29, 2.30, and 2.31 summarize all these findings. First, cognitive abilities and parental education are important predictors for a master, elite, or bachelor education and there are strong correlations among all three variables. Second, although there are many men who obtain degrees with highly educated parents and high cognitive ability, there is a substantial share of graduates that have highly educated parents but below

mean cognitive ability. Hence, having highly educated parents compensates for low cognitive ability in the process of obtaining a degree. Overall, it is important to note that we do not find parameter instability by parental earnings percentiles when controlling for cognitive ability and that this suggests that credit constraints are not a main driver for the intergenerational persistence in education when a child's cognitive ability and parental education are controlled for.

2.5.5 Intergenerational Mobility in Earnings

Above, we document that parental education is a very important predictor of different educational outcomes of children. Parental earnings, however, are a less important predictor once we account for parental education. Parental earnings and parental education might not necessarily reflect the same type of social status. Whereas a philosophy professor might be a man of some account through knowledge rather than earnings, some entrepreneurs earn a fortune without long formal educations. Hence, both parental education and earnings are likely imperfect proxies of social status. The question is whether the parental earnings percentiles are better predictors for children's earnings percentile than they are for children's education? Figure 2.32 shows the split points when analyzing a child's earnings percentile as an outcome. The first split is for the child's gender. Hence, there are different margins of parental characteristics that are important for boys and girls. For both genders, there are many points of parameter instability in father's and mother's earnings percentiles. For boys, the split points are both in the highest and the lowest earnings quintiles. Hence, very high paternal earnings provide an extra increase in the son's earnings percentile, whereas very low paternal and

maternal earnings are a hindrance for obtaining high earnings. Although there are no split points in father's education for boys, there are split points in mother's education for both genders. Because many mothers might not participate in the labor market or only to a small degree, their education might be a better proxy for their earnings potential. Moreover, girls whose fathers have more than 15 years of education (more than a bachelor degree) achieve higher earnings as adults. Overall, these results suggest that parental earnings percentiles are good predictors for children's earnings percentile and that well-earning parents transmit traits that enable children to achieve high earnings too, whereas children of poor parents are more likely to remain poor.

When focusing on boys and controlling for cognitive ability (see Figure 2.33), the first split is for above- and below-average ability. The remaining split points are similar to those in Figure 2.32: at high and low values of the father's earnings percentiles and at low values of the mother's earnings percentiles. In addition, there are splits at 12 years of mother's and 14 years of father's education. Hence, cognitive ability as well as parental earnings are good predictors for children's high earnings percentiles.

Overall, using our highly flexible specification, allowing for nonlinearities and interactions between parental education, parental earnings, the region of birth, year of birth and gender, most significant nonlinearities and inter-actions are for paternal earnings. Interestingly, there are no split points in birth years or regions. Hence, although parental earnings percentiles are relatively poor predictors for children's educational attainment, they are good predictor of children's earnings percentiles. This indicates that there is a strong intergenerational persistence in social status and, particularly, in

the dimension parents do best—highly educated parents are likely to have highly educated children and parents with high earnings percentiles are likely to have children that are well off financially.

2.6 Conclusion

Although the social welfare state in Nordic countries provides an environment where education is freely available and a generous scholarship system exists, parental education still plays an important role in explaining educational attainment, particularly for, university education. In this paper, we use a model-based recursive partitioning approach developed by Zeileis, Hothorn, and Hornik (2008) to provide empirical evidence at which margins along the educational distribution this strong persistence arises and whether there are specific dimensions of the family background or the interplay of various dimensions that are the most important drivers for this intergenerational persistence in education.

We document that obtaining a bachelor, a master, or an elite degree is not a continuous function of parental years of education. Our results show discontinuities at different parental education levels. In particular, more than 13 years of paternal education, which reflects education beyond high school, and more than 10 (more than compulsory education), 13, and 15 years (some master degree education) of maternal earnings are important margins at which there are clear jumps in the likelihood that a child obtains a degree. In addition, for girls' educational attainment and, for achieving elite degrees, fathers with more than 17 years of education (education beyond a master degree) are crucial. This suggests that there are social status traps at the very top of the educational distribution of parents that are only present for a long

and prestigious college education. Whereas having a father with more than 13 years of education creates the largest nonlinearity in the likelihood that an individual obtains a master or an elite degree, an individual's gender and having a mother with more than 10 years of education are more important for graduating with at least a bachelor degree. Although the levels at which we detect the nonlinearities and interactions are similar across the different educational outcomes, the hierarchy of their importance is different between individuals who get a short and a long college education. Moreover, our results suggest that borrowing constraints are not the main explanation for educational status traps. Parental earnings are a much less important predictor of children's education once mother's and father's education are controlled for. The only split point that might suggest that there are some borrowing constraints is that boys are significantly less likely to obtain a bachelor degree if their fathers' earnings are below the 14th percentile of the earnings distribution. For master or elite degrees or for girls, we do not find any split points at the lower end of the paternal earnings distribution. For elite degrees, we find a split point for girls at the top of the father's earnings distribution suggesting that girls from the top six percentiles do about equally well as girls whose fathers have 18 years of education or more. Furthermore, our findings suggest that high paternal earnings cannot compensate for low parental education. In addition, we present evidence that cognitive ability is a good and nonlinear predictor of educational attainment but not the only driver of the educational status traps. In particular, we show that having a cognitive ability level above the mean creates the strongest nonlinearities and that there are still parameter instabilities in father's education. Moreover, our findings suggest that a father's education can compensate for his son's lower cognitive ability. That is, the likelihood for obtaining a master degree is the

91

same for men with the highest level of cognitive ability and for low-educated fathers as it is for men with average cognitive ability and fathers with a Ph.D. There is also an interest in the situation in which we do not find any parameter instability: there are no split points in grandfathers', aunts', and uncles' education as well as interactions of all these dimensions. Although there is regional variation in educational attainment, there are no instabilities with respect to geographic location. In addition, the algorithm does not detect any splits by birth cohort. This is rather surprising because there was an expansion in the number of colleges and in access to college loans in Norway during the years of our analysis. Hence, many of the measures the social welfare state implements to equalize opportunities do not lead to significant instabilities.

Overall, the first contribution of this paper is that we capture more complex relationships between family background characteristics and are not bound to analyze a single aspect such as parental earnings or education. The second contribution is that we do not need to assume linearity as in most of the previous literature (see, e.g., Chetty, Hendren, Kline, and Saez 2014; Adermon, Lindahl, and Marten Palme 2019). It is important to note that the main limitation of our approach is that we can only interpret our results as suggestions for causal mechanisms. Although our results suggest that it is difficult to design public programs to lower status traps in the form of master and elite education by just increasing access to additional funds in the form of study loans or higher parental earnings, future research is needed to properly determine causal effects.

2.7 Tables and Figures

Table 2.1: Most Common Occupations by Education Level

Education level	Occupation	Mean earnings percentile
< **Bachelor's**	Personal service activities	0.305
	Sales	0.366
	Office workers	0.434
Bachelor's	Workers in business, administration and sales	0.398
	Occupations in culture and sports	0.593
	Health related occupations	0.393
Master's	Civil engineers	0.797
	Administrative and business leaders	0.792
	IT advisors	0.610
Elite	Medical occupations	0.855
	IT advisors	0.801
	Civil engineers	0.832

Note: The table shows the most common occupations in the full estimation sample in the years 2003–2015, i.e., men and women born in the years 1955–1980. We classify occupations using the STYRK08 standard and we aggregate occupations to the first two digits of the STYRK08 codes. The earnings percentiles are calculated at the national level by cohort and then averaged by occupation.

Table 2.2: Summary Statistics

Sample	Variable	N	Mean	SD
Full	Child has Bachelor's	952167	0.34	0.47
	Child has Elite Education	952167	0.03	0.17
	Child has Master's	952167	0.10	0.30
	Father's Earnings Percentile	952167	50.50	28.87
	Father's Years of Education	952167	11.59	2.92
	Male	952167	0.52	0.50
	Mother's Earnings Percentile	952167	50.50	28.87
	Mother's Years of Education	952167	11.01	2.46
	Year of Birth	952167	1968.48	6.96
Cognitive Ability	Child has Bachelor's	381805	0.31	0.46
	Child has Elite Education	381805	0.04	0.18
	Child has Master's	381805	0.10	0.31
	Child's Cognitive Ability	381805	5.10	1.81
	Father's Earnings Percentile	381805	50.86	28.79
	Father's Years of Education	381805	11.79	2.92
	Male	381805	1.00	0.00
	Mother's Earnings Percentile	381805	50.76	28.77
	Mother's Years of Education	381805	11.24	2.47
	Year of Birth	381805	1970.54	5.46
Extended	Child has Bachelor's	377151	0.37	0.48
	Child has Elite Education	377151	0.03	0.17
	Child has Master's	377151	0.09	0.29
	Father's Earnings Percentile	377151	50.00	28.73
	Father's Years of Education	377151	11.79	2.76
	Male	377151	0.52	0.50
	Max Years of Education among Aunts/Uncles	377151	13.58	3.00
	Mother's Earnings Percentile	377151	52.45	28.59
	Mother's Years of Education	377151	11.45	2.44
	Paternal Grandfather's Years of Education	377151	8.25	2.17
	Year of Birth	377151	1972.44	5.80

Table 2.3: Split Points for Different Outcomes by Gender

Education Level	Males					Females			
	Father's Years of Education	Mother's Years of Education	Father's Earnings Percentile	Mother's Earnings Percentile	Child's Cognitive Ability	Father's Years of Education	Mother's Years of Education	Father's Earnings Percentile	Mother's Earnings Percentile
Bachelor's	12 13	10 13 15	13			13	9 10 13	60 72	25
Bachelor's with IQ	13	10 13			3 5				
Master's	13	10 13 15				13 17	10 13		
Master's with IQ	13				5 7				
Elite	10 13 17 18	10				13 17	10	93	
Elite with IQ	13 17				6				

Figure 2.1: Shares of Individuals with Different Degrees by Father's Years of Education
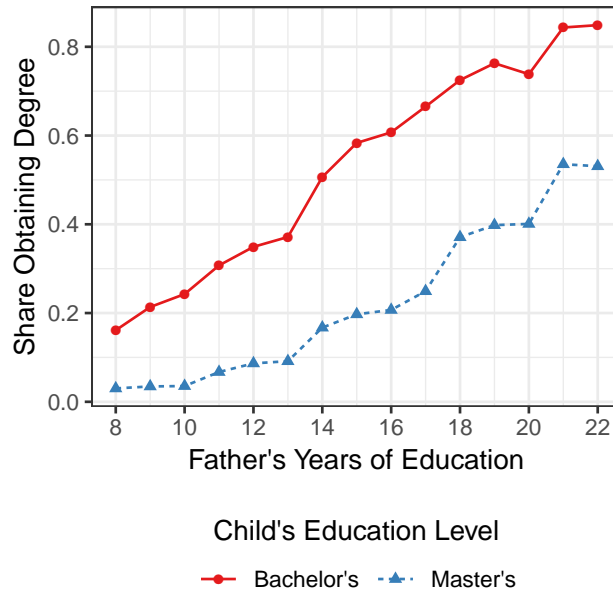
Note: The figure plots the share of individuals among birth cohorts 1950 to 1980 who have bachelor and master degrees by father's years of education.

Figure 2.2: Share of Individuals with Different Educational Degrees by Birth Cohort
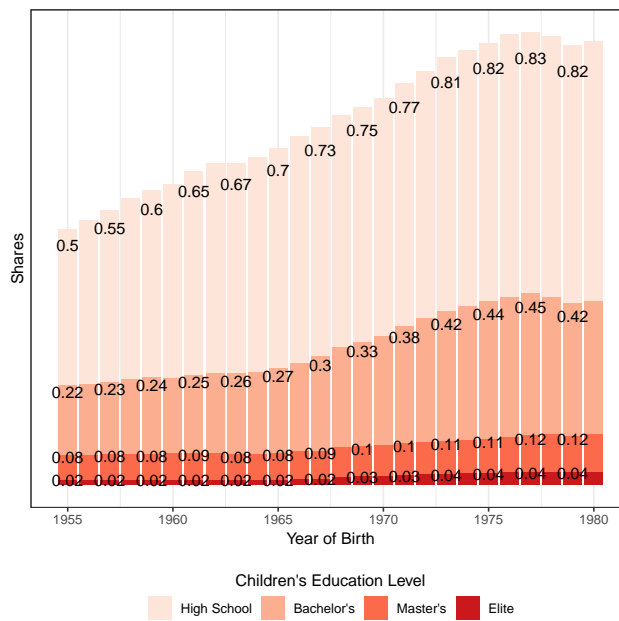
Note: The figure plots the share of individuals among each birth cohort 1955–1980 who have a high school diploma, a bachelor, or a master degree.

Figure 2.3: Share of Individuals with Different Educational Degrees by Birth Cohort and Father's Education Level
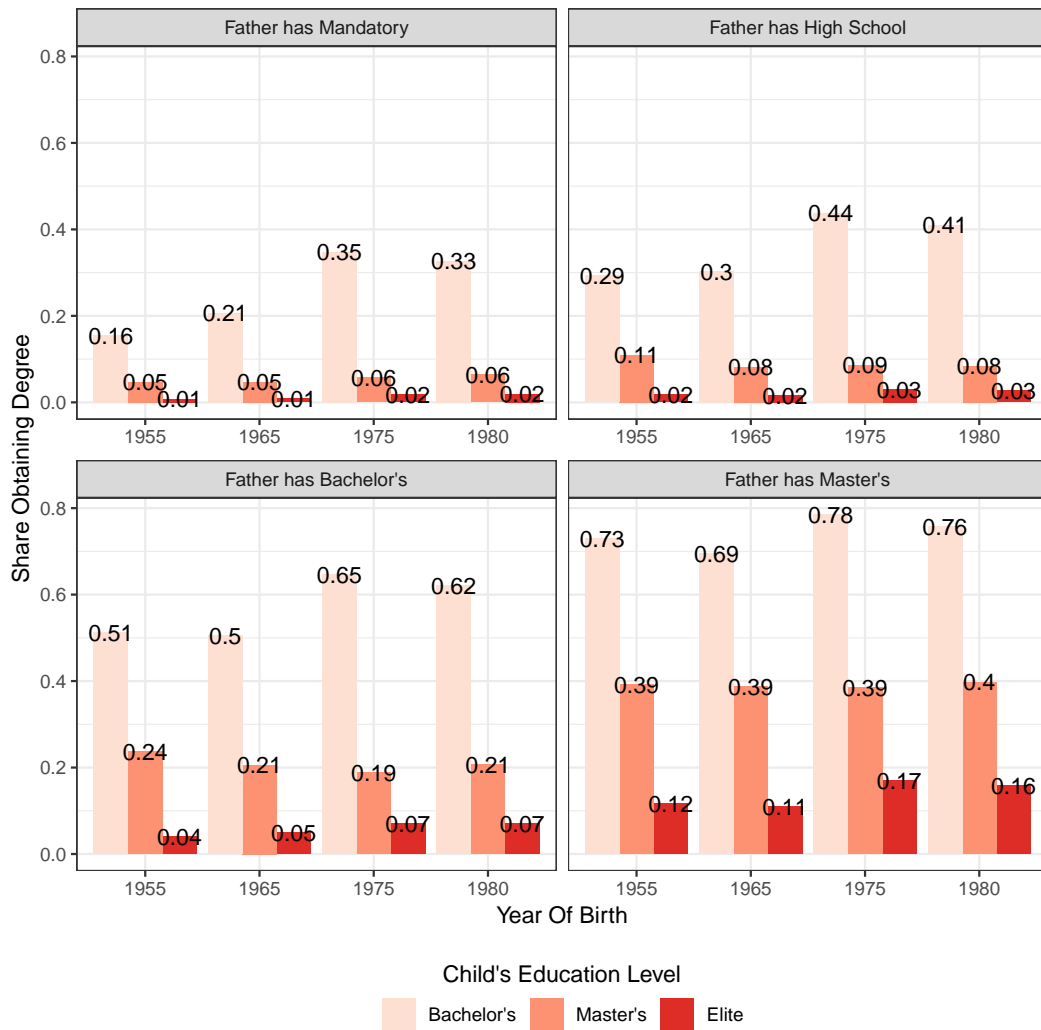
Note: The figure plots the share of individuals in four different birth cohorts (1955, 1965, 1975, and 1980) who obtain a bachelor, a master or an elite education conditional on father's education level.

Figure 2.4: Earnings Percentile Density by Educational Degree
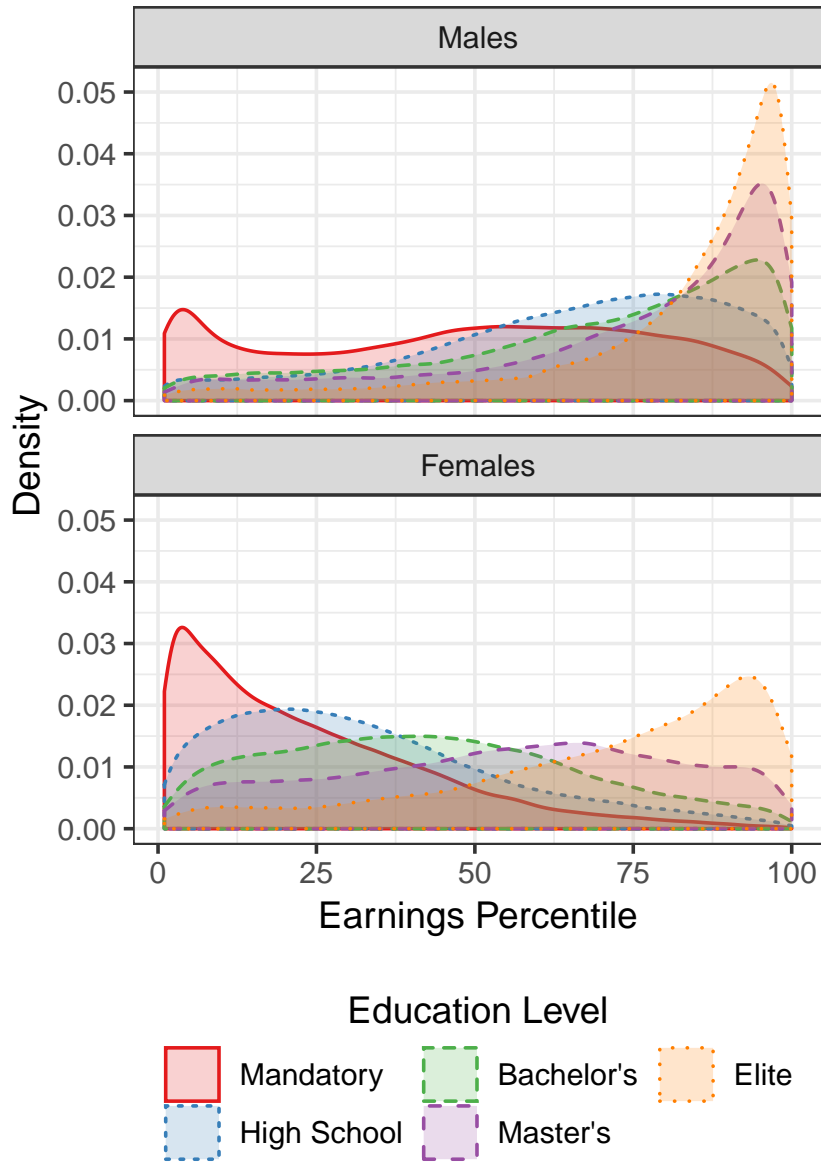
Note: The figure plots the distribution of earnings percentiles in the national income distribution conditional on education level and gender for our sample of individuals born 1955–1980. The earnings percentiles are calculated within birth years.

Figure 2.5: Cross-Validation Results
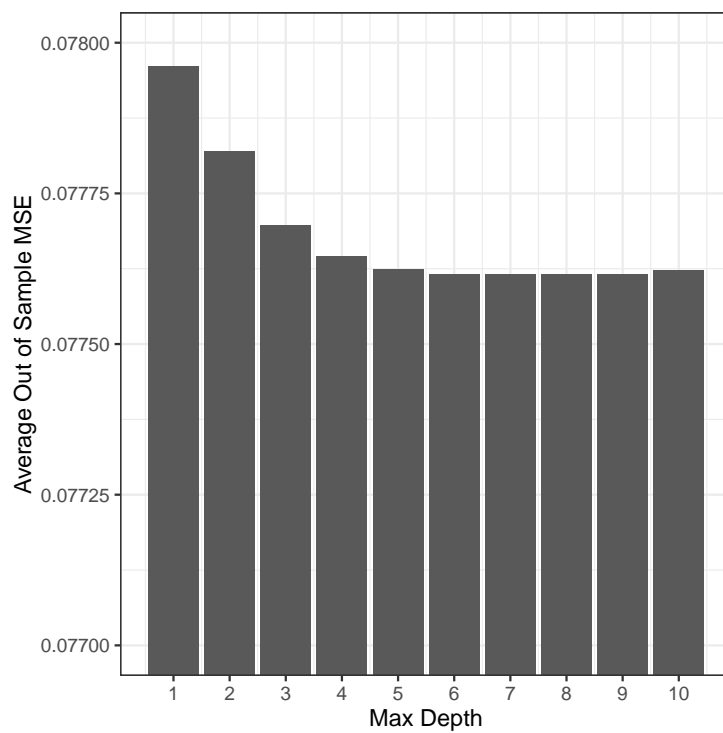
Note: The figure shows hold out mean squared error from applying the model-based recursive partitioning algorithm with various maximum tree depths. The results were obtained using 10-fold cross-validation cross-validation of predictions from Equation 2.1 with obtaining a master degree as the outcome.

Figure 2.6: Generalized Linear Model Tree for Master Degrees
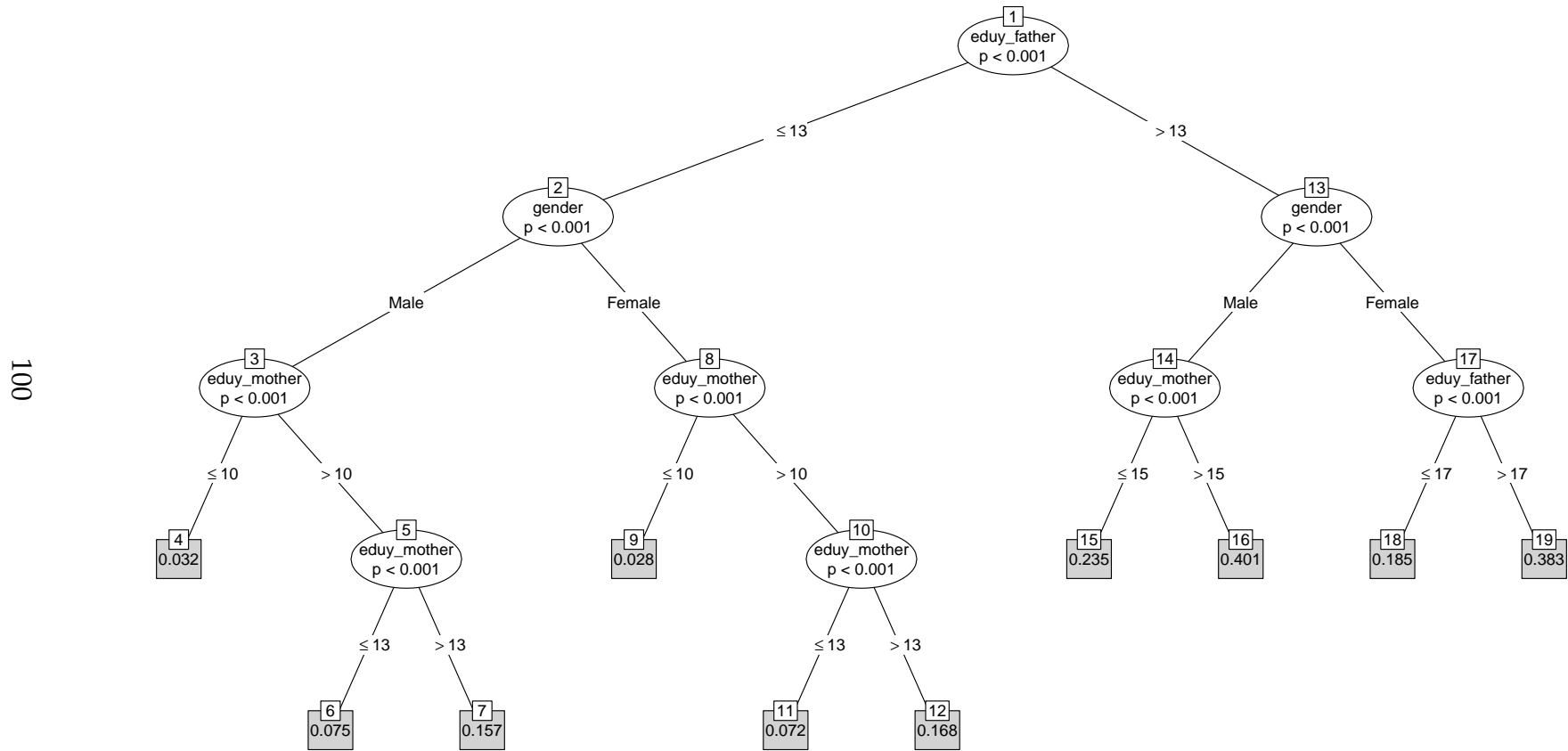
Note: The figure shows the split points and groups resulting from running the partitioning algorithm on Equation 2.1 with master degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining master degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.7: Marginal Effect of Father's and Mother's Years of Education, Master Degrees
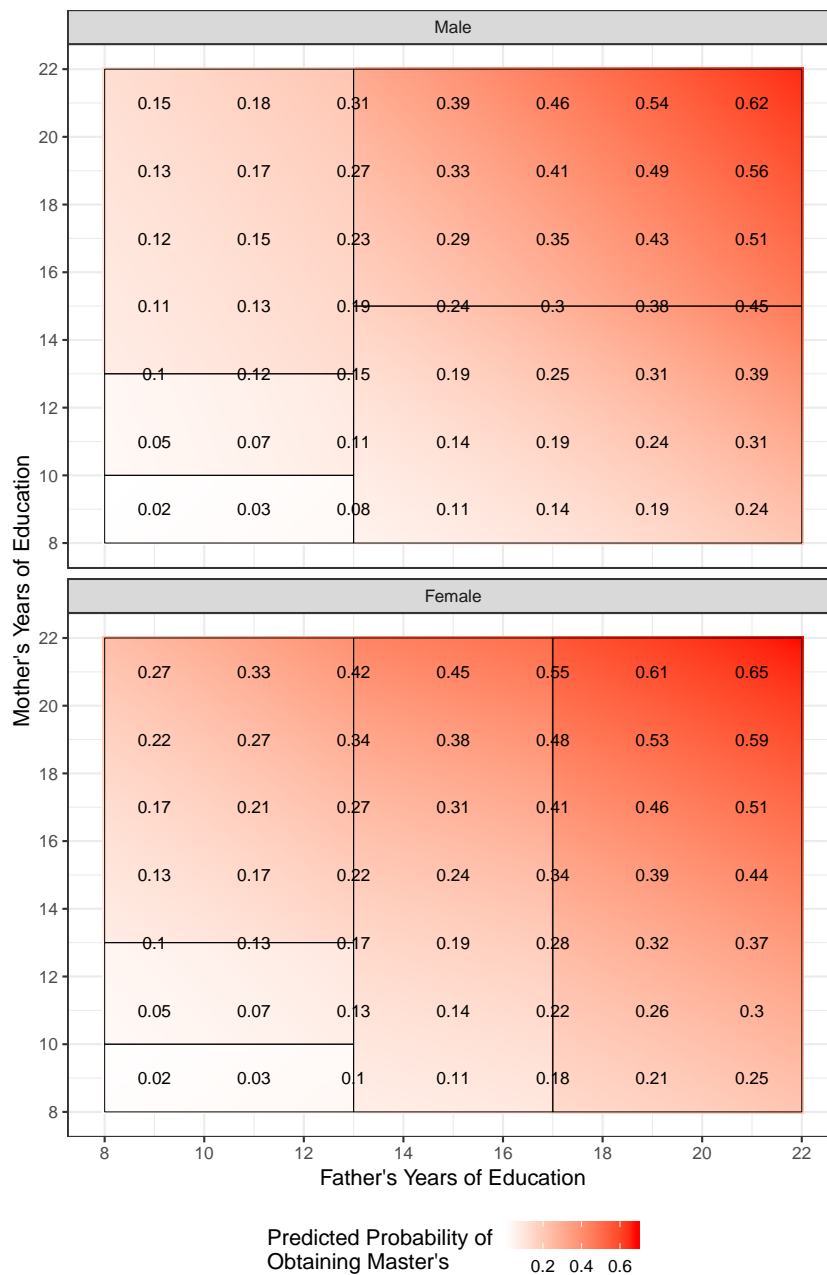
Note: The figure plots the predicted probability of obtaining a master degree depending on father's and mother's years of education. All other background characteristics (father's and mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.8: Marginal Effect of Father's Earnings Percentiles and Father's Years of Education, Master Degrees
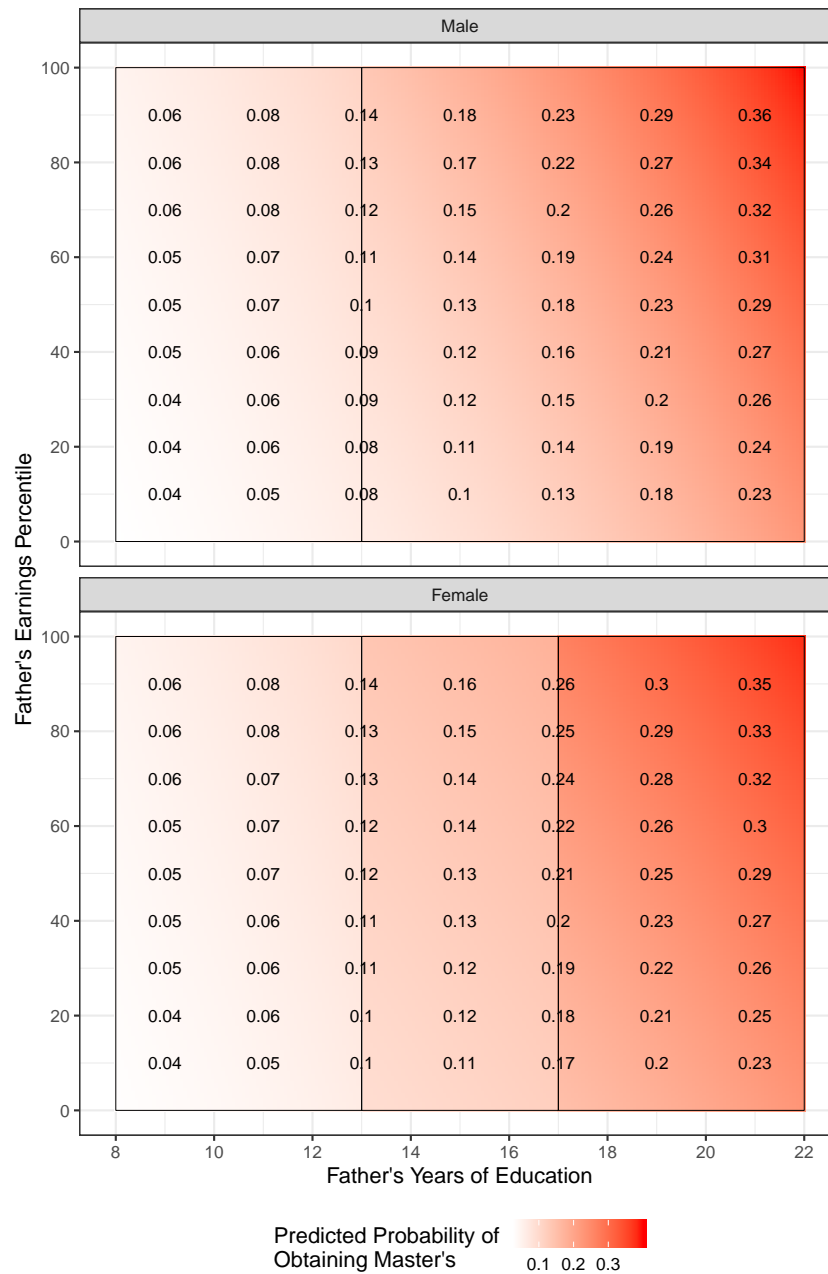
Note: The figure plots the predicted probability of obtaining a master degree depending on father's earnings percentile and father's years of education. All other background characteristics (mother's years of education, mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.9: Alluvial Diagram Master Degrees
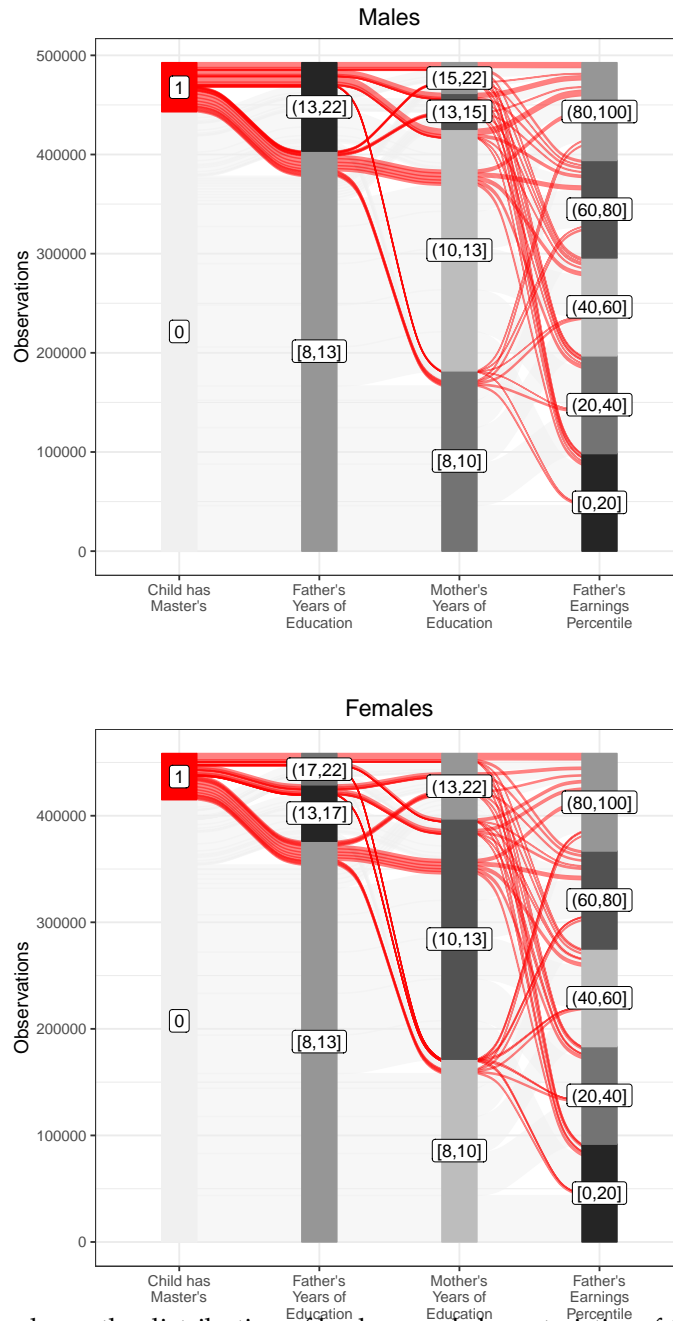
Note: The figure shows the distribution of background characteristics of those who do and do not obtain master degrees in terms of parental education and paternal earnings. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

Figure 2.10: Generalized Linear Model Tree for Elite Degrees
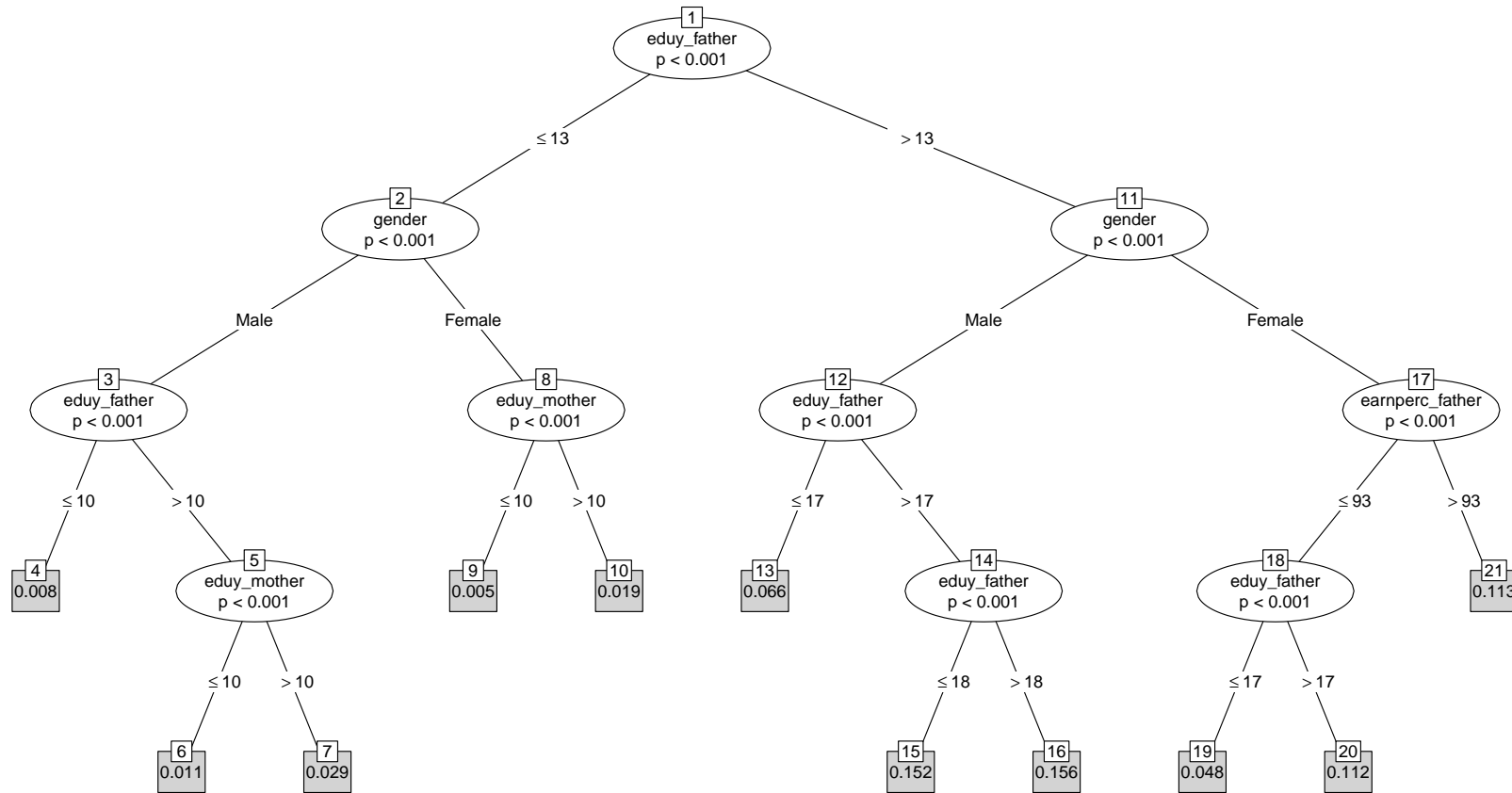
Note: The figure shows the split points and groups resulting from running the partitioning algorithm on Equation 2.1 with elite degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining master degrees. The nodes higher up in the tree show the split points selected by the algorithm.

**Figure 2.11:** Marginal Effect of Father's and Mother's Years of Education, Elite Degrees
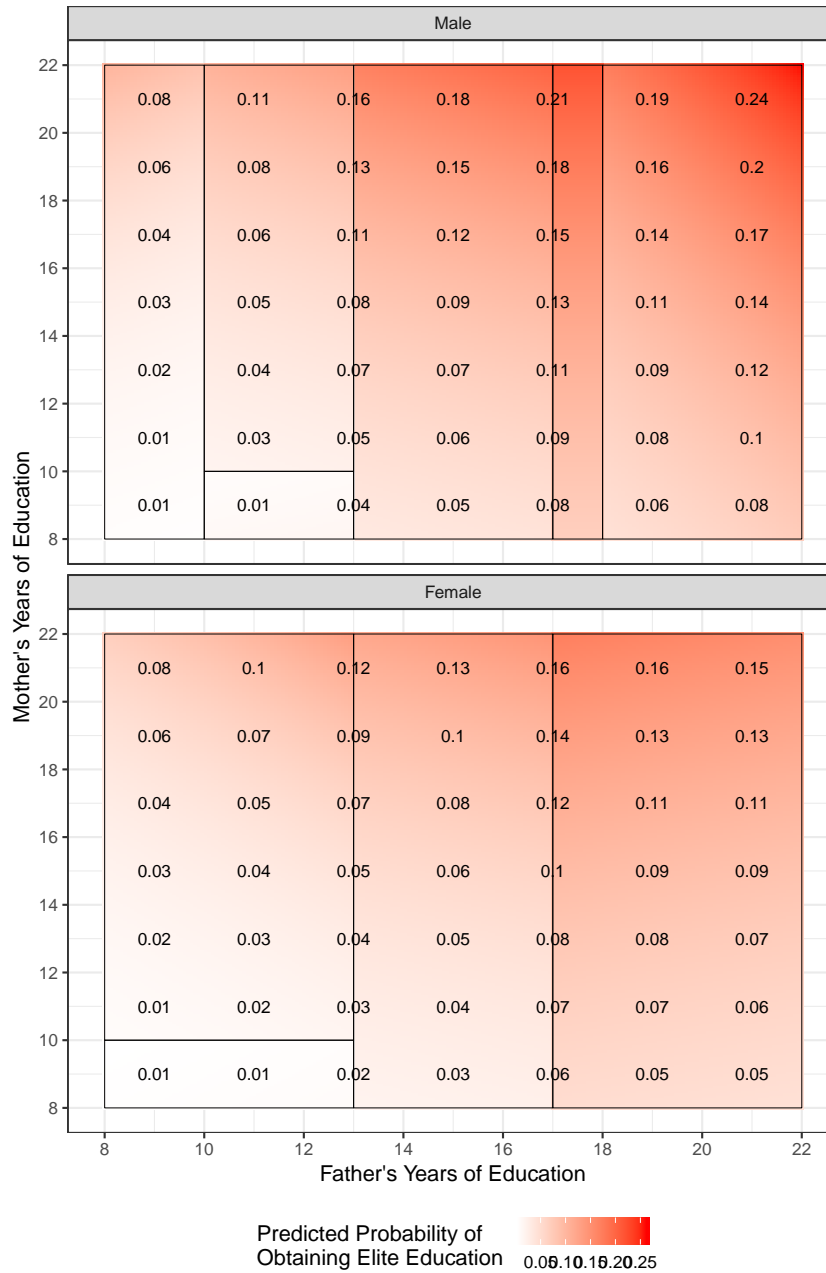
Note: The figure plots the predicted probability of obtaining an elite degree depending on father's and mother's years of education. All other background characteristics (father's and mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values

Figure 2.12: Marginal Effect of Father's Earnings Percentiles and Father's Years of Education, Elite Degrees
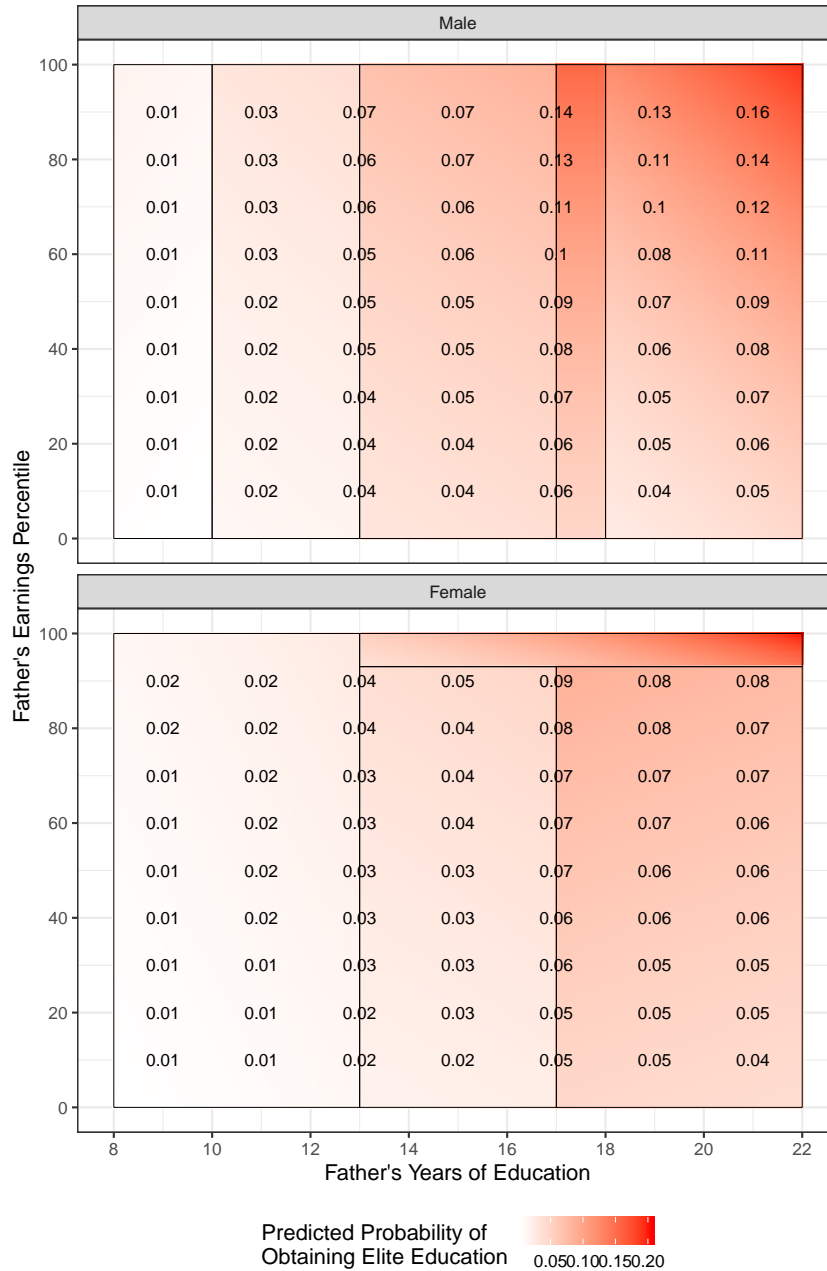
Note: The figure plots the predicted probability of obtaining an elite degree depending on father's earnings percentiles and father's years of education. All other background characteristics (mother's years of education, mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.13: Alluvial Diagram Elite Degrees
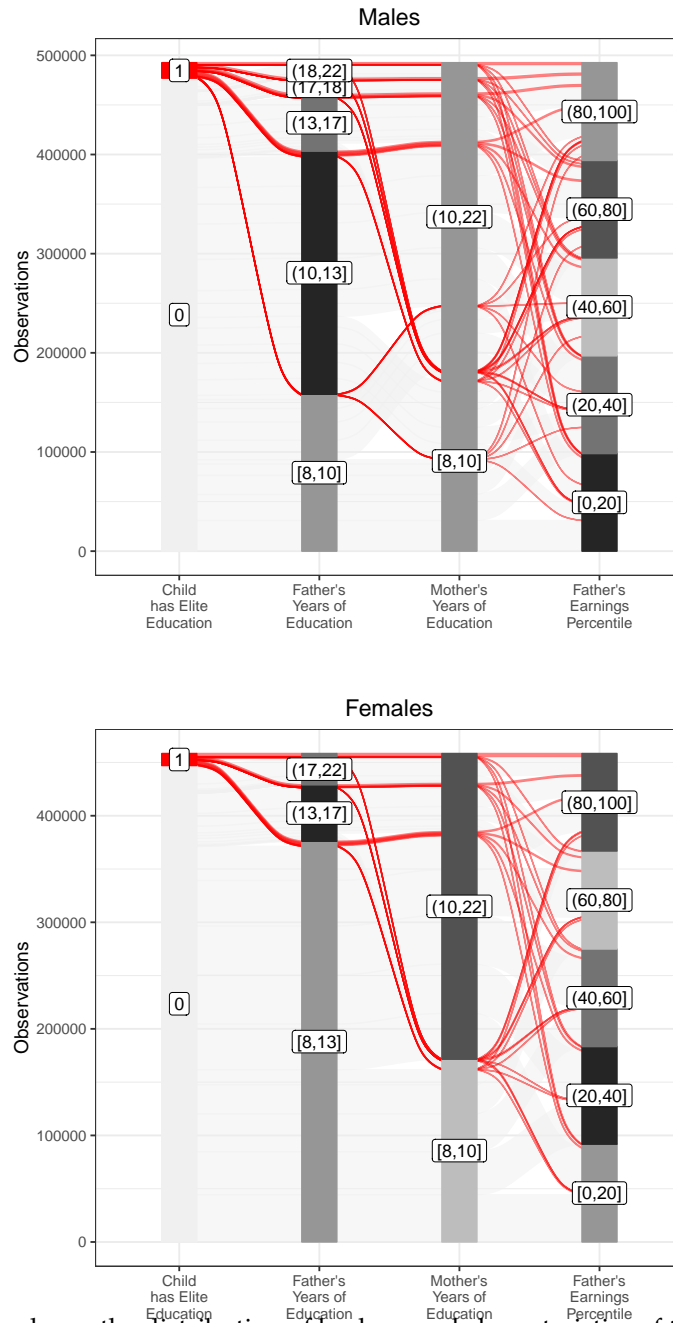
Note: The figure shows the distribution of background characteristics of those who do and do not obtain elite degrees in terms of parental education and paternal earnings. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

Figure 2.14: Generalized Linear Model Tree for Elite Degrees Controlling for Characteristics of the Extended Family
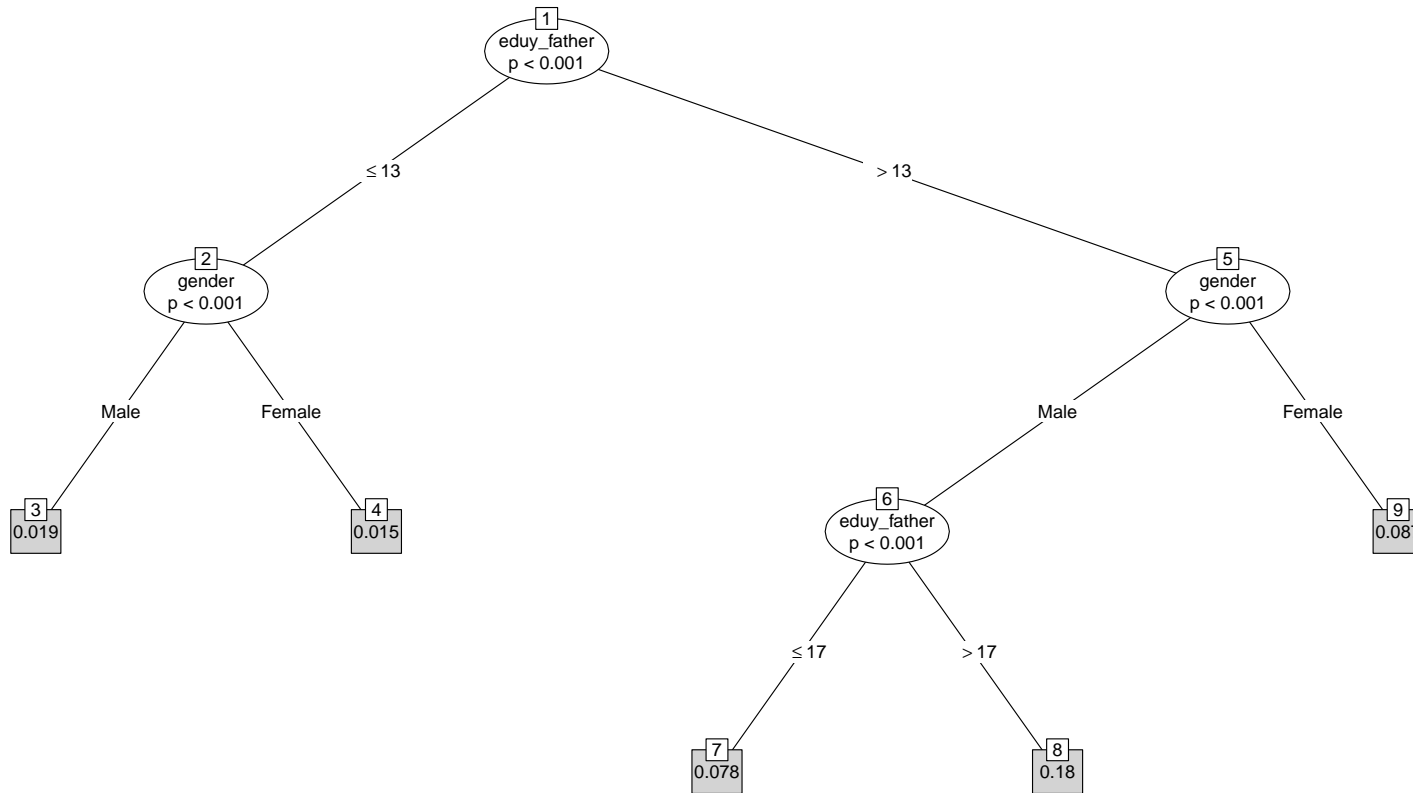
Note: The figure shows the split points and groups resulting from running the partitioning algorithm on Equation 2.2, i.e., while including extended family in the analysis, with elite degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining elite degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.15: Marginal Effects of Father's and Mother's Years of Education, Elite Degrees with Extended Family Controls
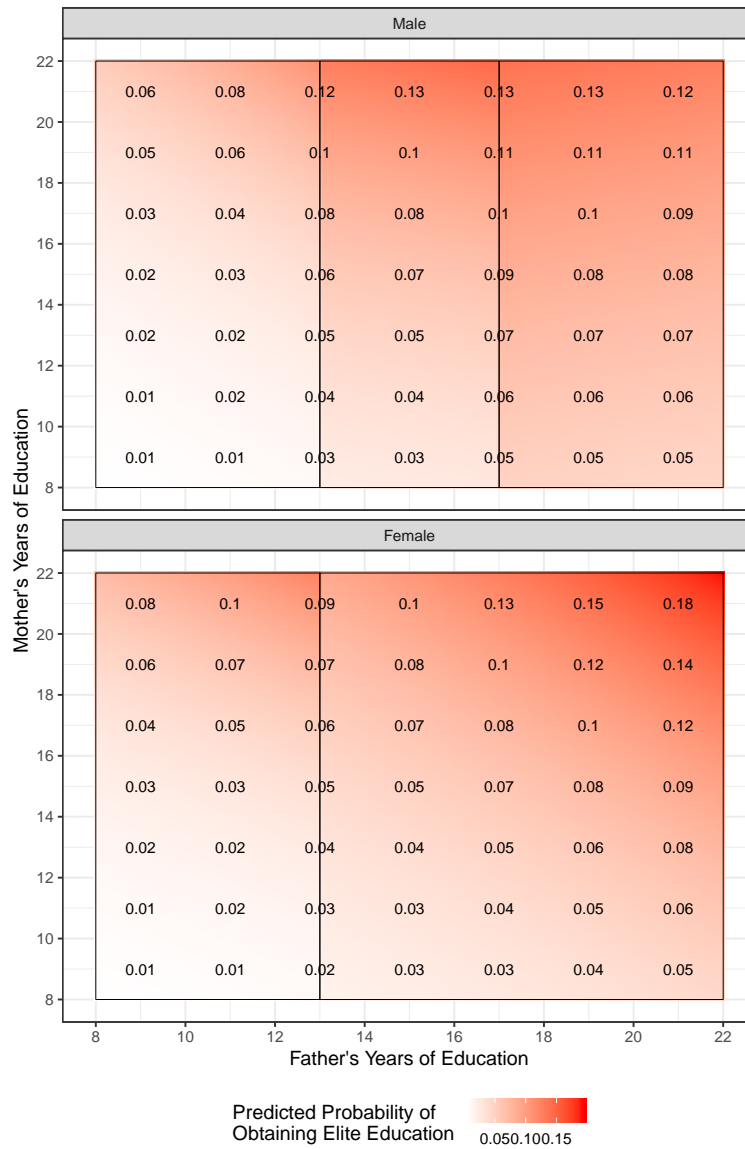
Note: The figure plots the predicted probability of obtaining an elite degree depending on father's and mother's years of education. All other background characteristics (father's and mother's earnings percentile, grandfather's, aunts' and uncles' years of education, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.16: Marginal Effects of Father's Earnings Percentile and Father's Years of Education, Elite Degrees with Extended Family Controls

Note: The figure plots the predicted probability of obtaining an elite degree depending on father's earnings percentile and father's years of education. All other background characteristics (mother's years of education, mother's earnings percentile, grandfather's, aunts' and uncles' years of education, year of birth, and labor market of birth) are held fixed at median values.

110

Figure 2.17: Marginal Effects of the Maximum Years of Education among Aunts' and Uncles' and Father's Years of Education, Elite Degrees with Extended Family Controls
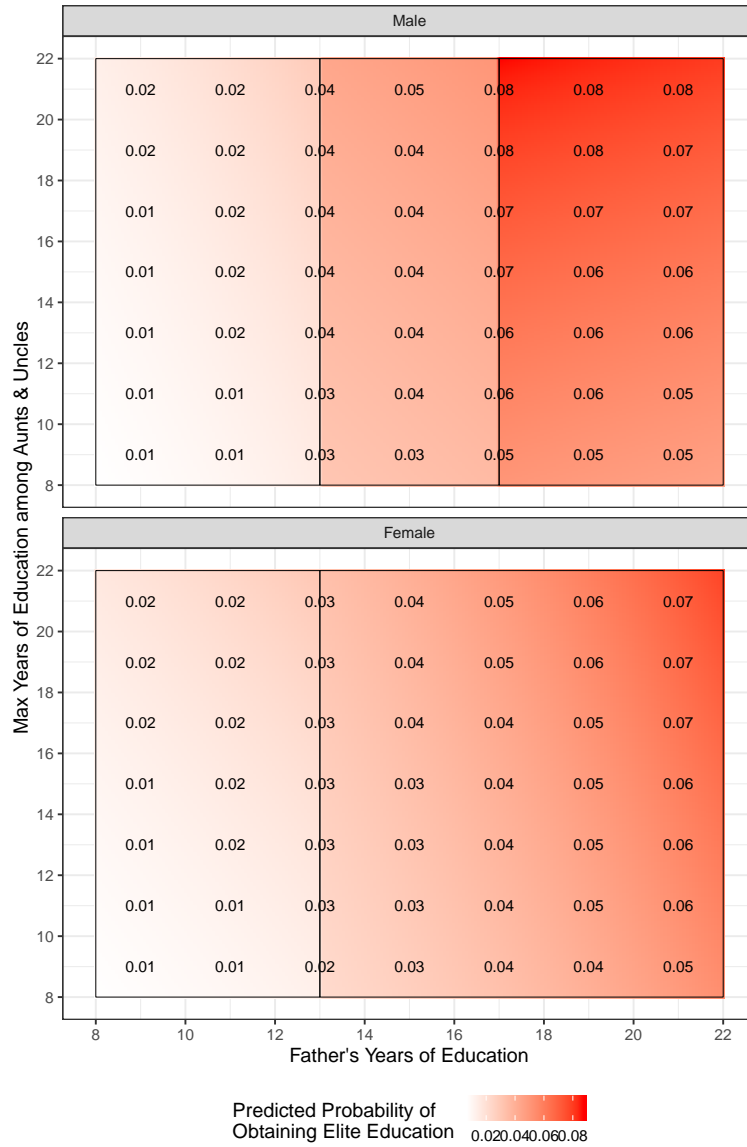
Note: The figure plots the predicted probability of obtaining an elite degree depending on the maximum years of education among uncles and aunts and father's years of education. All other background characteristics (mother's years of education, father's and mother's earnings percentiles, grandfathers' years of education, year of birth, and labor market of birth) are held fixed at median values.

**Figure 2.18:** Marginal Effects of Grandfathers' and Father's Years of Education, Elite Degrees with Extended Family Controls
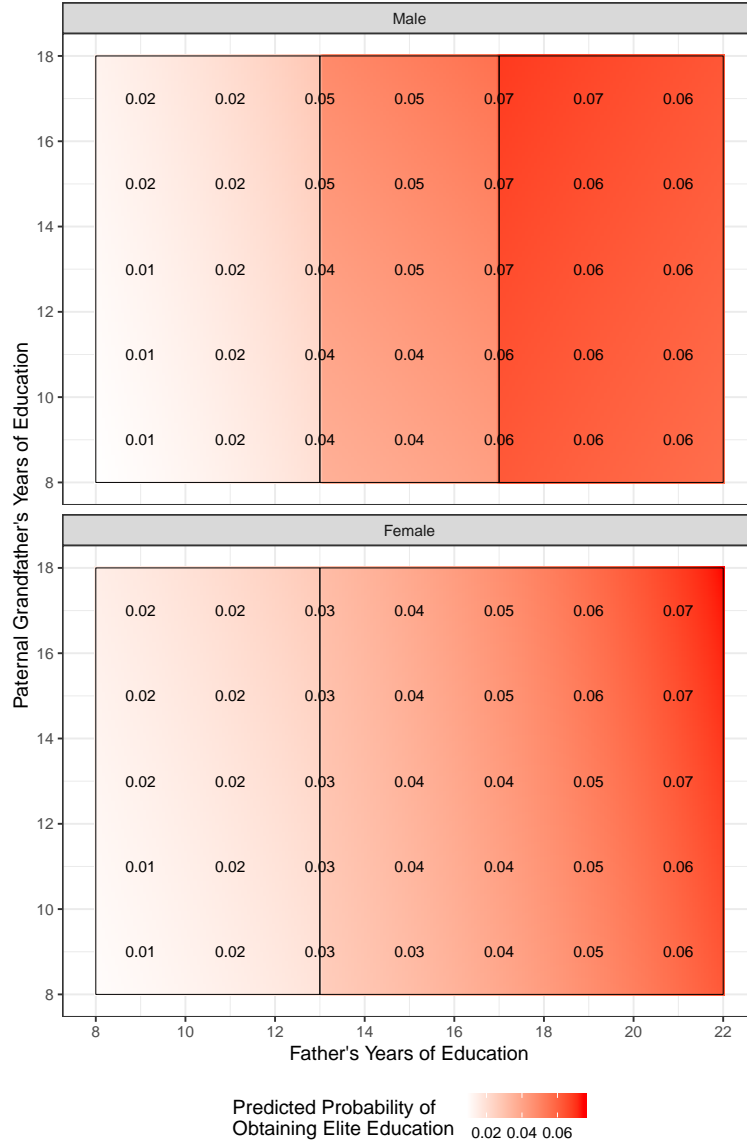
Note: The figure plots the predicted probability of obtaining an elite degree depending on grandfathers' and father's years of education. All other background characteristics (mother's years of education, father's and mother's earnings percentiles, aunts' and uncles' years of education, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.19: Generalized Linear Model Tree for Bachelor Degrees
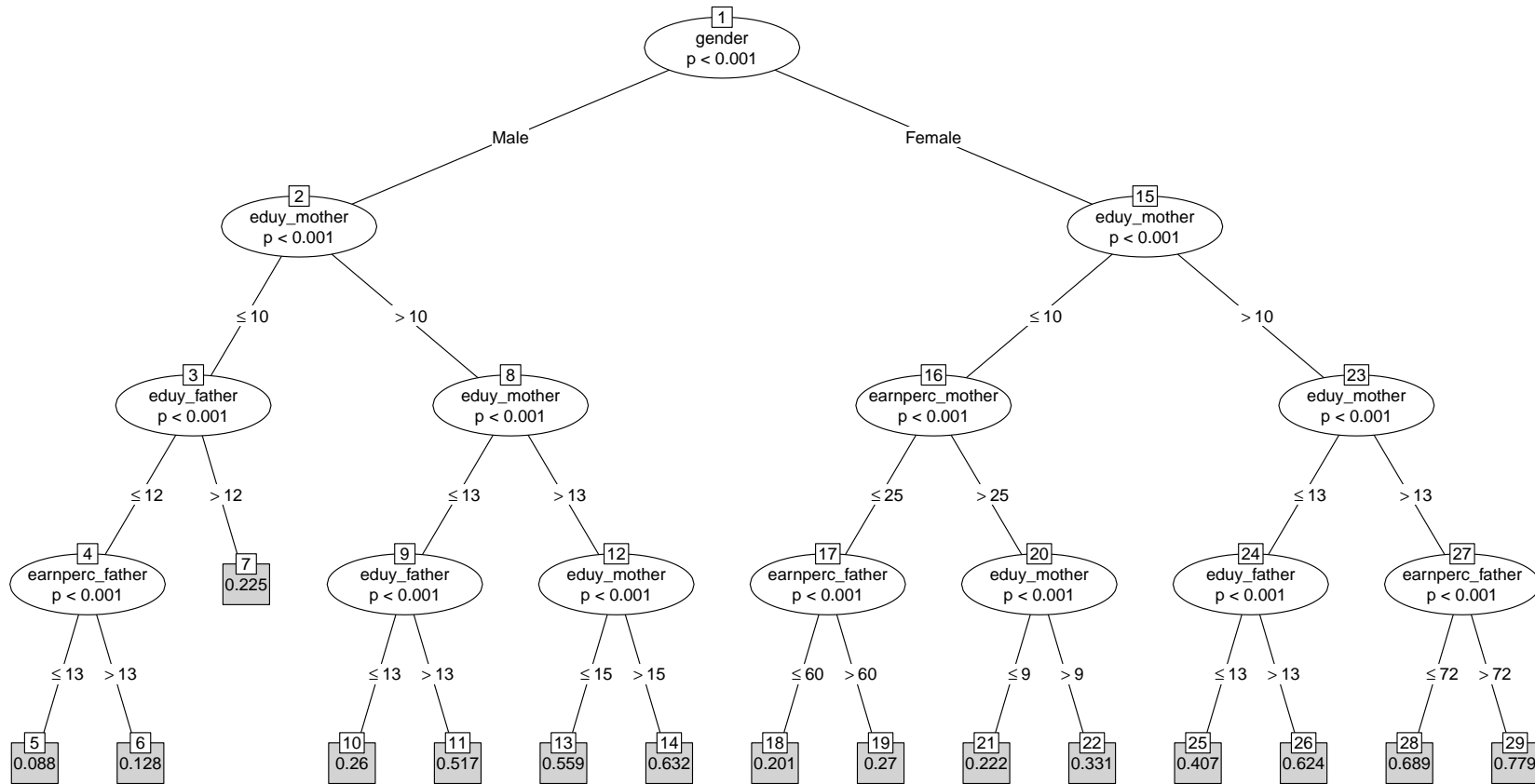
Note: The figure shows the split points and groups resulting from running the partitioning algorithm on Equation 2.1 with bachelor degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining bachelor degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.20: Marginal Effects of Father's and Mother's Years of Education, Bachelor Degrees
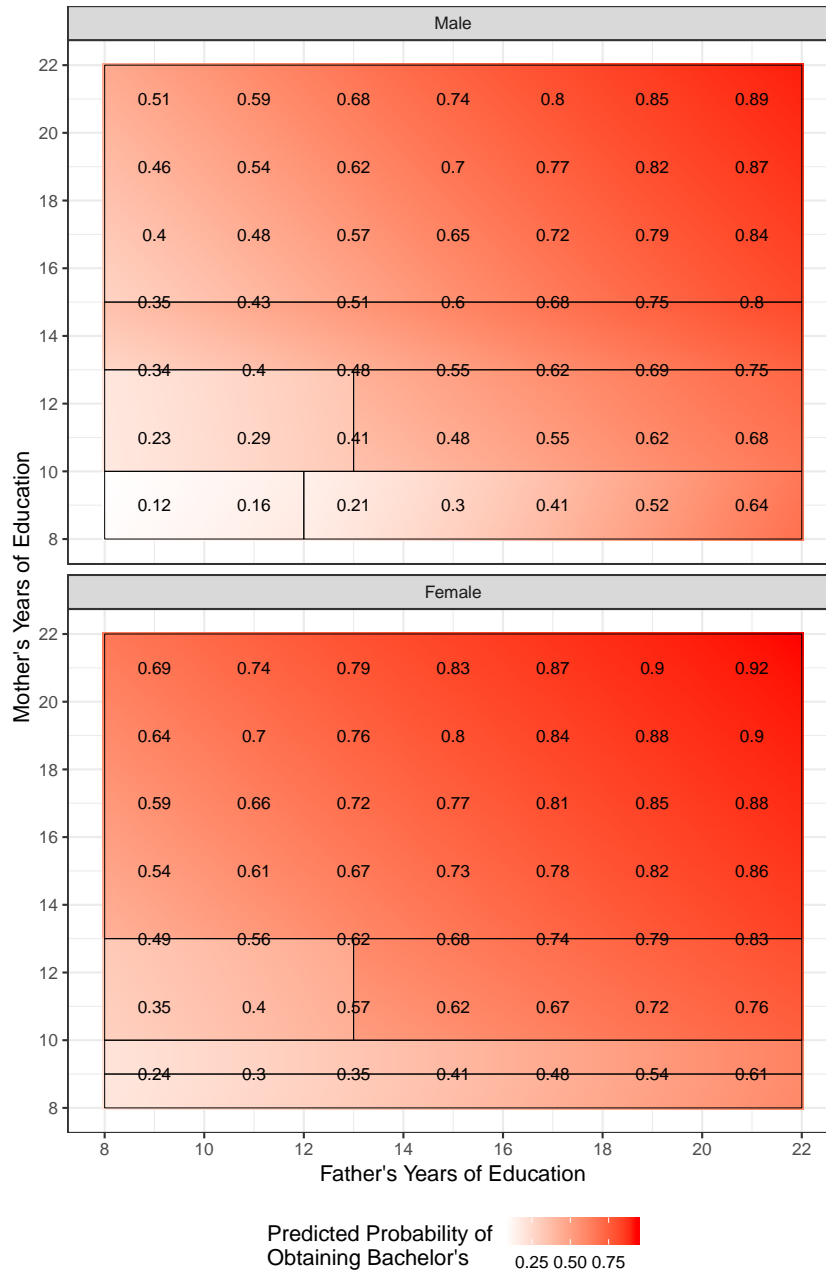
Note: The figure plots the predicted probability of obtaining a bachelor degree depending on father's and mother's years of education. All other background characteristics (father's and mother's earnings percentile, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.21: Marginal Effects of Father's Earnings Percentile and Father's Years of Education, Bachelor Degrees

Note: The figure plots the predicted probability of obtaining a bachelor degree depending on father's earnings percentiles and father's years of education. All other background characteristics (mother's years of education, mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.22: Alluvial Diagram Bachelor Degrees



Note: The figure shows the distribution of background characteristics of those who do and do not obtain bachelor degrees in terms of parental education and paternal earnings. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

116
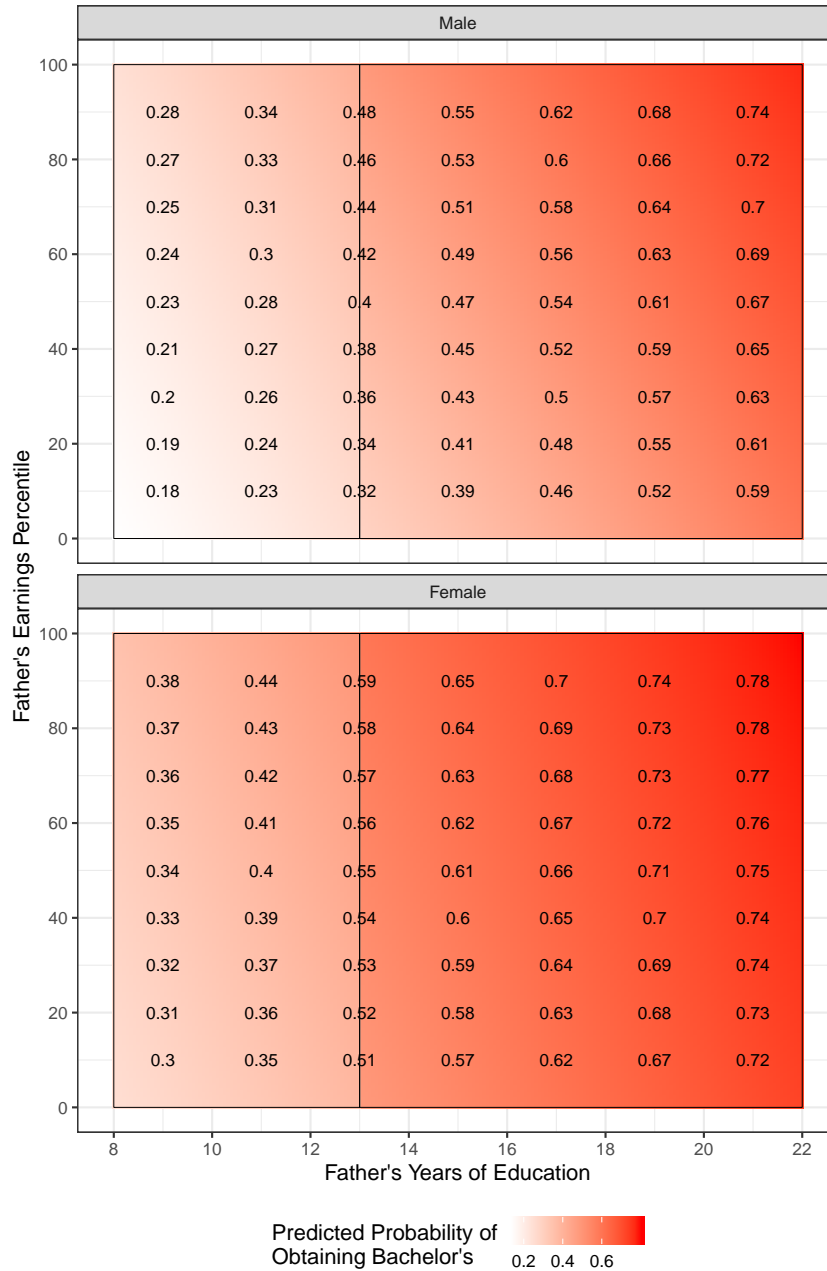
Figure 2.23: Generalized Linear Model Tree for Master Degrees when Controlling for Cognitive Ability, Master Degrees
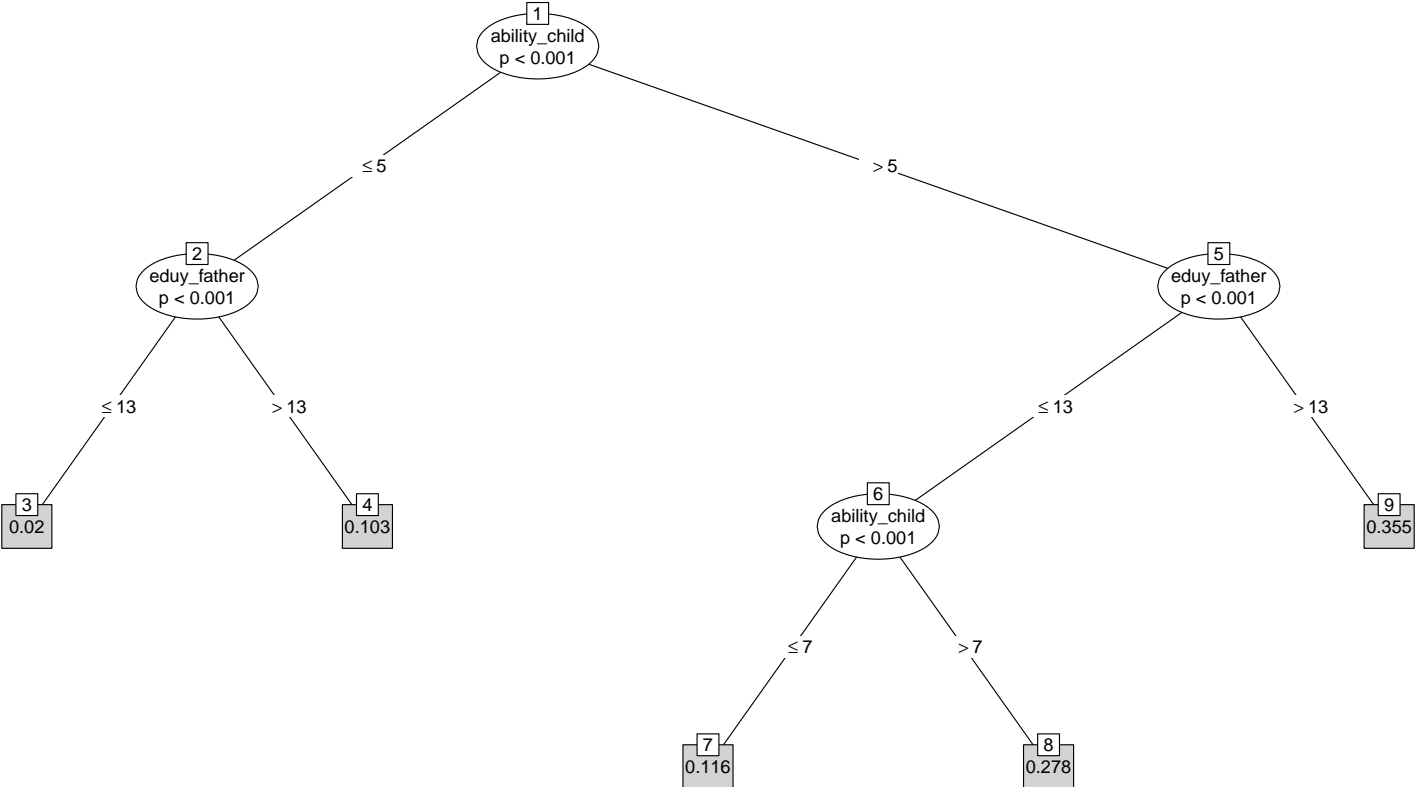
Note: The figure shows the split points and groups resulting from running the partitioning algorithm while including child's cognitive ability in the analysis, with master degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining master degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.24: Generalized Linear Model Tree for Elite Degrees Controlling for Cognitive Ability, Elite Degrees
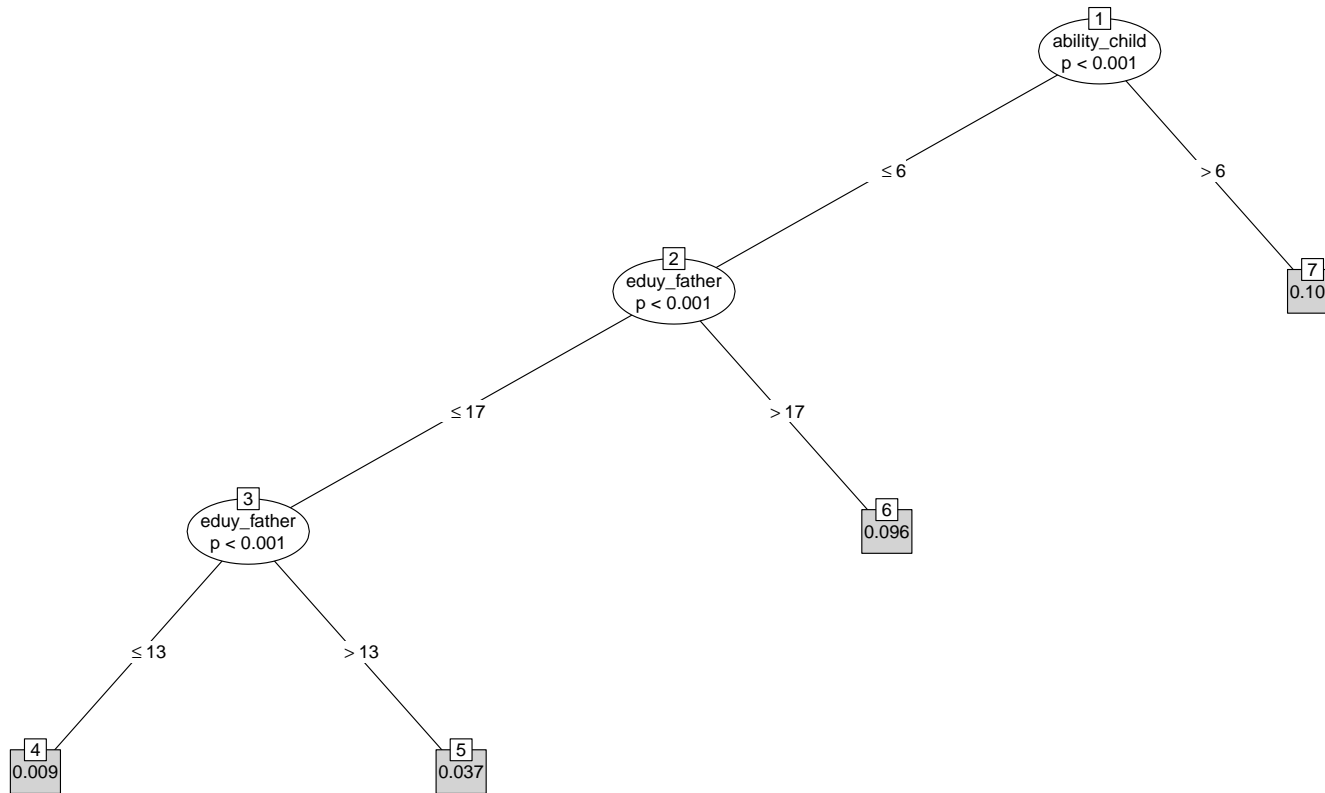
Note: The figure shows the split points and groups resulting from running the partitioning algorithm while including the child's cognitive ability in the analysis, with elite degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining elite degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.25: Generalized Linear Model Tree for bachelor Degree when Controlling for Cognitive Ability, Bachelor Degrees
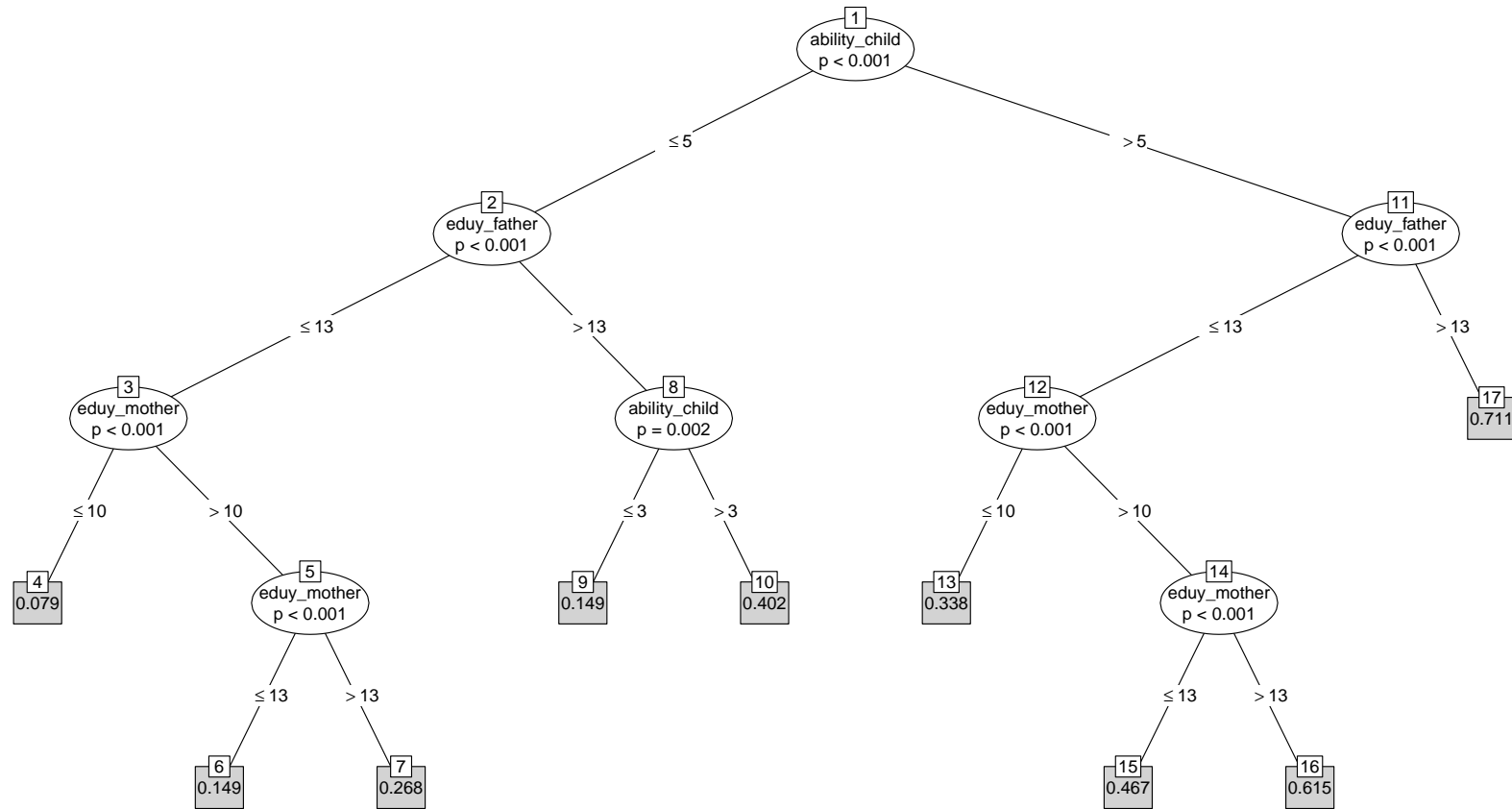
Note: The figure shows the split points and groups resulting from running the partitioning algorithm on Equation 2.3, i.e., while including child's cognitive ability in the analysis, with bachelor degrees as the outcome. The numbers in the terminal nodes at the bottom of the tree show the share of the group obtaining bachelor degrees. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.26: Marginal Effects of Son's Cognitive Ability and Father's Years of Education, Master Degrees
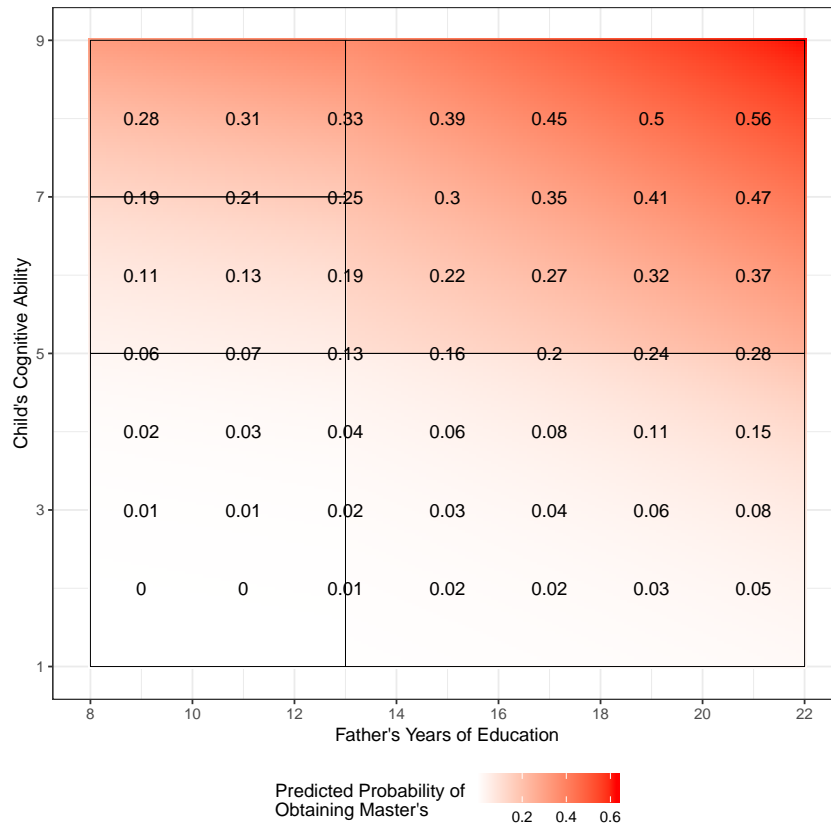
Note: The figure plots the predicted probability of obtaining a master degree depending on son's cognitive ability and father's years of education. All other background characteristics (mother's years of education, father's and mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.27: Marginal Effects of Son's Cognitive Ability and Father's Years of Education, Elite Degrees
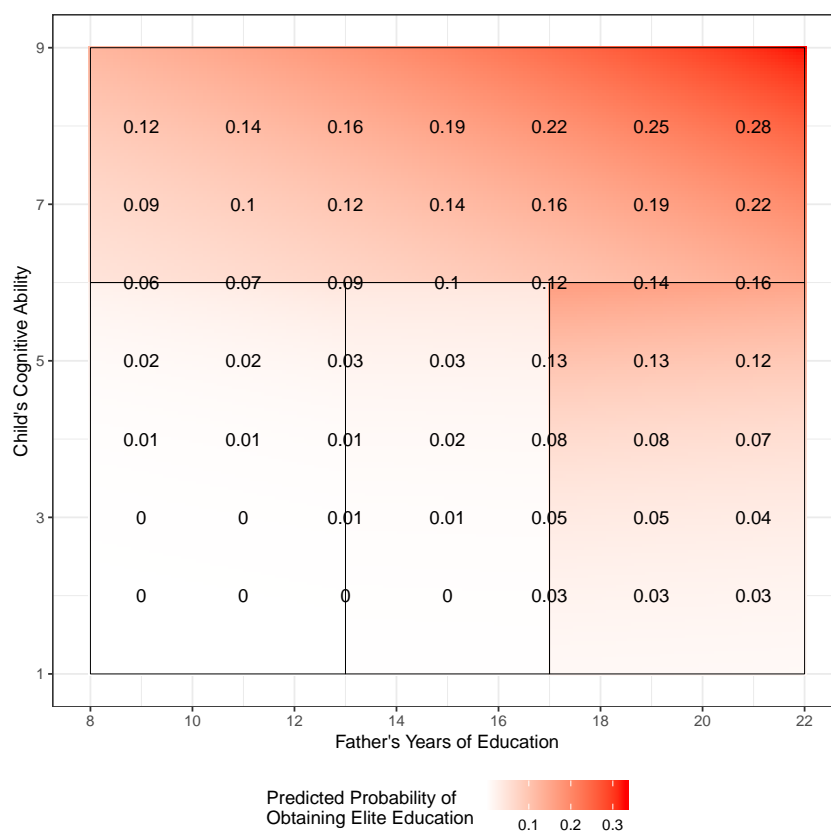
Note: The figure plots the predicted probability of obtaining an elite degree depending on son's cognitive ability and father's years of education. All other background characteristics (mother's years of education, father's and mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.28: Marginal Effects of Son's Cognitive Ability and Father's Years of Education, Bachelor Degrees
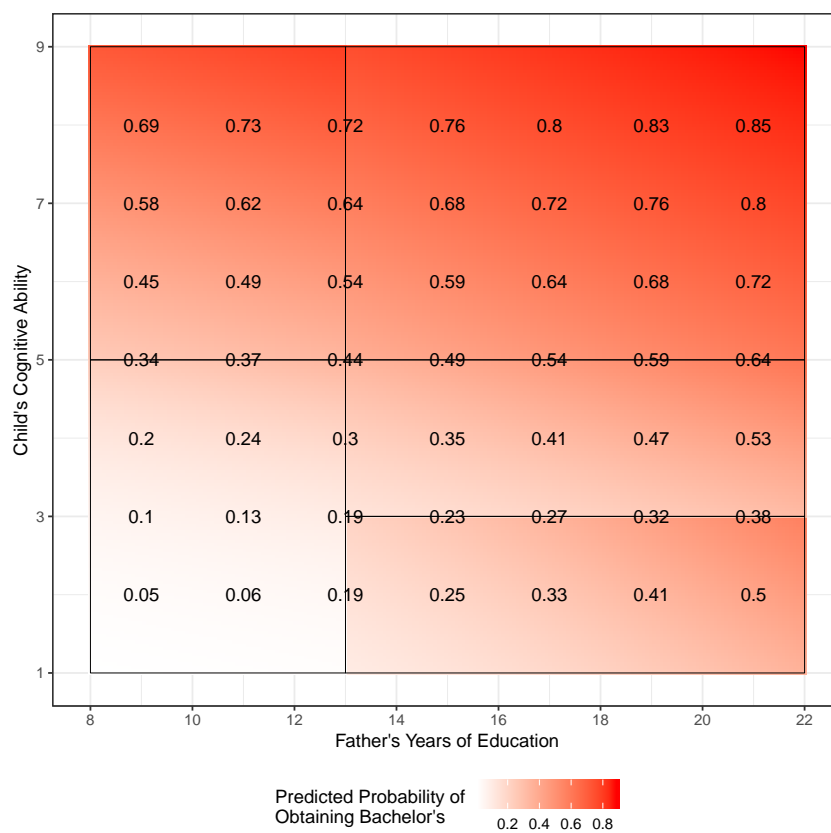
Note: The figure plots the predicted probability of obtaining a bachelor degree depending on son's cognitive ability and father's years of education. All other background characteristics (mother's years of education, father's and mother's earnings percentiles, year of birth, and labor market of birth) are held fixed at median values.

Figure 2.29: Alluvial Diagram Master Degrees when Controlling for Cognitive Ability
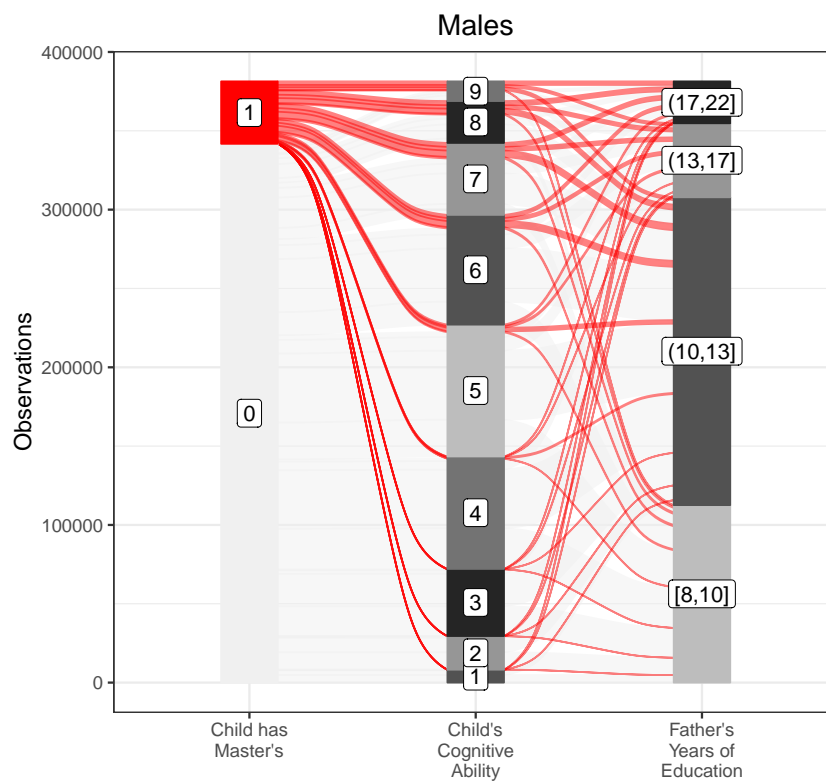
Note: The figure shows the distribution of background characteristics of those who do and do not obtain master degrees in terms of own cognitive ability and paternal education. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

Figure 2.30: Alluvial Diagram Elite Degrees when Controlling for Cognitive Ability
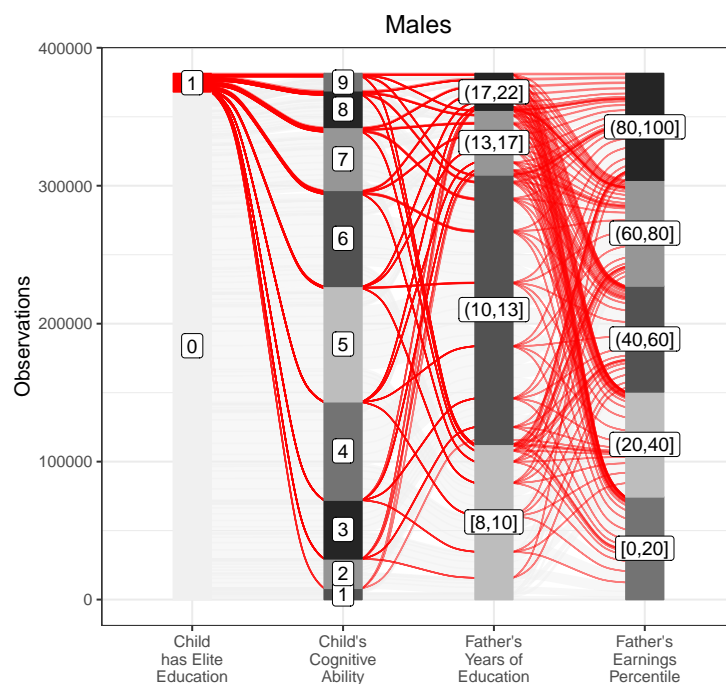
Note: The figure shows the distribution of background characteristics of those who do and do not obtain elite degrees in terms of own cognitive ability and paternal education. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

Figure 2.31: Alluvial Diagram Bachelor Degree when Controlling for Cognitive Ability
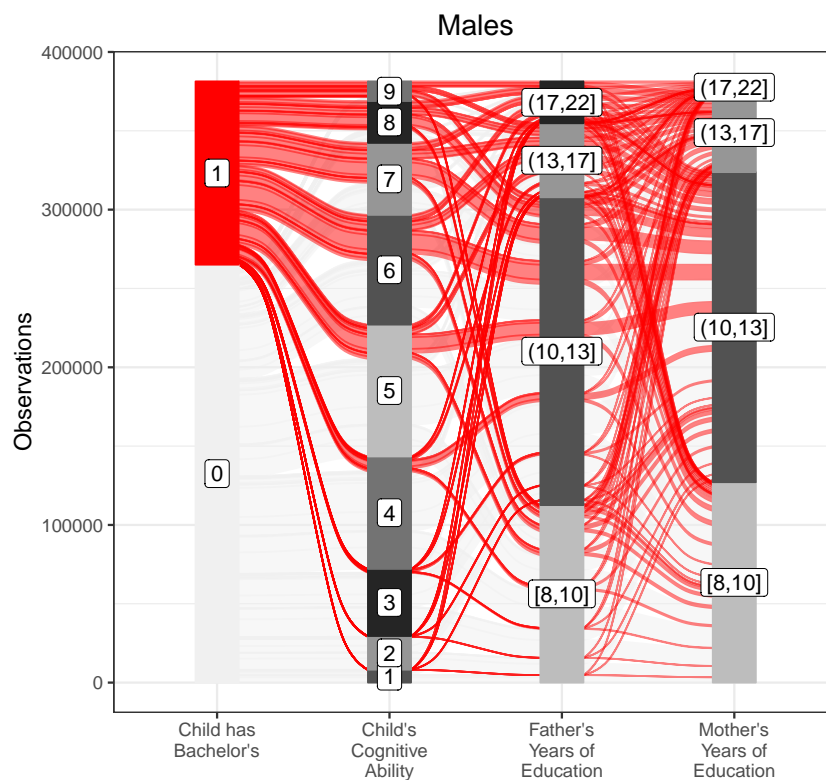
Note: The figure shows the distribution of background characteristics of those who do and do not obtain bachelor degrees in terms of own cognitive ability and parental education. Each of the columns show how the sample is distributed between the groups in that column. The flows between the columns show the prevalence of combinations of background characteristics and large flows indicate that a combination of characteristics is common whereas smaller flows indicate the opposite.

Figure 2.32: Linear Model Tree for Children's Earnings Percentiles
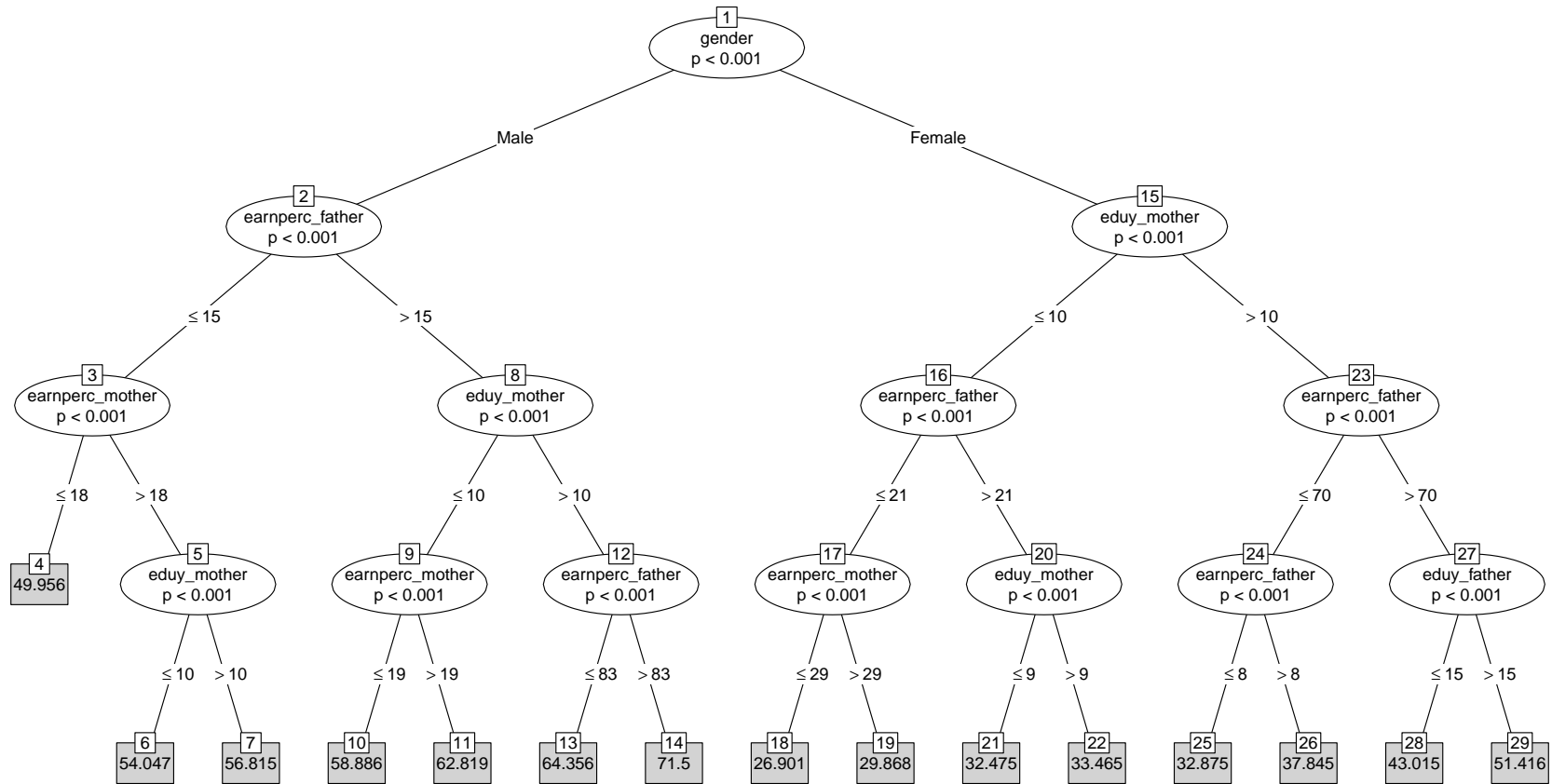
Note: The figure shows the split points and groups resulting from running the partitioning algorithm with children's earnings percentiles as the outcome. The numbers in the terminal nodes at the bottom of the tree show the average earnings percentile in the group. The nodes higher up in the tree show the split points selected by the algorithm.

Figure 2.33: Linear Model Tree for Children's Earnings Percentiles when Controlling for Cognitive Ability
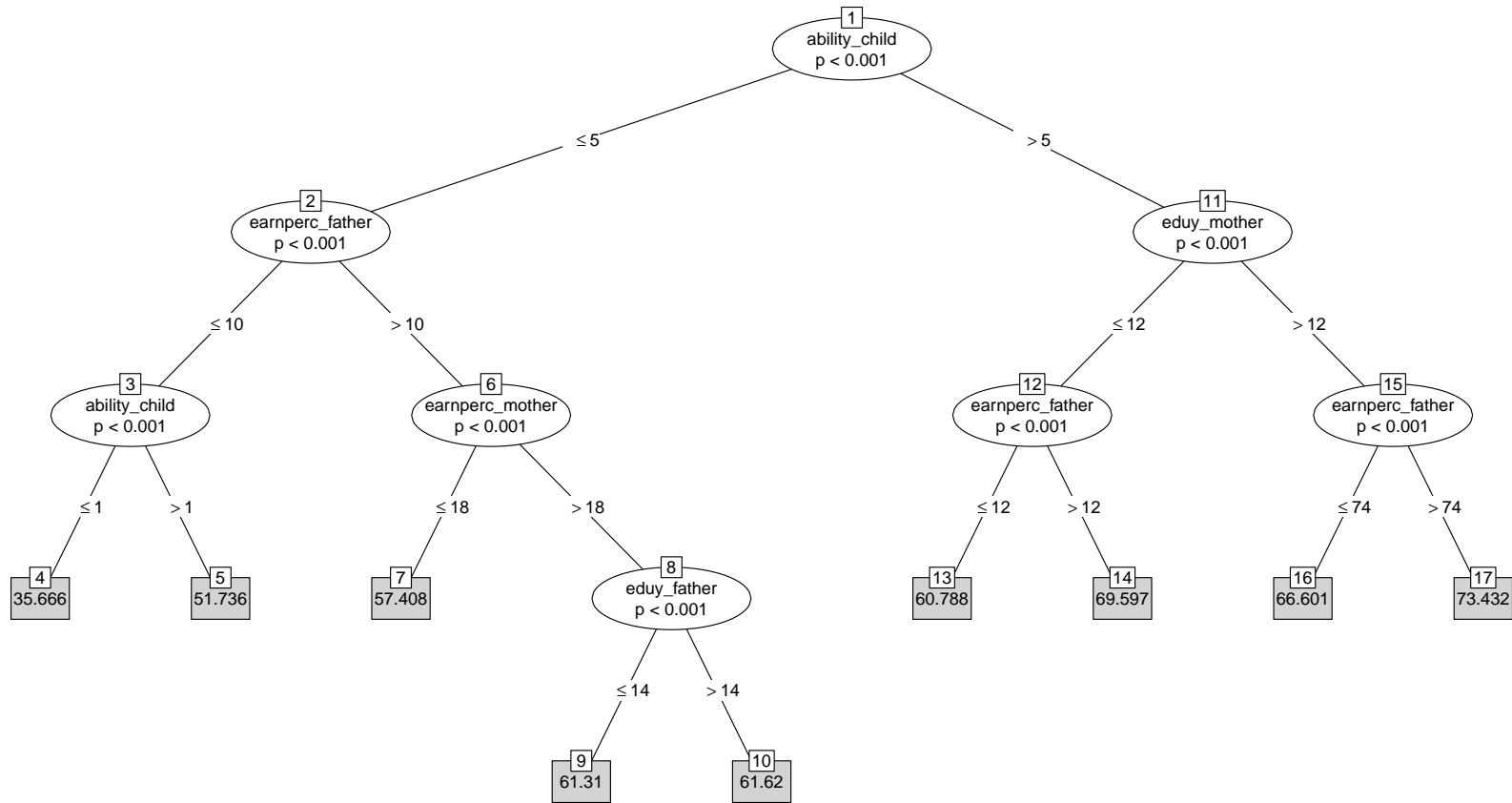
Note: The figure shows the split points and groups resulting from running the partitioning algorithm, while including the child's cognitive ability in the analysis, with children's earnings percentiles as the outcome. The numbers in the terminal nodes at the bottom of the tree show the average earnings percentile in the group. The nodes higher up in the tree show the split points selected by the algorithm.

Chapter 3

Intergenerational Mobility over Time and Across Regions in Norway

Pedro Carneiro, Sarah Cattan, Sonya Krutikova, Erling Risa and

Kjell G. Salvanes*

Abstract

In this paper we analyze intergenerational mobility in Norway for cohorts of children born from the mid 1950s until the mid 1980s and are grown up today. We focus on regional differences and changes across regions and within regions over time. We use several measures of income mobility, and in addition to relative mobility measures like rank-rank, we use measures to detect changes at different margins, like moving from the bottom to the top quintile and the share of sins have higher earnings than fathers. Next, we focus on the mechanisms behind the differences in mobility across regions and changes over time. We are particularly interested in the role of human capital investments, the role of the labor market and returns to human capital and

*Carneiro: Department of Economics, University College London; Cattan: IFS; Krutikova: IFS; Risa: FAIR, Department of Economics, NHH Norwegian School of Economics; Salvanes: FAIR, Department of Economics, NHH Norwegian School of Economics

characteristics of the industrial structure and other labor market characteristics. We use machine learning to identify regional differences and labor market differences. These to parts of the analysis will be analyzed together within a panel regression framework in the next step.

3.1   Introduction

Norway is among the OECD countries with the highest income mobility and lowest cross-sectional income inequality countries as measured by for instance the Gini coefficient (Corak 2013; Bratberg, Davis, Mazumder, Nybom, Schnitzlein, and Vaage 2017; Pekkarinen, Salvanes, and Sarvimäki 2017). However, are there regions that are more mobile and some that are poverty traps as has been found in a series of papers from the US (Chetty, Hendren, Kline, and Saez 2014; Chetty, Hendren, Kline, Saez, and Turner 2014), and for countries like Canada and Italy (Connolly, Corak, and Haeck 2019; Güell, Mora, and Solon 2018). Moreover, have there been changes over the last decades? And what are the mechanisms behind the differences in mobility across regions and changes over time, or more specifically what is the role of human capital investments, the role of the labor market and returns to human capital and characteristics of the industrial structure and other labor market characteristics.

In this paper we first document the patterns across local labor markets and regions in Norway of cohorts born from the mid 1950s to the mid 1980s and around the age of 30 today, by leveraging an extensive and population wide register data connecting families and regions over time along outcome dimensions as income and education. We also document the development across regions and time along different margins of mobility, using different

measures of intergenerational mobility, such as a relative measure as rank-rank which comprise both upward and downward mobility, the predicted rank for growing up poor and growing up rich (20 bottom and top parental percentile), and going from rags to riches, share of children going from bottom 20 percentile parental rank to the top 20 percentile, as well the share of sons with higher earnings than their fathers in order to measure welfare improvements. We are in particular interested in the role of human capital investments by parental background over time and across regions, and the role of the labor market opportunities for different socio-ecnomic groups across regions and over time, and we estimate the relationship between parental income rank and child educational attainment. We document this also for the bottom and top parental income percentile, and for different educational margins, high school, college, and postgraudate or master programs.

While Norway is characterized by a high degree of income mobility, it is more similar in terms of intergenerational mobility in education and especially at highest education level (Björklund and Salvanes 2011; Bütikofer, Risa, and Salvanes 2019).

Chetty and Hendren (2018a) and Chetty and Hendren (2018b) find support for that especially differences in human capital investments (as measured by high school completion, test scores, and school expenditures) are drivers correlation with the regional difference in income mobility. Rothstein (2019) and Kourtellos, Marr, and Tan (2018), however, find stronger support for the role of the labor market in explaining the regional pattern in the mobility pattern in the US by assess differences in returns to human capital for different groups across regions. As compared to these to major advances in the literature, we leverage a very rich data set and focus both on the regional

dimension but also the changes over time across regions. Other parts of this new wave of literature analyse the effect of industrial composition, growth and non-growth areas, social capital etc. (Connolly, Corak, and Haeck 2019; Güell, Mora, and Solon 2018).

In order to assess the importance of human capital investments for socio-economic groups, the role of the labor markets, and the role of industrial composition, growth etc, we next focus on understanding which characteristics of the regional labor markets that are important in explaining these patterns, and in particular the role of human capital formations and the role of the labor market. In order to do this, we will employ machine learning techniques in order to handle the challenges of high dimensional data. The first challenge posed by high dimensional data is the fact that it is not clear ex-ante which of the many potential variables belong in the analysis. A second and related challenge is the fact that including many variables increases the variance of the estimates, making statistical inference more difficult. This second point is especially relevant in this context because the number of regions and years of data available naturally limit the sample size in the analysis. Moreover, several of the variables, such as measures of income inequality within regions at different points in time, are correlated, further increasing the variance of the estimates. The plan for a next step, is to use a coherent panel data framework for analysing the relationship between income mobility across time and regions, and the what the main mechanisms are.

Our main findings so far are as follows. The rank-rank correlation of gross earnings has remained remarkably stable since the 1950s. It is about half as large as the association estimated by Chetty, Hendren, Kline, and Saez

(2014) for the US. The time-trend is similar across the regions, however, income mobility has been persistently highest in the Western and Southern regions and lowest in the Eastern region. Breakdowns by father's earnings quintile show that while the mobility gap between the Western and Eastern regions is driven by differences in mobility at the bottom of father's income distribution, there are differences at both the bottom and the top when comparing the Southern and Eastern regions. We find that around 12-13 percent of children goes from the lowest to the highest quintile as a national average, and increasing over time. There are large and quite stable differences across regions, again with the West as the most mobile from the bottom to the top, and the North and the Middle region with the lowest mobility. Interestingly, we also find that for all regions, the share of sons making more money than their fathers, is high and increasing over time, again with the Western region with a higher share than the rest. We see some convergence over time, except for the East region. At the national level, the share of sons with higher earnings than their fathers starts out in the 50s cohort at around 70 percent similar to the US, but in stead of steadily going down as in the US, it increases to about 85 percent in Norway.

In contrast to the stability in the overall income mobility over time - although changes over time and across regions especially with mobility at the bottom, education mobility measured by strength of association between father's income and child years of schooling has increased, especially sharply in the late 1950's and early 1960's - for the 1985 cohort it is very similar to Rothstein (2019)'s estimates for the US. While for the earliest cohorts there was a gap of nearly 2.5 years in the mean schooling of children of fathers in the top and bottom quintiles of the earnings distribution, for cohorts born in the late 70's it was a full year smaller; further, the difference in the proportion of children

completing high-school between the top and the bottom quintiles has nearly halved since the late 1950's.As with income mobility, education mobility is persistently highest in the Western region and lowest in the Eastern region. The time-trend is similar across most of the regions with the North as the exception. It starts out as one of the most mobile regions in the 1950's and 60's, alongside the Western region and ends up as the second least mobile region by the 1980's.We see the reverse pattern when looking at post graduate studies or master degree completion. Here mobility has, if anything, gone down over time and there has been little/no change in the difference in master degree completion rates between children of fathers in the top and bottom income quintiles. There is substantial regional variation with big differences in mobility trends for post graduate completion in Northern and Middle regions relative to the rest. The difference in trends is driven by differential trends at the top of parental income distribution rather than bottom.Using IQ as the child measure rather than schooling appears to yield similar results to years of education and probability of completing high school. However, currently scores for the earlier cohorts don't seem to be comparable to scores for the older cohorts. In order to assess the role of the labor market for human capital investments, we estimate returns to years of education in a simple Mincer equation framework. First, we notice that returns differ across regions, from close to zero to above ten percent, and with a tendency to an overall decline over time both for fathers and children. However, estimating returns to higher education, the returns to college vs high school, and returns to a master degree vs high school, the returns are stable over this time period. Again there are big differences across regions indicating an important role for the labor market in human capital investment. Especially the Western region has low returns to the post

graduate level.

The paper unfolds as follows. We start by providing a careful description over time and across regions of intergenerational mobility in Norway. We provide a relative measure of intergenerational mobility using ranks for parents and children, and then extend to measures focusing on where in the distribution there is strongest persistence, the share of children going from rags to riches, and then at last providing a measure of absolute mobility measuring the share of children with higher earnings than their fathers. This measure may be thought of as measuring whether the standard of living is increasing across generations. We will provide all of these measures will be present across cohort and across cohorts over regions. Then we present results for the role of parental income and educational achievements, both overall in years of education and for different educational margin. Simple Mincer equations are estimated with age controls, again both for returns to a year of education and at different margins. We then turn to a section using Machine Learning where we present a framework how to detect which background variables are important in explaining the regional and time patterns in intergenerational mobility. We then sum up our finding and suggest the next step for a coherent panel data framework for analysing the relationship between income mobility across time and regions, and the what the main mechanisms are.

3.2   Data and Variables

The core data we exploit in this paper is Norwegian administrative data from different administrative units such as tax authorities and educational institutions and census data from 1960 and 1980. We exploit two features of

these data in the present paper: the long panel structure and the parent–child link. The panel structure allows us to identify where and when people are born and follow them for several decades in the data - both in terms of educational and employment, but also across regions. Hence, we are, for example, able to construct characteristics of the regions they grew up and link those to medium- and long-term educational outcomes and earnings. The fact that the data contain unique personal identifiers enables us to link children and their parents, grandparents, and siblings, as well as aunts and uncles.

We focus on the children born from 1955 to 1985 to ensure that we have enough education and income data for their parents as well as adult outcomes of the children. A detailed description of the different variables in the data and descriptive statistics are provided below.

3.2.1 Earnings

We calculate most of the earnings measures used in the analysis based upon before tax gross earnings from national tax records. Earnings consists of some transfers such as unemployment benefits and parental leave benefits, i.e., benefits related people in the labor force. Earnings percentile ranks form the basis for many of the earnings measures we include. When calculating percentile ranks for the children in the analysis, we first calculate average annual earnings when they are between 30 and 35, before generating ranks within each cohort nationwide. We include both men and women and calculating ranks jointly for both genders. For the adults, the earnings percentiles are based upon average annual earnings when the fathers are between 30 and 50 years old and calculated separately for each cohort nationwide. These per-

centile ranks for parents and children are then used to estimate the rank-rank slopes used in the analysis.[2]

At the regional level, we calculate some additional earnings measures. Based on the percentile ranks, we calculate the share of the working-age population in each region with earnings below the tenth earnings percentile, giving us a measure of people with low earnings. Moreover, we calculate the average percentile ranks of the regions to obtain a measure of where the average individual in the region falls in the national earnings distribution.

The analysis also includes some additional earnings measures based on earnings levels rather than percentile ranks. One such measure is the Gini coefficient of pension generating earnings within each region. In order to obtain some further measures of inequality, we also calculated the 10th, 50th, and 90th earnings percentiles conditional on gender and education level for each region. When conditioning on education level we split into mandatory schooling, high school and college education.

Since the characteristics of a region may influence the outcomes of an individual both in childhood and adulthood, the analysis includes regional earnings measures calculated both in childhood and adulthood. We define childhood as when the individuals are 12 years old, while adulthood is when they are 25 years old.

3.2.2   Education

An individual's highest completed level of education is measured from the education register, and any educational institution is legally obligated to

---

[2]For the rank-rank slopes, we calculated earnings percentiles only among fathers

report any student who completes a degree to Statistics Norway. We classify individuals into three education groups: low- (basic and lower secondary education), medium- (upper secondary education), and high-educated (post-secondary level). We also construct education levels at the regional/local labor market level. The first measure is the share of individuals with mandatory education (ten years) or less as their highest level of completed education. The next measure is the average years of education within each region.

A little background on the education system in Norway may be useful since human capital will play an important part in the analysis. The Norwegian education system consists of four levels: primary school (grades 1–7), middle school (grades 8–10), high school (grades 11–13), and college and university education. Currently, Norwegian compulsory education starts at age six, lasts for 10 years and consists of primary and middle school. Compulsory schooling is organized by Norwegian municipalities and the vast majority (98%) of pupils attend public, local schools. All pupils are allocated to schools based on fixed school catchment areas within municipalities. Except for some religious schools and schools using specialized pedagogic principles such as Montessori schools, there is no school choice among publicly funded schools. The school system gets more competitive from middle school onward, where exit exams and grades from teachers are crucial for entry to the best high schools. High schools have two main tracks: vocational and academic. The academic track prepares students to attend college and there are different specializations within the academic track (e.g., science, languages, music). Within the vocational track, there are currently nine distinct programs such as "health and child development", "restaurant and food", and "construction" teaching students professional qualifications. High schools are administered at the county level (above the level of municipalities). High school is not

mandatory in Norway, however, since the early 1990s all students graduating from middle schools are guaranteed a slot in high school.

Higher education consists of universities, scientific schools, and university colleges. Until the late 1990s, when the majority of our sample attended higher education, Norway had three universities (Oslo, Bergen, and Tromsø) offering a very wide range of study subjects including medicine, law, humanities, and social science, and six scientific schools offering specific subjects at university level such as business schools, engineering schools, architect and design schools, and arts schools. In 2019, there were ten universities (most of them established after 2000) and nine scientific schools. Since the early 2000s, Norwegian universities and scientific schools offer three-year bachelor and five-year master degrees. Students would usually receive a bachelor degree after three years and then continue for two additional years to receive a master degree. Before 2000, a degree from these institutions would normally last 3–6 years. In addition, there are many regional university colleges offering mostly professional degrees in business, health, and teaching. Today, these institutions mostly offer bachelor degrees. Before 2000, university colleges offered two- to four-year degrees. Most students attend a public institution, and even private institutions are funded and regulated by the Ministry of Education and Research. There are generally no tuition fees for attending public higher education in Norway, and most students are eligible for financial support (part loan/part grant) from the Norwegian State Educational Loan Fund.

We supplement this individual-level panel spanning 1967–2010 with data on IQ test scores from Norwegian military records for cohorts born after 1949 as a proxy for cognitive ability (see Sundet, Barlaug, and Torjussen

2004; Thrane 1977). The military service in Norway is mandatory for men, and each male individual is tested for physical and psychological suitability during the examination around his 18 birthday. The IQ score is constructed as an unweighted mean of three tests–arithmetics, word similarities, and figures–and converted into a single digit number on a 0 to 9 scale, so that the scores are normally distributed with a mean of 5 and a standard deviation of 2.

3.2.3 Local Labor Markets and Regions

We employ two types of regional aggregations in this paper. The first type of aggregation is a classification by Gundersen and Juvkam (2013) which results in 160 regions. The regions are classified based on a range of characteristics, the three most important of which are "location in the centre structure; commuting frequency between municipalities; and travel between municipal centres" (Gundersen and Juvkam 2013, p.9), meaning that the classification is similar to the commuting zones used in Chetty, Hendren, Kline, and Saez (2014). The second level of geographical aggregation we use is five larger regions based on the counties. These five regions group regions based on whether they lie in the east, south, west, middle or north of Norway.

3.2.4 Industry Composition

The employment measures used in the analysis come from two separate sources. Employment measures when the individuals are adults come from annual employment records at Statistics Norway. In order to maximize consistency, we aggregate the employment codes to the highest aggrega-

tion level, which corresponds to 17 industries. The one exception to this is petroleum-related industries which we identify using codes at the second-highest aggregation level. We proceed by calculating the share of the labor force employed in each industry every year within each region. We assign adult employment shares to the individuals in our sample based on the region in which they were born and measure it when they are 31 years old. For a few of the youngest cohorts, we do not observe employment shares in the data when they are 31. In these cases, we use the most recent year of data we have available.

In order to have employment measures also during the individuals' childhood, we supplement the annual records with employment data from the 1960, 1970, and 1980 censuses. The occupation codes from the censuses follow the 1978 edition of the Standard for Industrial Classification. As with the annual employment records, we aggregate the occupation codes from the census data to the highest aggregation level, resulting in nine industries. Also, as with the annual records, we identify petroleum-related industries at the second-highest aggregation level. Because censuses were only conducted every ten years, we assign the employment shares from the 1960 census to the cohorts born 1955-1965, the employment shares from the 1970 census to those born1965-1975 and the shares from the 1980 census to those born 1975-1985.

3.3   Institutional Context and Descriptives

We start by providing a careful description over time and across regions of intergenerational mobility in Norway. We provide a relative measure of intergenerational mobility using ranks for parents and children, and then

extend to measures focusing on where in the distribution there is strongest persistence, the share of children going from rags to riches, and then at last providing a measure of absolute mobility measuring the share of children with higher earnings than their fathers. This measure may be thought of as measuring whether the standard of living is increasing across generations. We will provide all of these measures will be present across cohort and across cohorts over regions.

3.3.1 Intergenerational Mobility in Earnings over Time and the Income Distribution

We first show how the mobility of earnings changes over time by splitting into three cohort for cohorts born 1955-1985 with measured income at the age of 35-40. We includes several measures of intergenerational mobility in income. The first one, which we call the rank-rank income slope is obtained by estimating the following regression:

$$\text{RankEarnings\_Child}_i = \alpha + \beta \text{RankEarnings\_Father}_i + \varepsilon_i, \qquad (3.1)$$

where $\text{RankEarnings\_Child}_i$ signifies child i's earnings percentile in her cohorts national income distribution, Earnings_Father_i is the earnings percentile in the national income distribution of individual i's father and β is the coefficient of interest. The results are provided in Figure 1.

On average for the most recent cohorts born between 1980 and 1985, a 1 percentile difference in father's earnings is associated with a 0.168 percentile difference in child's eventual income. This is slightly lower than what

Figure 3.1: Rank-Rank Intergenerational Mobility National Average
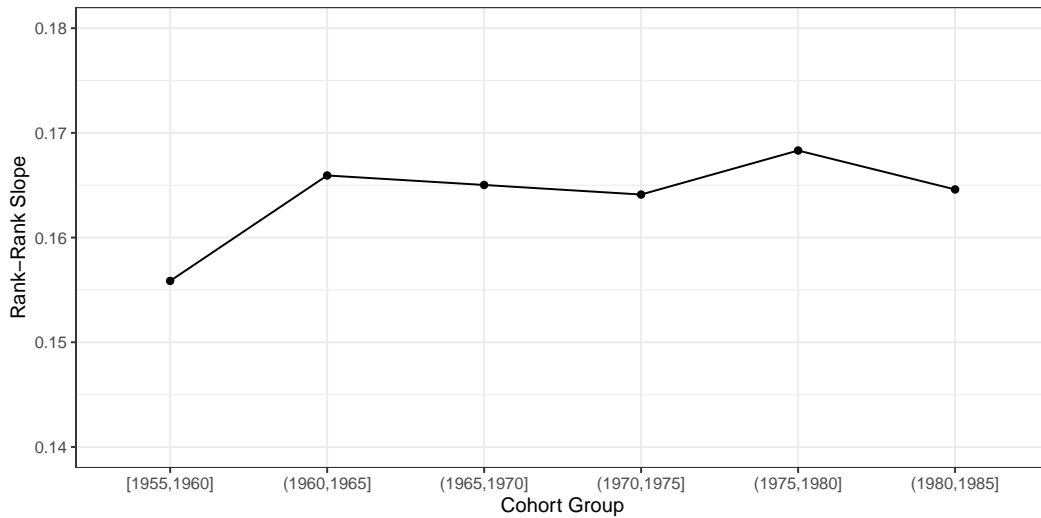
Landersø and Heckman (2017) find for Denmark (0.205 for wage earning including 0's) and much lower than what Chetty, Hendren, Kline, and Saez (2014) finds for the US, where on average a 1 percentile difference in parental income is associaes with 0.32 difference in child income. There has been little change over time - among the earlier cohorts the rank-rank association is around 0.162, peaking at 0.17 in early 1960's. This result support the result for Norway that intergenerational mobility was very stable over time in the post WWII period and in line with other what other report for Norway in this period (Pekkarinen, Salvanes, and Sarvimäki 2017; Butikofer, Dalla Zuanna, and Salvanes 2018; Bratberg, Davis, Mazumder, Nybom, Schnitzlein, and Vaage 2017). It is also similar for other Nordic countries and placing these countries among the high intergenerational income mobility countries (Black and Devereux 2011; Corak 2013).

In order to assess whether there is a difference in persistence across the father's income distribution, we estimate the three measures of absolute mobility. First, we predict by cohorts the rank for children for the lowest and the highest quintile for the fathers. Second, we estimate the share of children

going from the bottom to the top quintile. Third, we estimate the share of children earnings more than their fathers.

First, we estimate estimating the following regression for the same cohorts:

$$\text{RankEarnings\_Child}_i = \alpha + \beta \text{RankEarnings\_Father}_i + \varepsilon_i, \qquad (3.2)$$

where $\text{RankEArnings\_Child}_i$ signifies child i's earnings percentile in her cohorts national income distribution, Earnings_Father_i is the earnings percentile for the highest and lowest quintile in the national income distribution of individual i's father and β is the coefficient of interest which we use to predict the child's percentile. We present the results in Figure 3.2.

Figure 3.2: Mean child income ranks for top and bottom father's income rank quintiles - national average
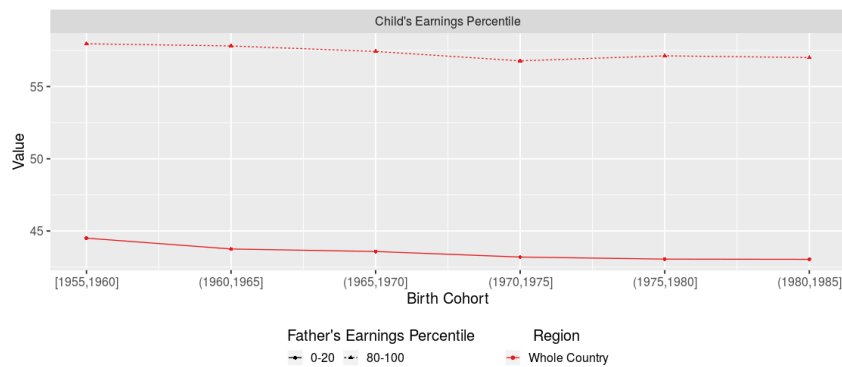
The persistence over time holds at both ends of the father's earnings distribution. Throughout the period, children of fathers in the top earnings quintile ended up in around the 57th/58th percentile, and those in the bottom in between 45th and 43rd. The result for the highest quintile is slightly lower than for the US, where Chetty, Hendren, Kline, Saez, and Turner (2014) report in their Figure 3 that it is about percentile 60-65. The predicted percentile for children from the bottom quintile is around the 35th percentile for the US and just above the 40th percentile for Canada (Connolly, Corak, and Haeck

2019).

In Figure 3.3 we present the absolute earnings level for the highest and lowest quintile in order to have sense of what the numbers means in terms of absolute (real) values in Norwegian Kroner (NOK).

Figure 3.3: Mean child income in real Norwegian Kroner for top and bottom father's income rank quintiles - national average
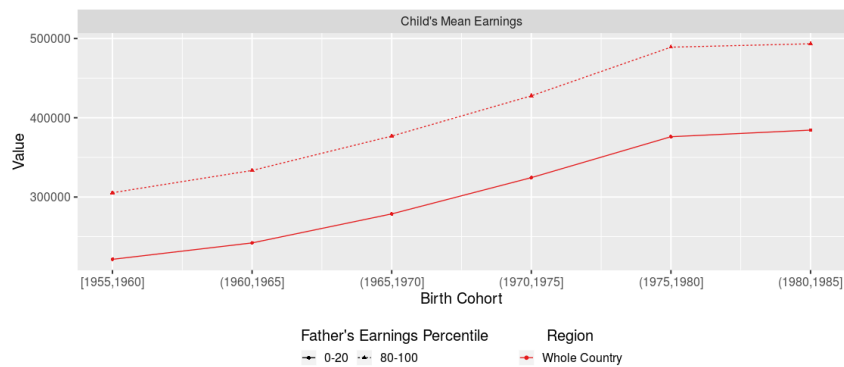
We see that on average in the most recent cohorts, children of fathers with earnings in the bottom 20% of the distribution earn 32% less than children of fathers with earnings in the top 20%. There is some evidence of an increase in inequality: the gap in mean earnings between children of fathers in the top and bottom earnings quintiles went up by 50% from around NOK 80,000 in late 1950's to NOK 120,000 in early 1980's.

Second, we estimate the share of children going from the bottom to the top by the following equation:

$$\text{Top\_Quintile\_Child}_i = \alpha + \beta \text{Top\_Quintile\_Father}_i + \varepsilon_i, \qquad (3.3)$$

where $\text{Top\_Quintile\_Child}_i$ is the share of bottom to top movers in her cohorts national income distribution, $\text{Top\_Quintile\_Father}_i$ is the earnings percentile for the highest and lowest quintile in the national income distribution

of individual i's father and β is the coefficient of interest which we use to predict the child's percentile. The result is presented in Figure 3.4 below.

Figure 3.4: Share Going from Bottom to Top Earnings Quintile
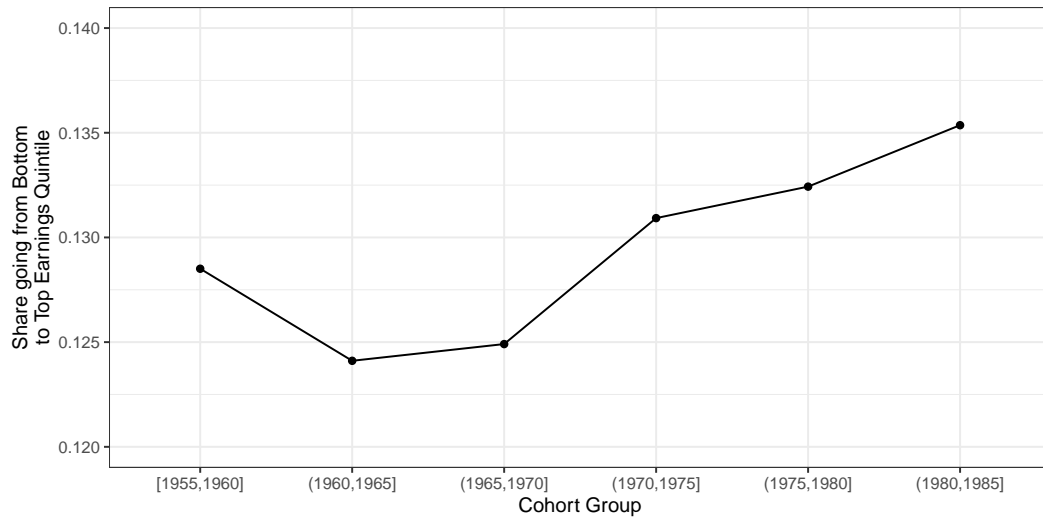
From Figure 3.4 that for the youngest cohorts, born 1980-1985, the share going from rags-to-riches is about 13 percent. This above compared with results for Canada where Connolly, Corak, and Haeck (2019) report a share of 11.2 percent, and is well above the US where Chetty, Hendren, Kline, and Saez (2014) report only 7.5 percent. There is a slight increase going from the bottom to the top quintile across cohorts, rising from a share of about 12 percent for the cohort born 1955-1960, to 14 percent for the most recent cohorts.

Lastly we calculate the share of sons with higher earnings than their fathers as measured at ages 34-36 and show the results in Figure 3.5.

Son's and Father's earnings are measured at the same ages (34-36) which is found to be the optimal age for measuring life time income for Norway (M. Bhuller, Mogstad, and Salvanes 2017). Interestingly, for Norway this share is increasing over time, going from a share of a little less than 70 percent for
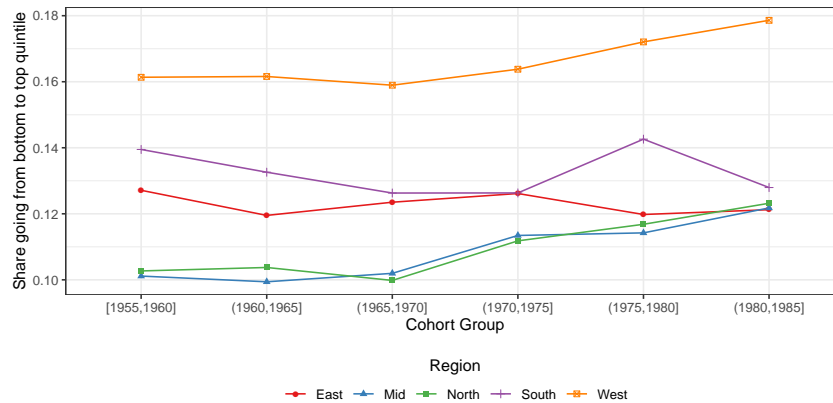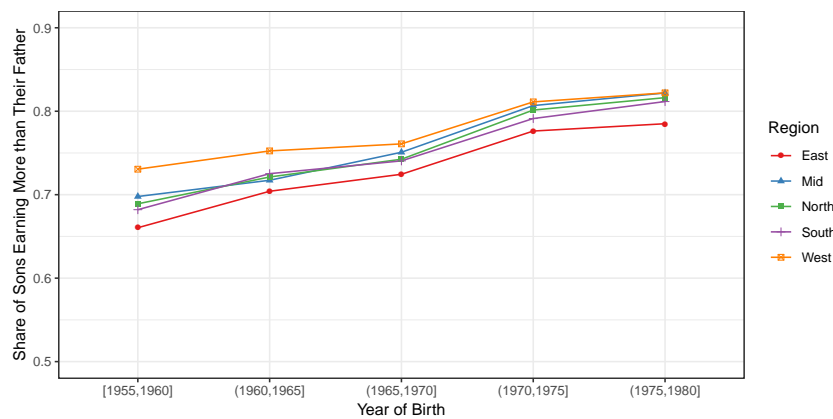
Figure 3.5: Share of Sons Earning More than Their Father- National average
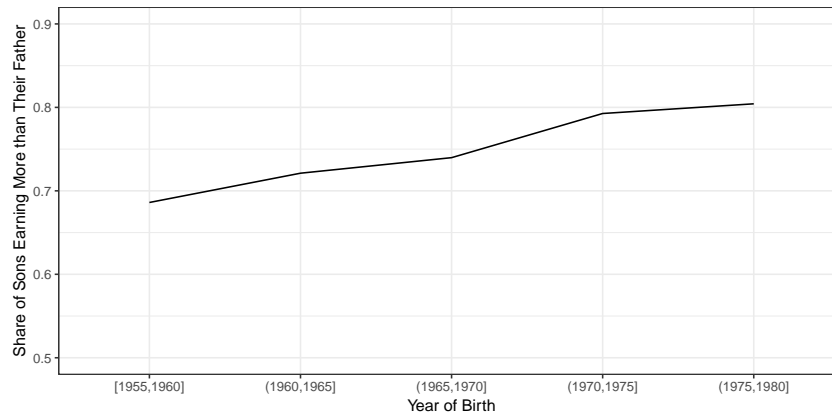
Note: Income is measured as mean earnings at age 34-36 for both sons and fathers

birth cohorts 1955 to a share of 85 percent for cohorts born in 1980. These number are very high in international comparisons. For instance (Chetty, Grusky, Hell, Hendren, Manduca, and Narang 2017) report that 70 percent of sons earn more that their fathers born in the mid 1950s and it is just above 50 percent for sons born around 1980. While we have seen an increasing welfare using this measure in Norway over time, in the US there has been a strong decline. In Appendix A, Figure 3.39 we provide the sensitive of results when earnings are measured at different ages. The main message is the the results at quite stable.

3.3.2   Regional differences in income mobility

We will now turn to regional differences in the rank-rank mobility where we also focus on differences over time. We are going to use two measure of regions, one including 161 commuting areas which is more in line what has been used for instance by Chetty, Hendren, Kline, and Saez (2014), and one where we aggregate the commuting zones into five distinct regions in Norway using the standard used in official statistics; The North, Mid Norway,

West, South and East.

We start by presenting detailed information for the rank rank measure of intergenerational mobility for the cohorts born 1980 to 1985 by the distribution across regions (161 commuting districts) in Figure 3.6, as compared to the US in Chetty, Hendren, Kline, and Saez (2014) for 741 commuting zones.

Figure 3.6: Rank-Rank slope distributions labor markets in Norway and the US
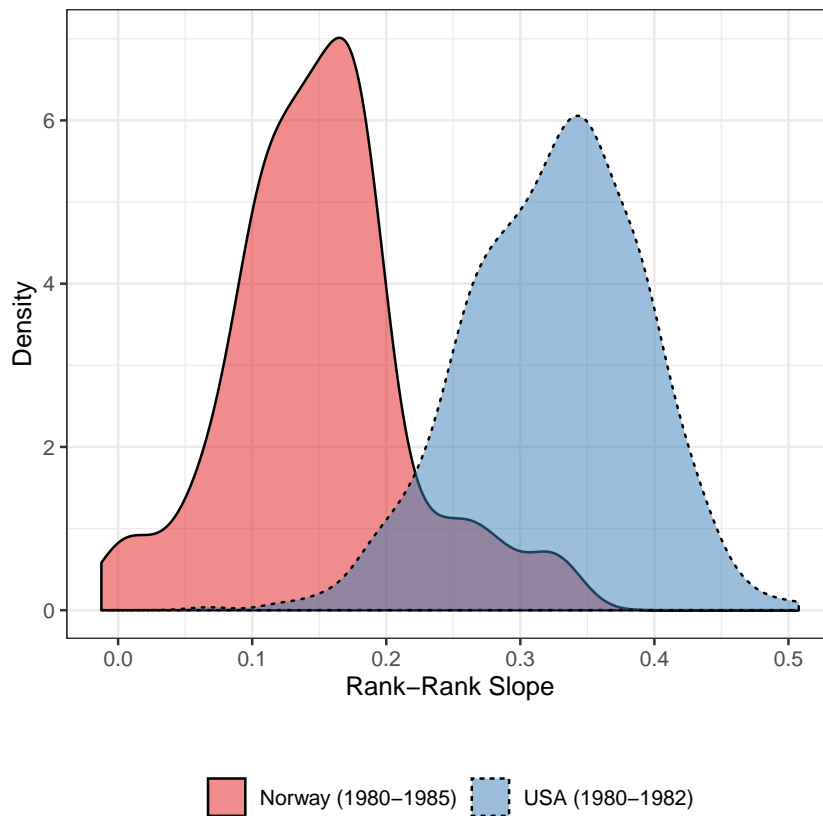
There are a couple stark differences in intergenerational income mobility between the US and Norway as reported in Chetty, Hendren, Kline, and Saez (2014) for commuting districts. First, the distribution is skewed to the right for the US implying the for most regions the income mobility is higher than in Norway. Second, there is a higher variance in the US across regions, from about 0.05 to above 0.5, while it is from about 0 to 0.35 in Norway, indicating that the highest level of persistence is not so far from the US mean.

This being said, there exists substantial variance in mobility across the 160 commuting zones also in Norway. In the Appendix A Figure 3.55, we present the distribution of rank-rank estimates for larger labor markets (around 50 instead of 150).[3]

Next we present the results for first for the rank-rank mobility measure for the five broader regions in Figure 3.7.

Figure 3.7: Rank-Rank Slopes between parent's and child's income ranks - regional averages
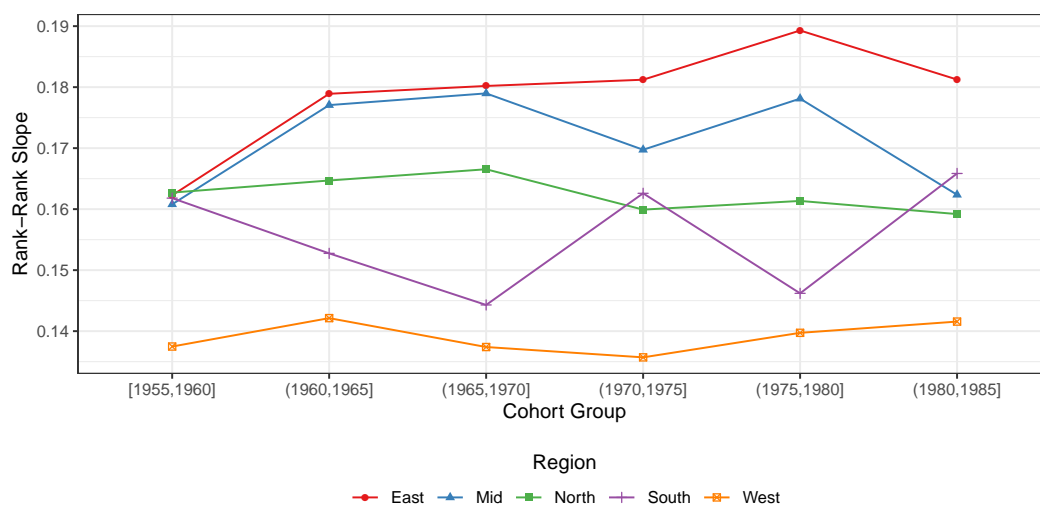
We notice that the income rank-rank mobility is persistently highest in the Western region and lowest in the Eastern region. Focusing on the cohorts born in the early 1980's, a 1 percentile increase in father's income is associated with a 0.18 percentile increase in child's earning percentile in the Eastern Region and only about 0.14 in the Western region. This gap has remained fairly stable over time.

Breakdowns by father's earnings quintile is presented in Figure 3.8.

From this Figure we notice that the difference between the regions, and

---

[3]In Appendix C we also provide a discussion on the role of measure error when calculation the regional mobility measures as well as results form different shrinkage procedures.

Figure 3.8: Rank-rank slopes by father's ranks - regional averages
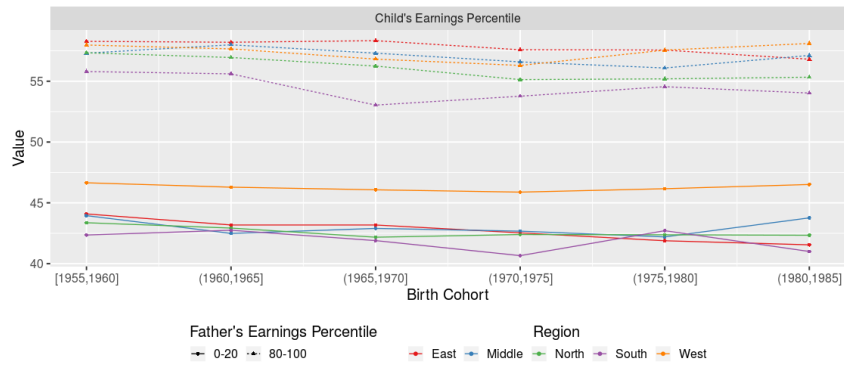
especially the Western region,is driven by higher mobility at the bottom of the parental earnings distribution rather than at the top. For example, in the most recent cohorts, fathers in the bottom earnings quintile born in the Western region are more than 5 percentage points higher in the earnings distribution than children born in the eastern region. In contrast there is almost no gap in mean earnings percentiles of children born to fathers in the top income quintile in the Western and Eastern regions. The picture looks different when comparing the second highest mobility region (South) to the East. Here mobility is higher at both the top and the bottom of the distribution.

In Figure 3.3 we provided the absolute numbers for real Earnings by these quintiles in NOK earnings.

Moving from the 20th to 80th percentile of father's income therefore translates into a 10.2 percentile increase in child income - in the Norwegian case this translates into a 20 percent increase in child income.

We next present the share of children going from the 20 percent lowest percentiles and to the top quintile by region. The results are presented in Figure 3.9

Figure 3.9: Share going from bottom to top quintile - regional averages

Figure 3.10: Share of Sons Earning More than Their Father- regional averages

Note: Income is measured as mean earnings at age 34-36 for both sons and fathers

Again a very similar pictures emerges. In the western region the share going from rags to riches is bar far the highest and it is increasing over time. The Northern region is again at the bottom together with the middle of the country. Interestingly, from cohorts born between 1965 and 1970, we see an increase an increase over time-

The last results on income mobility is the regional version of sons having a higher earnings than their father. We show this in Figure 3.10.

For all regions we have a similar pattern as the national average; the share of sons earning more than their father is higher and increasing over time. Again

we notice that the Western region is above the other regions although some convergence is noticeable from early to mid 1970s. It is a bit remarkable that the East is quite a bit below the other regions for the first cohorts and never converge to the rest of the country.

3.4 Parental Income to Child Education

We now turn to the relationship between father's income percentile and educational attainment, focusing on both national and regional patterns. We then assess different educational margins, from high school completion, to college and post graduate studies or Master degrees.

3.4.1 Years of Education: National and Regional Patterns

We are first going to present results for the correlation between father's income rank and the child educational attainment in terms of years for education for the whole country. In order to obtain this rank-education slope, we estimate the following regression:

$$\text{EducationYears}_i = \alpha + \beta \text{Earnings\_Father}_i + \varepsilon_i, \quad\quad (3.4)$$

where EducationYears_i signifies individual i's years of education, Earnings_Father_i is the earnings percentile in the national income distribution of individual i's father and β is the coefficient of interest.

In figure 3.11 we present the correlation between father's income rank and the child's years of education based on results form estimating equation 3.4.

Figure 3.11: Correlation between father's income rank and child years of education- National Average
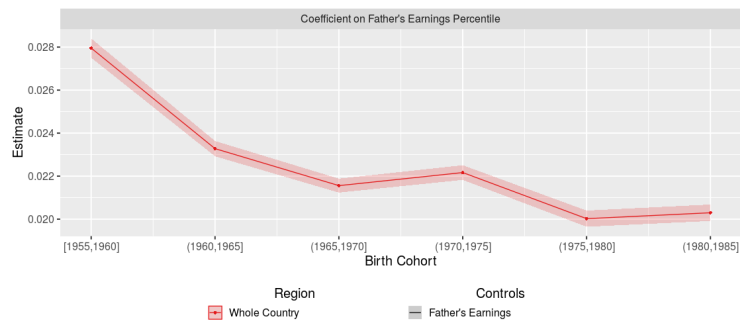
For the most recent 1985 cohort, nationally a 1 percentile increase in parental income is associated with 0.0215 increase in years of schooling. This is quite close to Rothstein (2019)'s estimates for the US: 0.019 for the 1985-86 birth cohort. There is a downward trend over time with an association of around 0.028 for the cohorts born in 1950's. The sharpest reduction seems to have taken place during the late 50's and early 60's. While for the earliest cohorts there was a gap of nearly 2.5 years in the mean schooling of children of fathers in the top and bottom quintiles of the earnings distribution, for cohorts born in the late 70's it was a full year smaller and for the most recent cohorts it is around 1.75 years.

In order to assess whether there is a difference in persistence in educational attainment across the father's income distribution, we estimate a similar measure absolute mobility as for income rank. We predict by cohorts the Years of Education of children for the lowest and the highest quintile of the fathers:

$$\text{EducationYears\_Child}_i = \alpha + \beta \text{RankEarnings\_Father}_i + \varepsilon_i, \qquad (3.5)$$

where $\text{EducationYears\_Child}_i$ signifies child i's Years of education, Earnings_Father_i

is the earnings percentile for the highest and lowest quintile in the national income distribution of individual i's father and β is the coefficient of interest which we use to predict the child's Years of Education. The results are presented in Figure 3.12. We present the results in Figure 3.12.

Figure 3.12: Correlation between father's income rank and child years of education for the highest and the lowest quintiles National Average
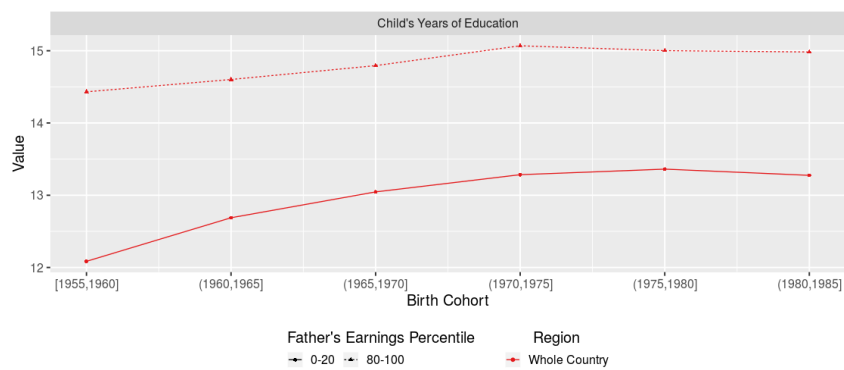
We notice from Figure 3.12 that the level of education is increasing for children from high and low socioeconomic background if we look at the whole country. Moreover, reflection the decrease in persistence noticeable from the previous Figure, we observe a quite considerable degree of convergence over time in that the children from the lowest quintile is closing in on the high SES group. However, still there is a quite big difference.

We now estimate the same equations by the five large regions, and present the results in Figure 3.13 and 3.14.

In line with the finding for income mobility, education mobility is highest in the Western region (where for the 1985 cohort a 1 percentile increase in father's income is associated with a 0.016 increase in years of schooling) and among the lowest in the Eastern region (where this association is 0.025). The time-trend is similar across most of the regions with the North as the exception. It starts out as one of the most mobile regions in the 1950's and

Figure 3.13: Correlation between father's income rank and child's years of education - Regional Averages
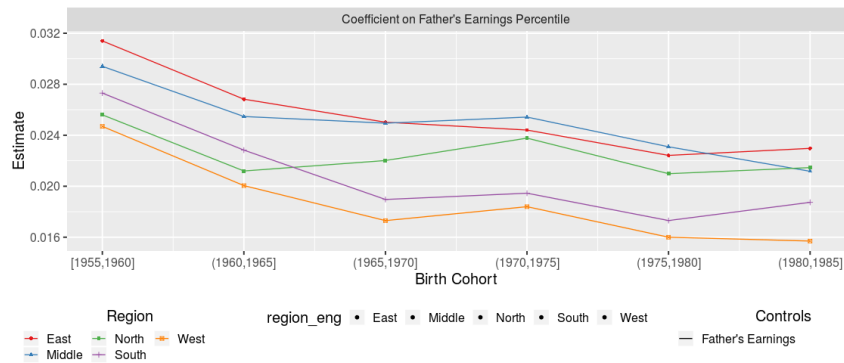
Figure 3.14: Correlation between father's income rank and child's years of education by highest and lowest quintils - Regional Averages
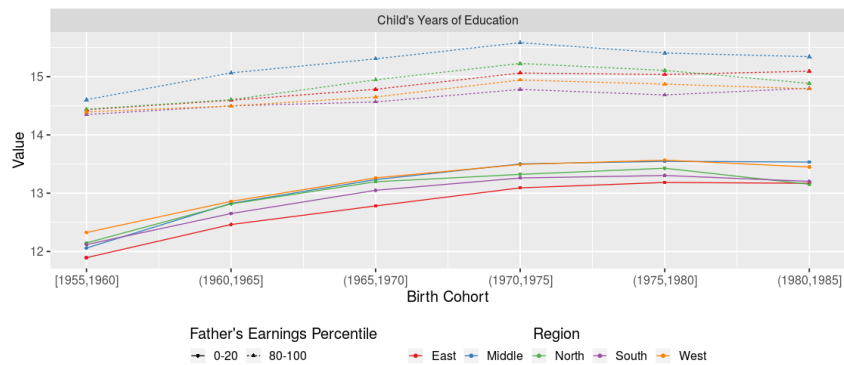
60's, alongside the Western region and ends up as the second least mobile region by the 1980's. This change in rank happens in the decade between the mid 1960's and 70's when mobility decreases in the North while remaining stable or increasing everywhere else.

In addition to the rank-Years of education slope, we include a series of measures showing intergenerational persistence at various levels of education in order to understand at which margins the increse in mobility takes place, at which cohorts and whether there are different patterns over regions.

We obtain these by estimating a series of regression of the following form:

$$\text{EducationLevel}_i = \alpha + \beta \text{Earnings\_Father}_i + \varepsilon_i, \quad (3.6)$$

154

where EducationLevel$_i$ is a dummy variable indicating whether individual i has attained a particular education level or not, Earnings_Father$_i$ is the earnings percentile in the national income distribution of individual i's father and β is the coefficient of interest. We estimate this equation three times, once with completing high school as the outcome, once with completing college completion as the outcome, and once with Postgraduate Degrees or Master Level degrees. This exercise results in the mobility measures we call the rank-high school slope, the rank-short college slope, and the rank-long college slope. We also estimate the predicted levels of education.

3.4.2   High school completion

We present the results for high school completion rates in Figures 3.15 to Figure 3.18. Nationally, among the 1980/85 birth cohorts, about 60% of children of father's in 10th earnings percentile completed high-school, 72% of children of father's in the bottom earnings quintile and 90% in the top. The gap between the top and the bottom quintiles has nearly halved since the cohorts born in the late 1950's.

Figure 3.15: Correlation between father's income rank and proportion of children finishing high school - National Average
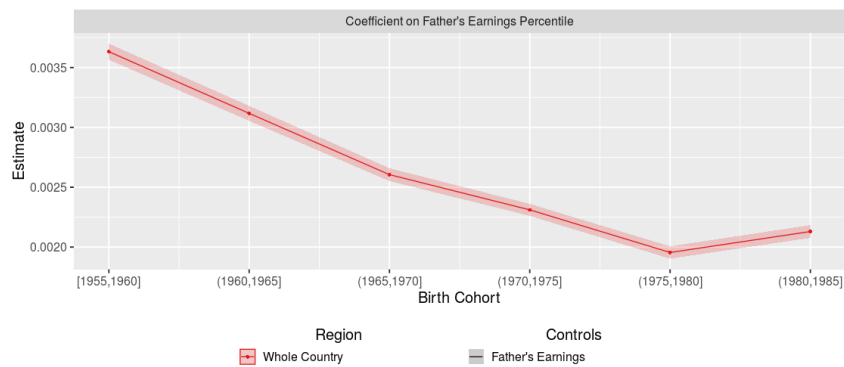
On average, nationally, as with years of education, education mobility mea-

sured by high school completion has increased. For the most recent 1985 cohort, a 10 percentile increase in father's income is associated with a just over 2 percentage point increase in the probability of completing high-school. This is a bit more than half of the strength of the association for cohorts born in the mid 1950's. In Figure 3.16 we present results for the highest and lowest quintile.

Figure 3.16: Proportion of children finishing high school in top and bottom parental earnings quintiles - Natioanl Average
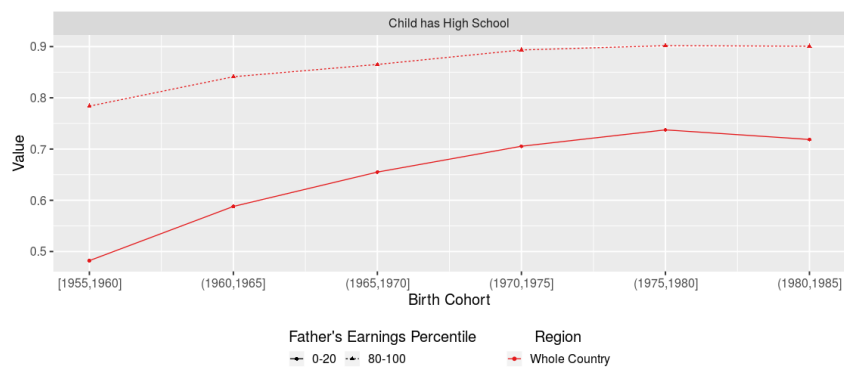
it is striking that there is a huge Socio-economic gradient in completing high school of about 30 percentage points. Half of the children born in the mid 1950s complete high school (both vocational and academic) for children from the lowest parental quintile and 80 percent for the high SES group. There is an increase in completion for both groups and a quite strong convergence. However, even for cohorts born in 1980 there is 20 percentage points in difference.

We then present the results for the relationship between father's income percentile and children's high school completion. the results are presented in Figure 3.17

Again throughout the period the association is weakest in the Western region and strongest in the Eastern region. The Southern region follows the pattern

Figure 3.17: Correlation between father's income rank and proportion of children finishing high school - Regional Averages
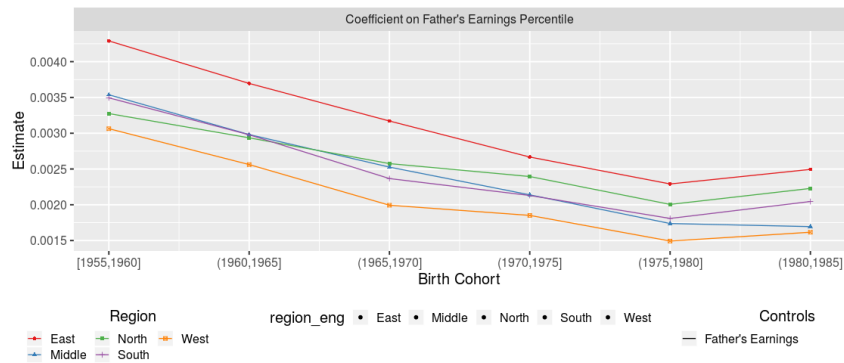
of the Western and the Middle of the East and there has been a least change over time in the Northern region.

We then presented the predicted high school completion rates by high and low SES by father's income in Figure 3.18.

Figure 3.18: Correlation between father's income rank and proportion of children finishing high school in top and bottom income quintiles - Regional Averages
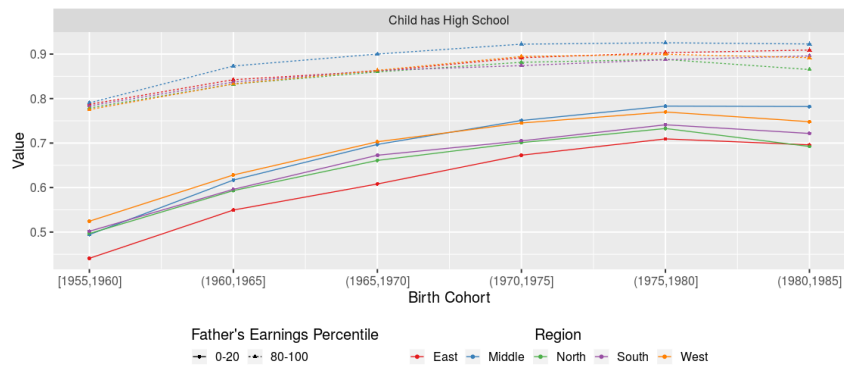
The regional differences are also here present, and there is an overall converge over time, but not really so much convergence across regions.

3.4.3 College completion

We present the correlation between father's income rank and proportion of children completing college at the national level in Figure 3.19.

Figure 3.19: Correlation between father's income rank and proportion of children finishing college - National Average
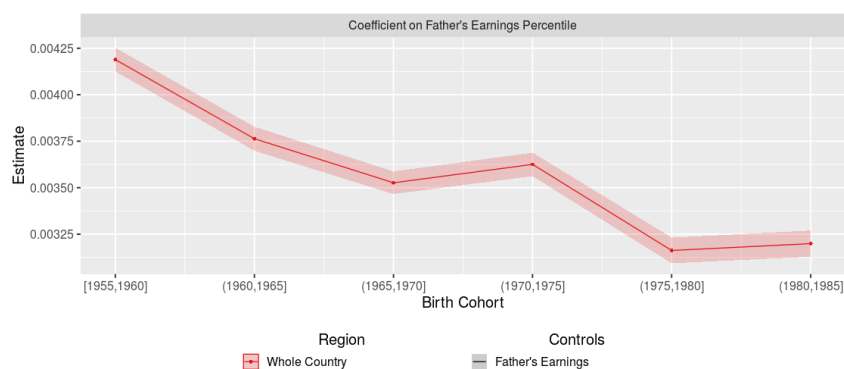
As for High School there is a quite strong increase in mobility over time at the national level. We then show the predicted shares completing college by the highest and lowest quintiles in Figure 3.20.

Figure 3.20: Proportion of children finishing college in top and bottom quintiles - National Average
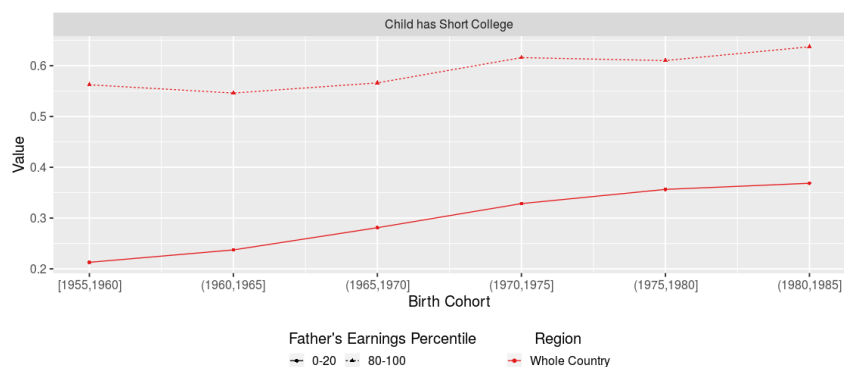
Starting out with a huge difference in college completion for cohorts born in the mid 1950s - 35 percentage points in difference, both groups increase the college attendance, and by the mid 1980s the difference is still big but reduced to below 30 percentage points.

Turning to Regional differences, we report the correlation by cohort and region in Figure 3.21.

Figure 3.21: Correlation between father's income rank and proportion of children finishing college for the highest and lowest quintiles - Regional Averages
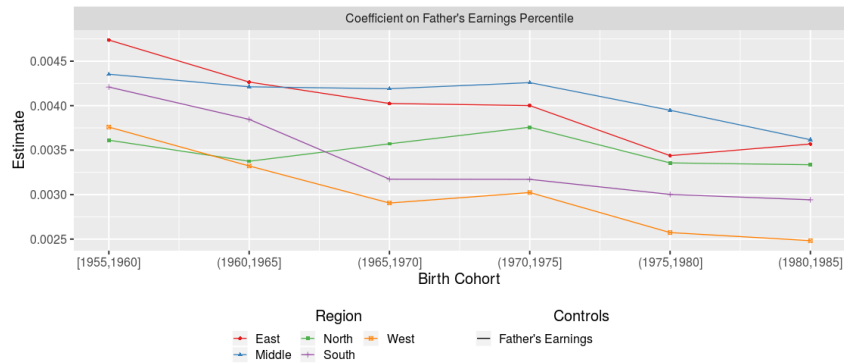
Again we notice that the Western region both start out in cohort from the mid 1950s as the most mobile interestingly with the most Northern region, and have the strongest increase in mobility where college attendance becomes less and less important over time. In the Northern region there is no change in the importance of family background over time in the North, parental income is as important over the the whole period. The Middle and East regions have the strongest persistence although also here parental income becomes less important over time.

Figure 3.22 presents the predicted share by cohort and region for the lower and upper quintiles.

The Western region had the lowest correlation over time, and also the largest change in the gradient, and we see from this figure that all the change is in the lowest quintile. We see convergece over time among the highest and lowest SES, but not really and convergence across regions amog the top and the bottom.

Figure 3.22: Proportion of children finishing college in top and bottom quintiles- Regional Averages
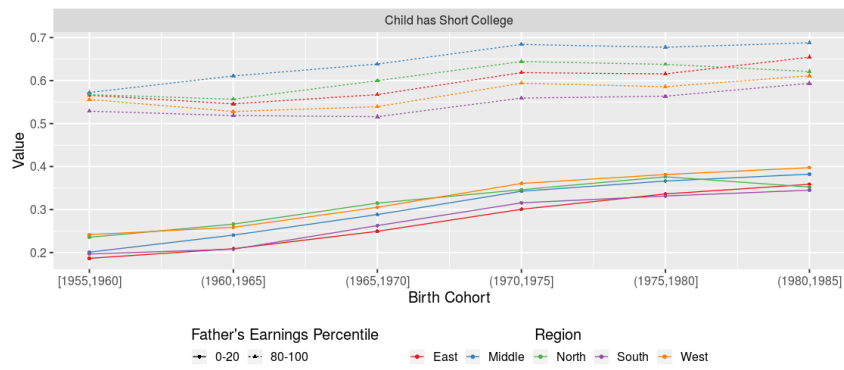
3.4.4 Master degree completion

A master degree or post graduate degree is important in the European countries including Norway. We know from other studies including our own work that graduates with a master degree end up in higher income ranks since they land high level jobs both in the public and private sector (Bütikofer, Risa, and Salvanes 2019). In Figure 3.23 we present again the pattern in the correlation between father income rank and the proportion obtaining a master degree (or a PhD).

Figure 3.23: Correlation between father's income rank and proportion of children finishing master's - National Average
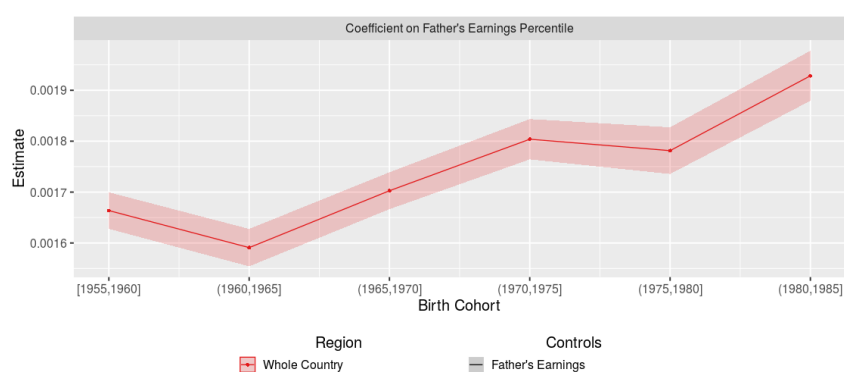
The direction in the mobility trend for completion of a Master degree is the opposite of that for years of schooling and high school completion, declining over time. For example, for those born in the early 1960's a 10 percentile increase in father's earnings percentile resulted in a 1.6 percentage point increase in the probability of completing long-college compared to nearly 2 percentage points for those born in the early 1980's. From Figure 3.24 we see that there has been little/no change in the difference in Master degree completion rates between children of fathers in the top and bottom income quintiles. This result is in stark contrast to the result for the other margins we have looked at, and is also in line with the results when assessing the role

161

of parental education and and the child's educational attainment; access to a master degree is strongly dependent also on parental education with really no change over time (Bütikofer, Risa, and Salvanes 2019).

Figure 3.24: Proportion of children finishing master's in top and bottom parental income quintiles - National Average
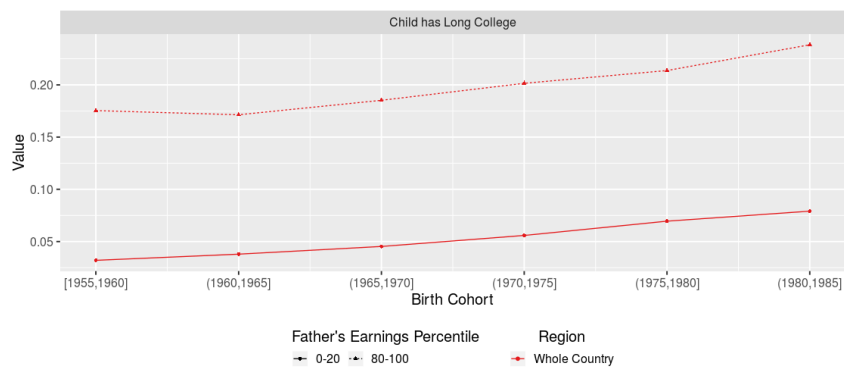
In Figure 3.25 we present the correlation by region.

Figure 3.25: Correlation between father's income rank and proportion of children finishing master's - Regional Averages
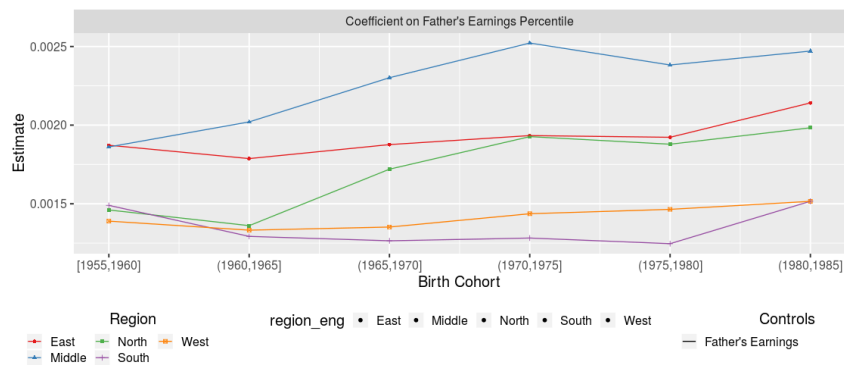
We notice that there are quite big differences across regions of the importance of parental earnings as well as a big difference in patterns over time. Interestingly the Middle region has the strongest correlation and it becomes less mobile over time. The South and the West has the lowest correlation and basically with no change over time. In Figure 3.26 we present the proportions completing a master degree by region and across cohorts by highest and lowest parental income quintile.

The figure shows very minor differences for the low SES, but a quite strong difference across regions among the top 20 percent. The Middle regions which had the highest correlation naturally also has the highest master degree attendance followed by the Northern region. This implies that in these two regions we have the biggest difference in the importance of family income in determining a child's master degree. We notice that the west and the South have the lowest difference. Important to note is also that, if anything, there has been a divergence over time in a master level attendance across regions.

We have now analyzed the intergenerational mobility in income for difference measures over time and regions, as well as a child's education and parental income. In Appendix B we present similar results for childhood outcomes such as Cognitive ability, Child BMI, and Child obesity,for boys measured at the age of 18. As before the West is persistently among the most mobile regions (along with South) by this measure - middle region is the least. The North was similar to the West in the 50s and early 60's, but is at the level of the Eastern region by the 1970's.

3.5 Returns to Schooling

Both intergenerational income mobility and the importance of SES as measured by parental income on educational attainment, may be explained both by family, characteristics of the region (and time period) when growing up (access to good schools, health care etc), but also to the characteristics of the labor market. Differences in returns to education over time and region is an important characteristics of the labor market determining educational choice. We will now present some simple measures of returns to education

over cohort and across region, and for different educational margins. The measure of returns to schooling comes from estimating a very simple Mincer equation:

$$ln(\texttt{earnings})_i = \alpha + \beta \cdot \texttt{education}_i + \gamma \cdot \texttt{polynomial in age}_i + u_i, \quad (3.7)$$

separately for each of the regions and each of the cohort groups. The earnings measure is the log of average annual earnings when the children are 30-35 years old. `Education` is years of education and `polynomial in age` is second order polynomial in age. In some of the equations we control for a cognitive capacity measure we have for all boys when they are being evaluated for military service at age 18-19. The βs then give us a measure of returns to schooling for each cohort group in each region.

First, we present a Figure 3.28 showing the estimates across the commuting areas of returns to education for the different regions and cohort groups when controlling for cognitive ability in the mincer equation. Similarly, Figure 3.27 shows estimates when not controlling for cognitive ability.

Figure 3.26: Proportion of children finishing long college in top and bottom parental income quintiles
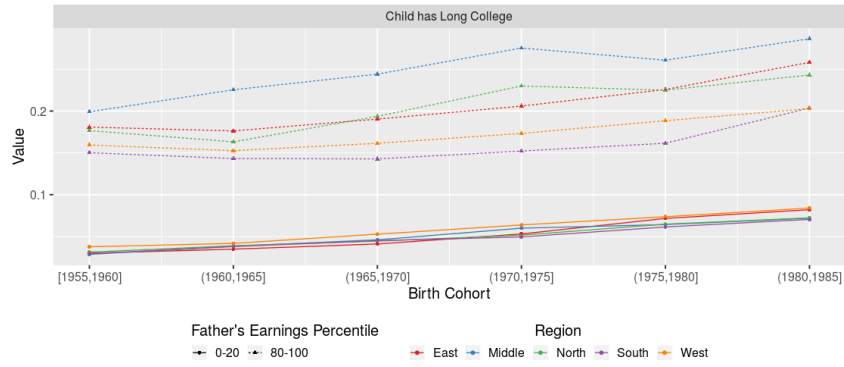
Figure 3.27: Returns to Education - not controlling for cognitive ability
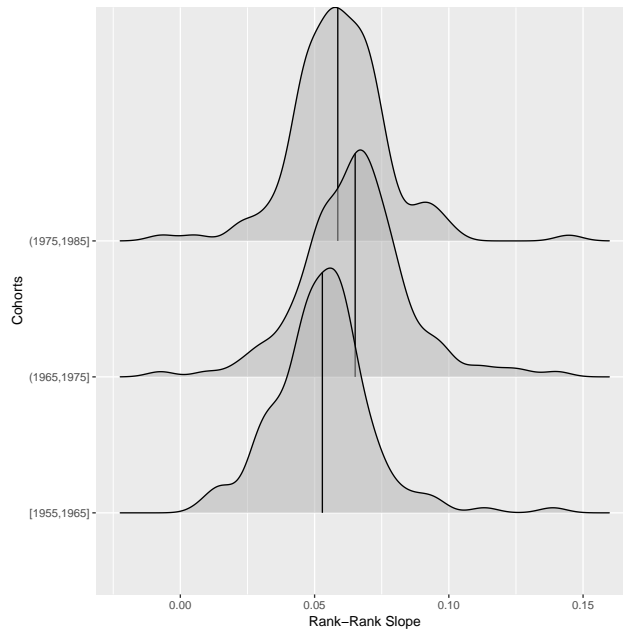
Figure 3.28: Returns to Education - controlling for cognitive ability
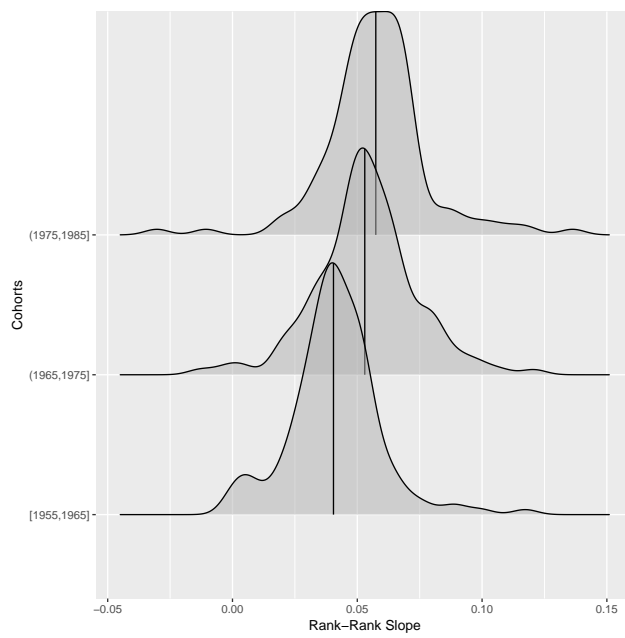
Figure 3.29: Returns to Education by region when the sample are adults
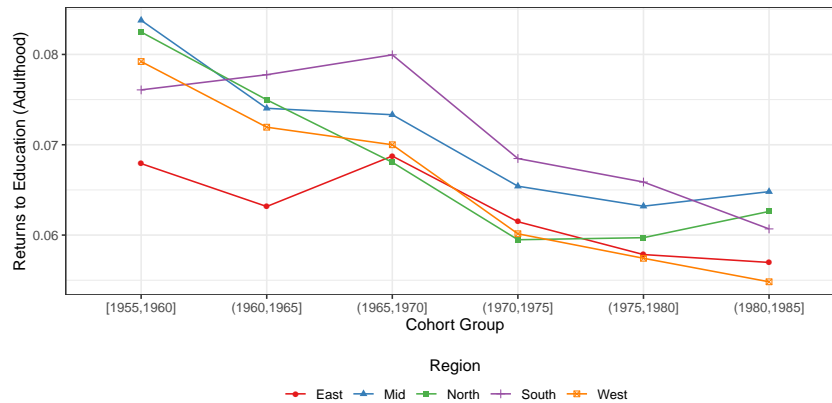
We notice that there is a quite large spread in returns to education across the commuting areas both with and without ability controls. Returns to a year of education varies from close to zero up to above 10 percent. We do notice a change over time with a small and insignificant tendency to declining returns especially when ability is controlled for.

Next we present results over time for the five regions in Norway, from the North to the South. In Figure ?? we present returns to one more year of education only with age control by region over time for fathers of children born from the mid 1950s onwards. While there is a spread between 6.5 percent and 8.5 percent at the start of the period, this difference has narrowed down to 5.5 and 6.5 at the end of the period. There is clearly a convergence over time, however, the East appears to do worse over the whole period. The West has the strongest decline, with a decline even after the mid seventies children cohort. These numbers are for the adult at as measured in their mid 30s, and is thus measured from the mid 1960s onwards. The same results are presented for their children, also measured in their mid 30s, and thus as measured from 90s onwards. The returns are overall higher than for the parents generation, but also with a slight decline in the beginning, but

Figure 3.30: Returns to Education by region when the sample are children
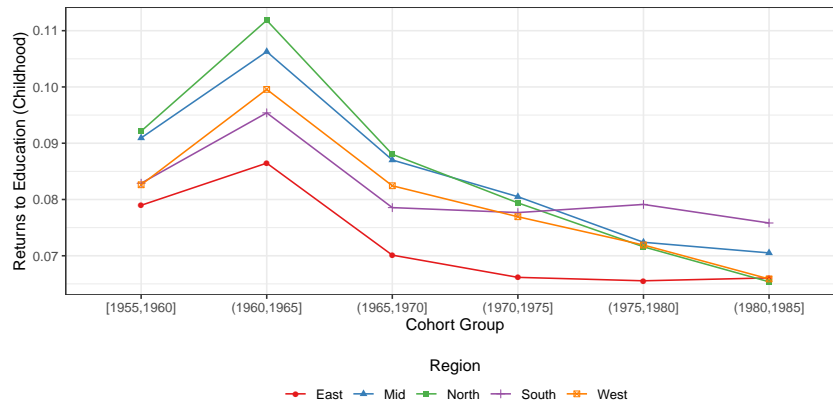
Figure 3.31: Returns to College Degree
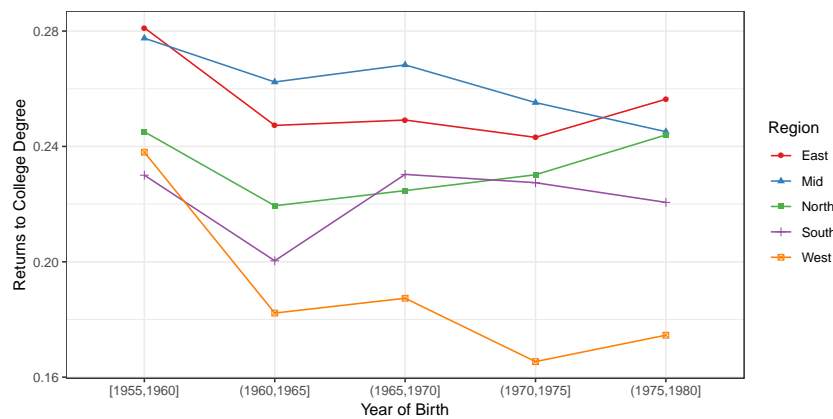
flattening out from cohorts born from the mid 1960s with the exception of the West. Still the East is underperforming as compared to the other regions. In Figure ??, we present the returns to college as compared to all degrees below for the children.

The returns to college is overall quite high, between 25 and 30 percent, and quite stable for all regions. The exception is the West, we there is a considerable drop in returns to college over time. The West is the center of the oil industry, and a strong negative effect on returns to higher education has been noted in the literature for the oil industry also in Norway (ref butiokofer, dalla zuanna, salvanes). In Figure ?? we present the results

168

Figure 3.32: Returns to Master's Degree
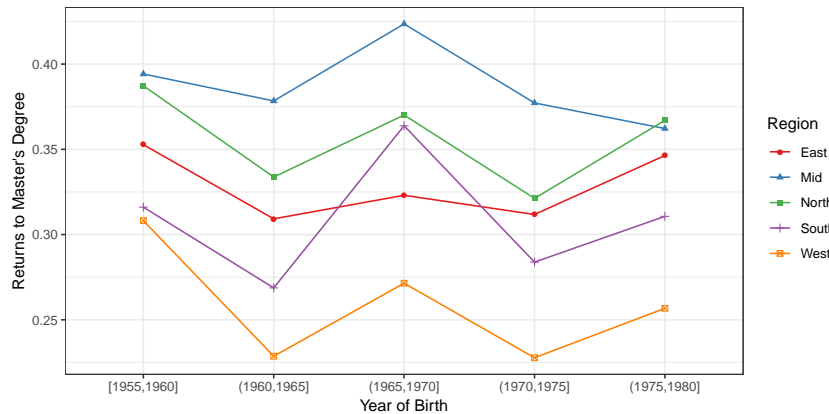
for a master degree also as compared to high school and below. A similar pattern is true here. The returns is high and stable with quite distinct regional differences. The West has the lowest returns, and the Middle region has the highest. The Middle regions has by far the most important engineering universities in Norway, and is the main hub for high tech industry.

3.6 Empirical Approach

We have provided detailed descriptive information on regional patterns over time for intergenerational income mobility measures in terms of both relative measures and for different parts of the income distribution. Moreover, we have provided information on human capital formation and returns to human capital over regions and time and for different education margins and for different parts of the parental income distribution.

We are now interested in understanding which characteristics of the regional labor markets that are important in explaining these patterns, and in particular the role of human capital formations and the role of the labor market. In order to do this, we will employ machine learning techniques in order to

169

handle the challenges of high dimensional data. The first challenge posed by high dimensional data is the fact that it is not clear ex-ante which of the many potential variables belong in the analysis. A second and related challenge is the fact that including many variables increases the variance of the estimates, making statistical inference more difficult. This second point is especially relevant in this context because the number of regions and years of data available naturally limit the sample size in the analysis. Moreover, several of the variables, such as measures of income inequality within regions at different points in time, are correlated, further increasing the variance of the estimates.

3.6.1 Machine Learning

The machine learning algorithm we use in the analysis, Glmnet[4], handles these challenges by imposing a penalty on the estimated coefficients, thereby shrinking them towards zero. For linear regression, the objective function takes the form of (Hastie and Qian 2014):

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \left[(1 - \alpha) ||\beta||_2^2 / 2 + \alpha ||\beta||_1 \right], \quad (3.8)$$

where the first term is the standard minimization of the squared residuals and the second term is the additional penalty term. In Equation 3.8, there are two parameters in addition to the estimated coefficients, namely α and λ. Alpha takes values between zero and one and determines what type of penalty gets applied to the coefficients. If $\alpha = 1$, the penalty is the same as in a lasso regression, whereas $\alpha = 0$ implies a ridge regression. Consequently, values between zero and one imply a mix of the two. Moreover, λ controls

---

[4]https://cran.r-project.org/package=glmnet

the overall size of the penalty and takes non-negative values. If $\lambda = 0$, the objective function becomes minimization of the squared residuals, resulting in ordinary least squares, whereas $\lambda > 0$ yields regression with shrinkage. What values of α and λ are most appropriate in a given setting is an empirical question. We used 10-fold cross validation to compare different model specifications and determine the optimal values for α and λ.[5] Because the elastic net algorithm penalizes the size of the coefficients, we standardize the x-variables to be mean zero and have a standard deviation of one in order to ensure that they are on the same scale so that the effect sizes are comparable.

---

[5]In order to reduce the variance of the performance metrics, we repeated the cross validation procedure ten times

3.7 Empirical Results

3.7.1 Variable Description

Table 3.1: Variable Description

Inequality	Share with Low Earnings (Tenth Percentile or Lower)
	Share with Low Education (Mandatory or Less)
	Average Years of Education
	Income Gini Coefficient
	Returns to Education
Earnings	Tenth Earnings Percentile (By Gender and Education Level)
	Median Earnings (By Gender and Education Level)
	90th Earnings Percentile (By Gender and Education Level)
	Average Earnings Percentile
Industry	Regional Industry Employment Shares
Education Mobility	Rank-High School Slope
	Rank-Short College Slope
	Rank Long College Slope
Region	Region Fixed Effects
Cohort	Cohort Fixed Effects

3.7.2 Machine Learning Results rank-rank

Figure 3.33 and 3.34 show the average r-squared and mean absolute value in the hold-out datasets during cross-validation at optimal α and λ values as chosen by cross-validation. Each column on the x-axis in the figure shows a model specification, increasing in complexity from left to right. There are several things to note in the figures. First, the increase in predictive performance between columns one and two indicates that both inequality measures, such as the share with low education, and the earnings measures, such as 90th percentile earnings by industry, are predictive of rank-rank slopes in a region. Moreover, the difference in predictive performance between columns two and three indicate that regional characteristics both in

childhood and adulthood, carry predictive information. The figures also indicate that industry employment shares carry predictive information. When it comes to the centrality and county controls added in columns six and seven, they do not appear to improve predictive performance. However, focusing on column eight yields two interesting observations. First, there is a marked increase in predictive performance for the elastic net model. Second, there is if anything a decrease for OLS, indicating that OLS starts to overfit when estimating the region fixed effects. Lastly, column ten shows that the measures of intergenerational mobility in education improve the predictive performance of the models, indicating that they carry information beyond what the other variables contain.

Figure 3.33: Model Comparison - R-squared
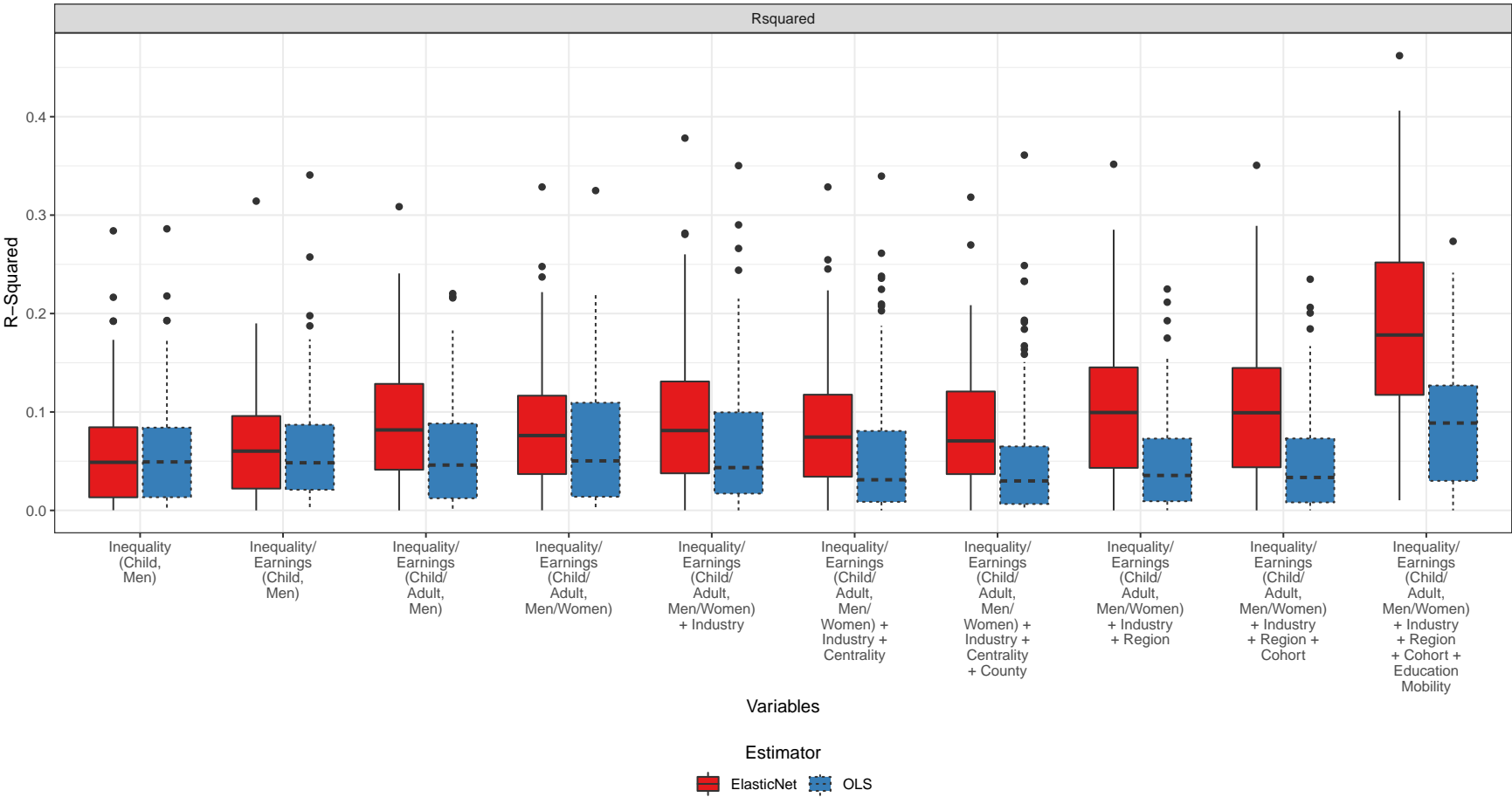
Figure 3.34: Model Comparison - Mean Absolute Error
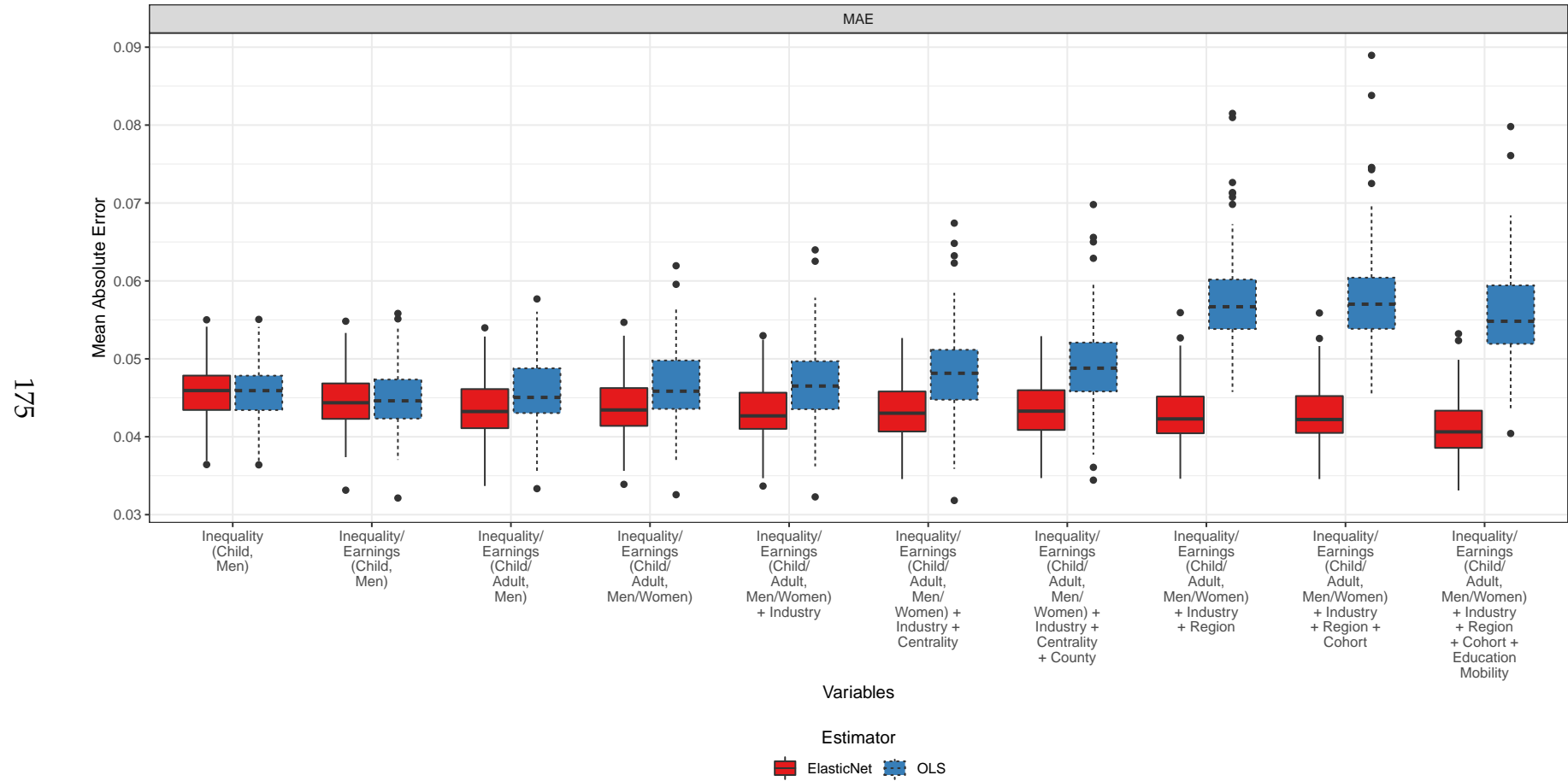
Because Figure 3.33 and 3.34 reveal that the model in column ten best describes how rank-rank slopes vary between regions, we focus on estimates from this model when proceeding to investigate what predicts rank-rank earnings mobility. The selected tuning parameters for the model in column ten are $\alpha = 0.1538462$ and $\lambda = 0.017$. The fact that cross-validation selects $\alpha > 0$ is informative because it indicates that it is optimal for the coefficients on some of the variables to be shrunk to precisely zero, which is something that only the lasso penalty achieves. This observation implies that some of the included variables do not help predict rank-rank slopes and are omitted from the model by the elastic net algorithm.

Given that the variables in the analysis are standardized, one way of determining which variables are most predictive of rank-rank mobility is to look at the absolute values of the estimated coefficients. Table 3.2 shows statistics for the ten variables, excluding region fixed effects, with the largest estimated effect sizes in our preferred elastic net model. The column labeled "Elastic Net Coef" contains the estimated coefficients from the preferred elastic net model. In order to get a sense of the variability of the elastic net estimates, we also show the average, the 2.5 percentile and the 98.5 percentile of the distribution resulting from 999 bootstrap replications of estimating the elastic net model with the optimal tuning parameter values in column three four and five. For comparison, we also include the point estimates and standard deviations from OLS in columns six and seven. For reference, we also include the sample means and standard deviations of the variables in column eight and nine. For clarity, we also plot the elastic net coefficients and the 2.5 and 98.5 percentiles of their bootstrap distributions in Figure 3.35.[6]

The first thing to note in Figure 3.35 is that the 2.5 and 98.5 percentiles of the

---

[6]We plot the whole distributions of coefficients in Figure 3.51 in the appendix

Table 3.2: Most Predictive Variables - Rank-Rank

Variable	Elastic Net Coef	Boot Coef	Lower Boot CI	Upper Boot CI	OLS Coef	OLS SE	Sample Mean	Sample SD
empshare_oil_gas_adult	-0.0024	-0.0022	-0.0046	0.0000	0.4186	2.1670	0.0085	0.0116
meanperc_mandatory_male_adult	-0.0017	-0.0013	-0.0045	0.0000	-0.0078	0.0313	0.4539	0.0638
empshare_construct_adult	-0.0011	-0.0010	-0.0039	0.0000	0.9136	4.7103	0.0797	0.0252
meanperc_college_female_child	0.0008	0.0008	0.0000	0.0037	0.0236	0.0175	0.4447	0.0514
lowearn_male_adult	0.0012	0.0008	0.0000	0.0037	0.0018	0.0145	0.0599	0.0261
meanperc_college_female_adult	0.0020	0.0016	0.0000	0.0049	-0.0300	0.0250	0.4571	0.0403
empshare_rtl_hotel_cens	0.0024	0.0013	0.0000	0.0042	0.1342	0.1689	0.1101	0.0417
lowearn_male_child	0.0036	0.0035	0.0000	0.0072	0.0133	0.0114	0.0484	0.0258
rank_longcollege_slope	0.0070	0.0063	0.0017	0.0106	0.0155	0.0032	0.0014	0.0007
rank_highschool_slope	0.0142	0.0133	0.0090	0.0178	0.0226	0.0041	0.0023	0.0011

bootstrap distributions include zero for all but two of the variables. While the bootstrap distributions should be used for statistical inference in the manner which is common for ordinary least squares, this observation does tell us that the estimated coefficients are zero in at least 2.5 percent of the bootstrap replications. To be conservative, we do not put any emphasis on the variables for which there are zero effect estimates in at least 2.5 percent of the bootstrap replications. Following this approach, we conclude that the only variables that are sufficiently predictive of rank-rank slopes to warrant discussion are the rank-long college slope and the rank-high school slope. Consequently, the primary characteristic of the regions with high rank-rank mobility is that they have a weaker link between parental earnings and children's educational outcomes.

We conclude this part of the analysis by exploring the regional variation in rank-rank slopes as predicted by our preferred model. We do this by plotting the estimated coefficients on the region dummies on a map in Figure 3.52 in the appendix. The magnitude of these coefficients shows how much each region differs on average from the predicted rank-rank slopes based on the other characteristics included in the model. The fact that the majority of the regions in the figure have a similar color tells us that the actual rank-rank

Figure 3.35: Top Ten Elastic Net Coefficients
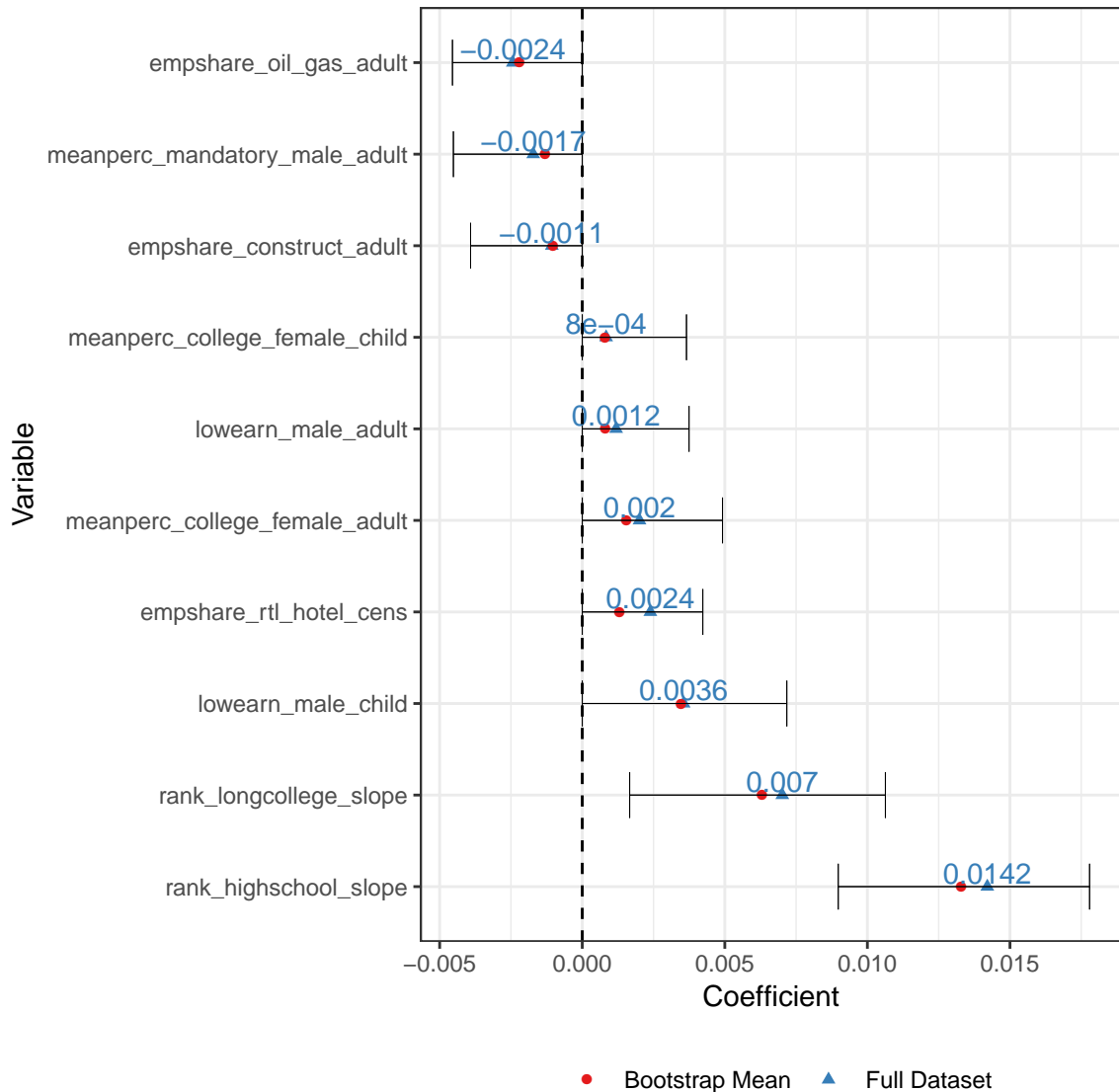
slopes in the regions are on average reasonably close to the ones predicted by the other variables in the model. A possible interpretation of this is that the scope for achieving additional predictive accuracy is limited. Moreover, part of the reason why there seem to be few strong predictors of rank-rank mobility might be the compressed earnings distribution in Norway. Specifically, it is possible that and that movements in the middle of the earnings distribution are somewhat more random. While movements in the middle of the earnings distribution may be somewhat random, movements in the

tails of the distribution may be more systematic. To investigate this, in the next section, we analyze a different outcome, namely the share of individuals whose parents are in the bottom 20 earnings percentiles end up in the top 20 earnings percentiles themselves.

3.7.3 Empirical Results: Absolute Intergenerational Mobility on characteristics of Local Labor Markets in Child and Adulthood

Machine Learning Results for absolute measures

Figures 3.36 and 3.37 show predictive performance when predicting the share going from the bottom 20 to the top 20 earnings percentiles, measured by r-squared and mean absolute error on the hold out datasets from cross-validation. The patterns in the figure are broadly the same as seen for rank-rank in Figure 3.33 and 3.34. Specifically, we see that regional characteristics in both childhood and adulthood are predictive, that employment structure also appears to matter, that geographical controls capture additional information and that the measures of intergenerational mobility in education also carry predictive information. Moreover, we see that also in the case of the bottom to top mobility, there is a gradual divergence in performance between OLS and elastic net, indicating that OLS starts to overfit as we add predictors. It is also worth discussing the selected tuning parameters which are $\alpha = 0$ and $\lambda = 0.043$ in the case of bottom to top mobility. The fact that cross-validation now shows an optimal $\alpha = 0$ indicates that there is less need for setting coefficients precisely equal to zero. This further implies that more of the variables included in the analysis are predictive of bottom to top mobility than was the case with rank-rank mobility.

As with rank-rank mobility, we now proceed to explore the key predictors of bottom to top mobility by tabulating the ten variables with the largest

Figure 3.36: Model Comparison - R-squared
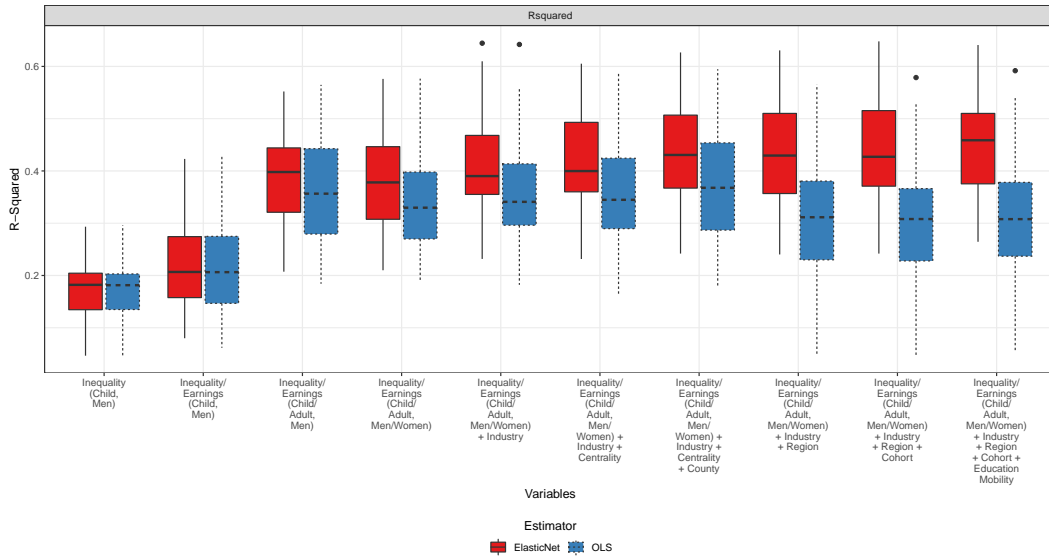
Figure 3.37: Model Comparison - Mean Absolute Error
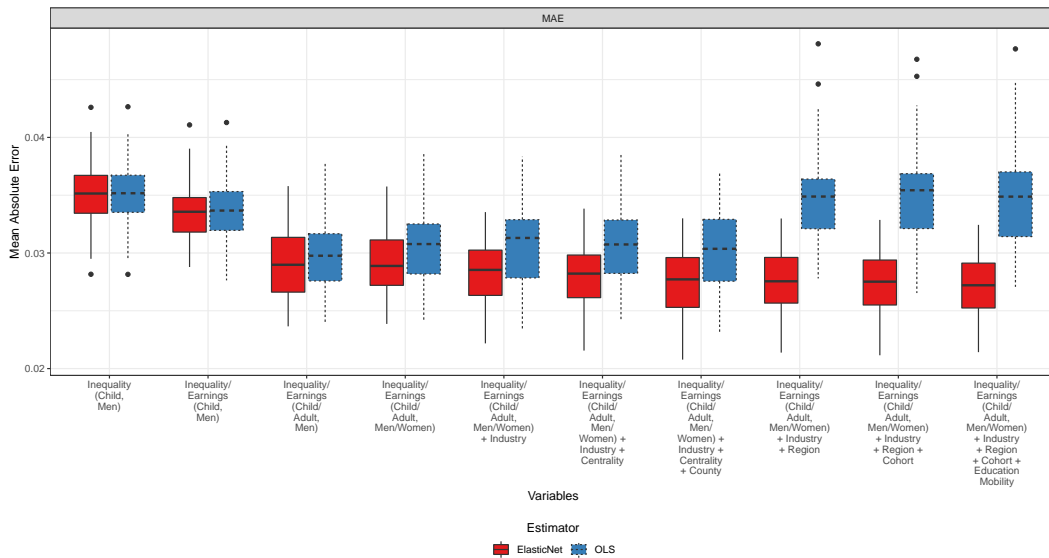
estimated effect sizes in Table 3.3 and plotting them in Figure 3.38.[7]

As seen in the table and figure, the tails of the distributions of bootstrapped coefficients for all but one variable lie far from zero. This observation indicates that there more variables that consistently predict bottom to top mobility in the bootstrap replications. Moreover, we see that four variables

---

[7]We plot the whole distributions of coefficients in Figure 3.53 in the appendix

Table 3.3: Most Predictive Variables - Bottom to top

Variable	Elastic Net Coef	Boot Coef	Lower Boot CI	Upper Boot CI	OLS Coef	OLS SE	Sample Mean	Sample SD
rank_longcollege_slope	-0.0027	-0.0026	-0.0040	-0.0011	-0.0072	0.0020	0.0014	0.0007
rank_highschool_slope	-0.0026	-0.0026	-0.0040	-0.0012	-0.0064	0.0026	0.0023	0.0011
empshare_health_social_adult	-0.0024	-0.0022	-0.0033	-0.0011	0.7402	6.6840	0.2476	0.0568
lowearn_male_child	-0.0021	-0.0020	-0.0028	-0.0011	-0.0049	0.0072	0.0484	0.0258
empshare_electr_wtr_cens	0.0019	0.0018	-0.0001	0.0036	-0.0245	0.0392	0.0117	0.0154
meanperc_male_adult	0.0020	0.0020	0.0016	0.0024	0.0508	0.0487	0.5658	0.0639
p90_mandatory_male_adult	0.0021	0.0020	0.0014	0.0027	-0.0058	0.0134	453.2475	91.9240
meanperc_mandatory_male_adult	0.0025	0.0025	0.0018	0.0032	0.0187	0.0197	0.4539	0.0638
empshare_oil_gas_adult	0.0028	0.0028	0.0018	0.0038	0.1622	1.3663	0.0085	0.0116
empshare_transport_adult	0.0030	0.0029	0.0017	0.0041	0.3413	2.9361	0.0748	0.0249

Figure 3.38: Top Ten Elastic Net Coefficients
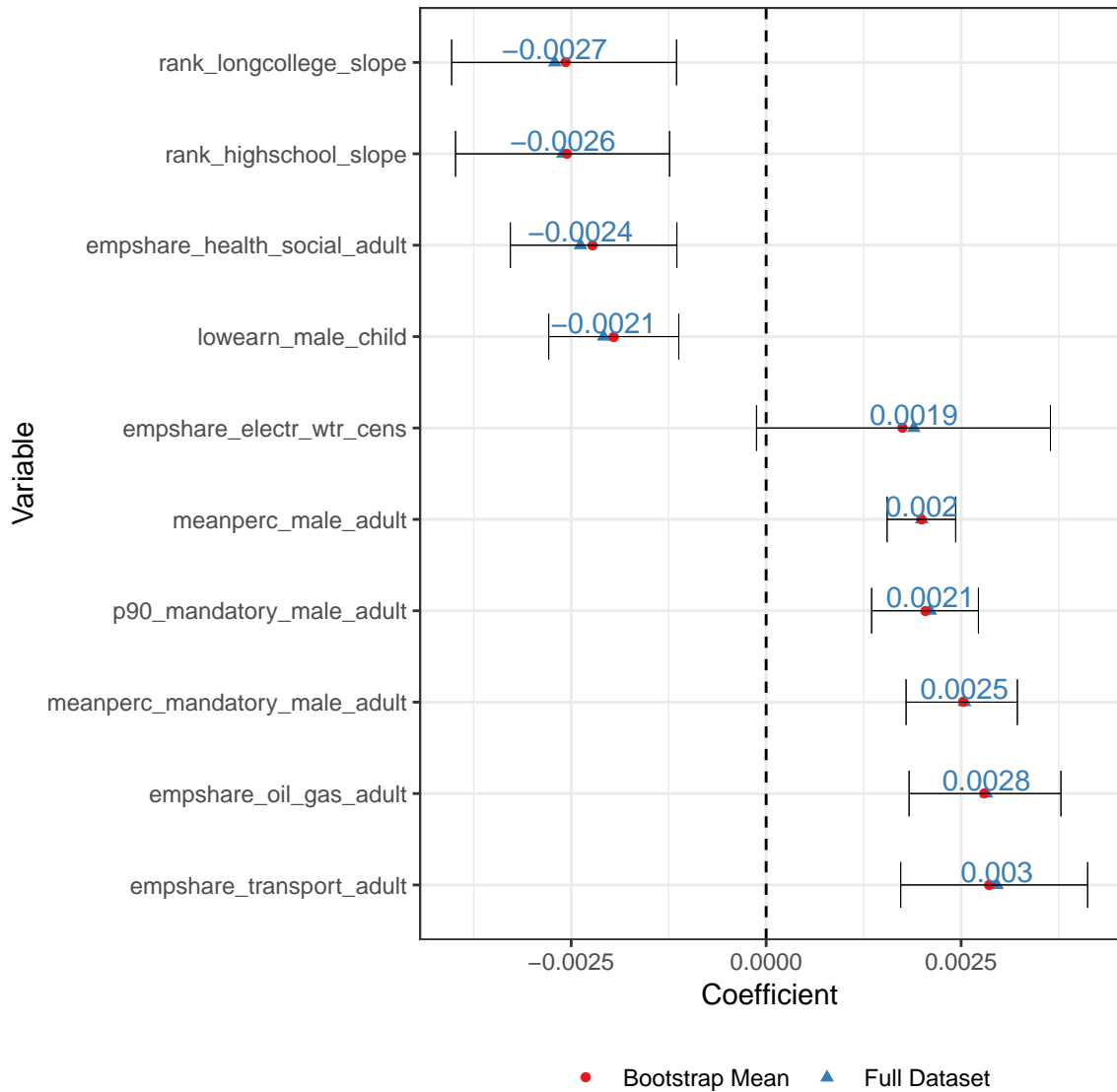
consistently predict lower bottom to top mobility and five variables that predict higher bottom to top mobility.

The four variables negatively associated with bottom to top mobility are the rank-long college slope, the rank-high school slope, the regions' employment share in health and social industries and the share of men with low earnings in the region during childhood. The five variables that are all related to the labor market in the region when the individuals in the sample are adults. Three of them have to do with the earnings of men in the region, and in particular men with no more than mandatory schooling. The last two have to do with industry structure, and indicate that regions with large employment shares in the oil and gas industry and the transportation industry tend to have high mobility from the bottom to the top of the earnings distribution.

3.8   Concluding Remarks

In this paper we first document the patterns across local labor markets and regions in Norway of cohorts born from the mid 1950s to the mid 1980s and around the age of 30 today, by leveraging an extensive and population wide register data connecting families and regions over time along outcome dimensions as income and education. We also document the development across regions and time along different margins of mobility, using different measures of intergenerational mobility, such as a relative measure as rank-rank which comprise both upward and downward mobility, the predicted rank for growing up poor and growing up rich (20 bottom and top parental percentile), and going from rags to riches, share of children going from bottom 20 percentile parental rank to the top 20 percentile, as well the share of sons with higher earnings than their fathers in order to measure welfare

improvements. We are in particular interested in the role of human capital investments by parental background over time and across regions, and the role of the labor market opportunities for different socio-ecnomic groups across regions and over time, and we estimate the relationship between parental income rank and child educational attainment. We document this also for the bottom and top parental income percentile, and for different educational margins, high school, college, and postgraudate or master programs.

We find that the rank-rank correlation of gross earnings has remained remarkably stable since the 1950s. It is about half as large as the association estimated by Chetty, Hendren, Kline, and Saez (2014) for the US. The time-trend is similar across the regions, however, income mobility has been persistently highest in the Western and Southern regions and lowest in the Eastern region. Breakdowns by father's earnings quintile show that while the mobility gap between the Western and Eastern regions is driven by differences in mobility at the bottom of father's income distribution, there are differences at both the bottom and the top when comparing the Southern and Eastern regions. Interstingly, we find that around 12-13 percent of children goes from the lowest to the highest quintile as a national average, and increasing over time. There are large and quite stable differences across regions, again with the West as the most mobile from the bottom to the top, and the North and the Middle region with the lowest mobility. Moreover, we also find that for all regions, the share of sons making more money than their fathers, is high and increasing over time, again with the Western region with a higher share than the rest. We see some convergence over time, except for the East region. At the national level, the share of sons with higher earnings than their fathers starts out in the 50s cohort at around 70 percent similar to the US, but in stead of steadily going down as in the US, it increases to about 85 percent in

Norway. In contrast to the stability in the overall income mobility over time - although changes over time and across regions especially with mobility at the bottom, education mobility measured by strength of association between father's income and child years of schooling has increased, especially sharply in the late 1950's and early 1960's - for the 1985 cohort it is very similar to Rothstein (2019)'s estimates for the US. While for the earliest cohorts there was a gap of nearly 2.5 years in the mean schooling of children of fathers in the top and bottom quintiles of the earnings distribution, for cohorts born in the late 70's it was a full year smaller; further, the difference in the proportion of children completing high-school between the top and the bottom quintiles has nearly halved since the late 1950's.As with income mobility, education mobility is persistently highest in the Western region and lowest in the Eastern region. The time-trend is similar across most of the regions with the North as the exception. It starts out as one of the most mobile regions in the 1950's and 60's, alongside the Western region and ends up as the second least mobile region by the 1980's.We see the reverse pattern when looking at post graduate studies or master degree completion. Here mobility has, if anything, gone down over time and there has been little/no change in the difference in master degree completion rates between children of fathers in the top and bottom income quintiles. There is substantial regional variation with big differences in mobility trends for post graduate completion in Northern and Middle regions relative to the rest. The difference in trends is driven by differential trends at the top of parental income distribution rather than bottom. In order to assess the role of the labor market for human capital investments, we estimate returns to years of education in a simple Mincer equation framework. First, we notice that returns differ across regions, from close to zero to above ten percent, and with a tendency to an overall decline

over time both for fathers and children. However, estimating returns to higher education, the returns to college vs high school, and returns to a master degree vs high school, the returns are stable over this time period. Again there are big differences across regions indicating an important role for the labor market in human capital investment. Especially the Western region has low returns to the post graduate level.

The plan for a next step, is to use a coherent panel data framework for analysing the relationship between income mobility across time and regions, and the what the main mechanisms are.

3.9 Appendix A

3.9.1 Measurement ages

Figure 3.39: Sensitivity of Income Estimates with Respect to Measurement Age
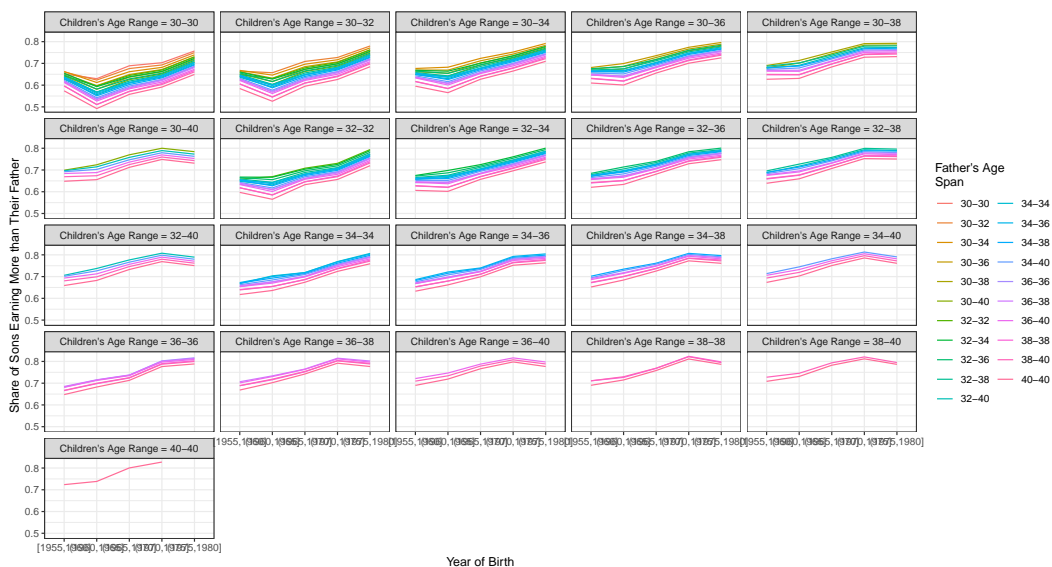
3.9.2 Measurement error

An important aspect of which regions to analyze is the sample size we are able to get within each region-cohort cell. As figure 3.40 and 3.41 show, the standard errors in the least populated regions are fairly large. The question is where to draw the line for excluding regions due to small sample sizes. Chetty, Hendren, Kline, and Saez (2014) chose 250 observations. Figure 3.42 and 3.43 show that the estimated rank-rank slopes start looking more "reasonable" somewhere around 125 observations, but we will need to discuss how much noise we are willing to accept. Clearly, if we put very strict restrictions on the sample size within each region we will have too few regions to analyze. This trade-off becomes more challenging for mobility measures which require us to only analyze parts of the population, such as transitioning from the bottom to top quintile.

Figure 3.40: Standard Errors Versus Sample Size
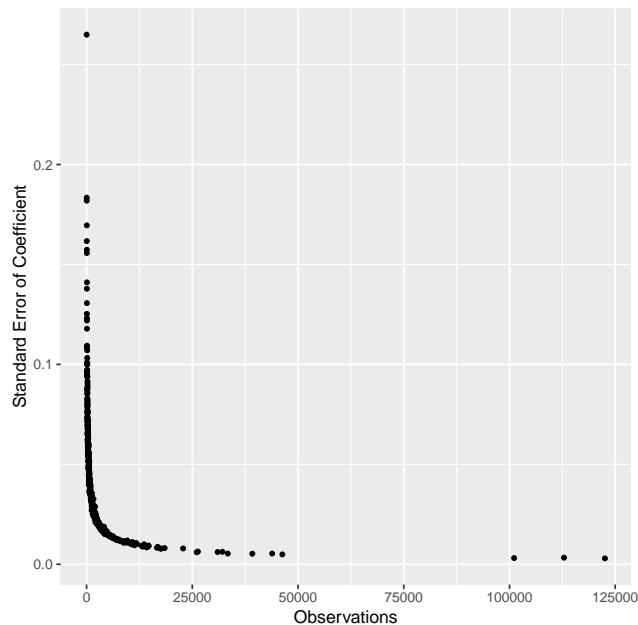
Figure 3.41: Standard Errors Versus Sample Size
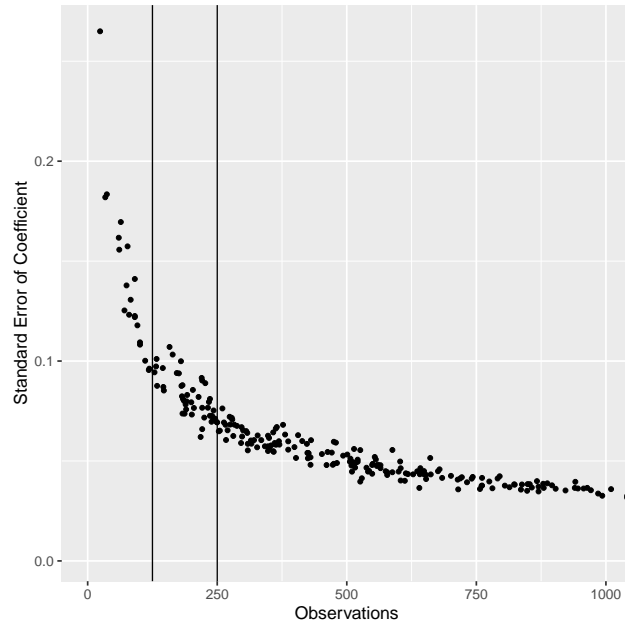
Figure 3.42: Rank-Rank Slopes Versus Sample Size
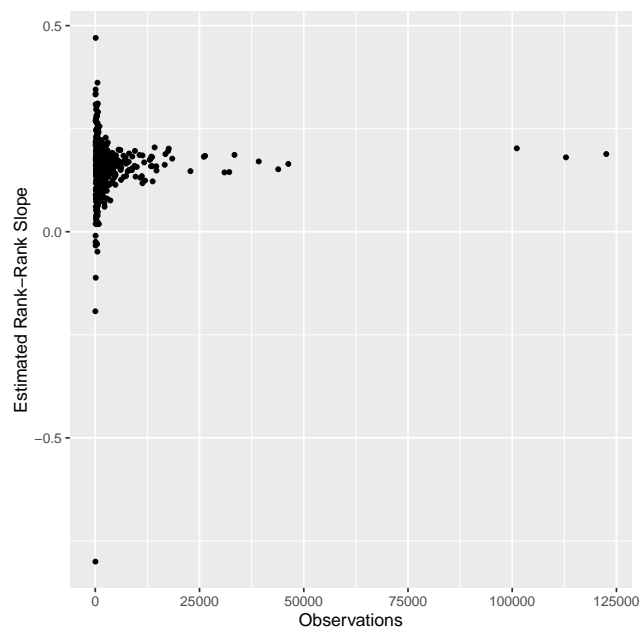
Figure 3.43: Rank-Rank Slopes Versus Sample Size
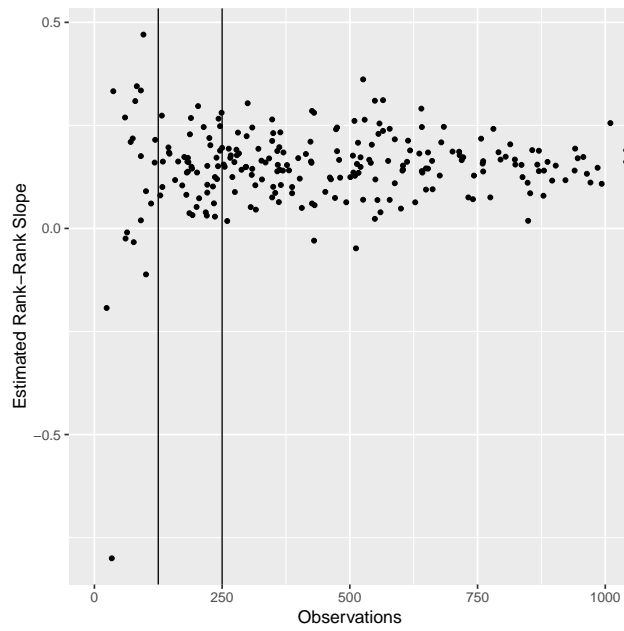
Shrinkage

The current shrinkage procedure comes from estimating multilevel models with two levels for each of the cohort groups. One way of writing up a main equations for the models is as follows:

$$y_{irt} = \beta_{0t} + \beta_{1t} \cdot x_{irt} + u_{0rt} + u_{1rt} \cdot x_{irt} + \varepsilon_{irt}$$

Where we assume:

$$u_{0rt} \sim N(0, \sigma_{0t}^2)$$

$$u_{1rt} \sim N(0, \sigma_{1t}^2)$$

This means that we are assuming that there is an overall countrywide intercept (β_{0t}) and rank-rank slope (β_{1t}) for each cohort group (t). Moreover, we assume that the region specific deviations from these coefficients are normally distributed with a mean of zero and variances σ_{0t}^2 and σ_{1t}^2. The countrywide parameters β_{0t} and β_{1t} are obtained through separate rank-rank regressions for each of the cohort groups (these are called the level one equations). The distributions for the coefficients u_{rt} and u_{1rt} are then estimated using restricted maximum likelihood in the level two equations. Finally, predicted values for u_{rt} and u_{1rt} are obtained using the empirical Bayes method to combine the raw estimates we would obtain from separate regressions for each region and the estimated distributions (we can think of the estimated distributions as priors). This is illustrated in Figure 3.45, 3.46 and 3.47, where we see that the shrunk estimates are a combination of the raw estimates and the empirical/estimated prior distribution. This implies that the estimates are shrunk towards the countrywide rank-rank slope. Moreover, the amount

of shrinkage depends on the noisiness in regional estimates, meaning that regions with noisy estimates are shrunk to a larger extent than regions with precise estimates. This can be seen in Figure 3.48, where we see that regions with more observations lie closer to the plotted "no shrinkage" lines. Figure 3.49 and 3.50 show maps of the shrunk rank-rank slopes.

Figure 3.44: Raw Versus Shrunk Rank-Rank Slopes
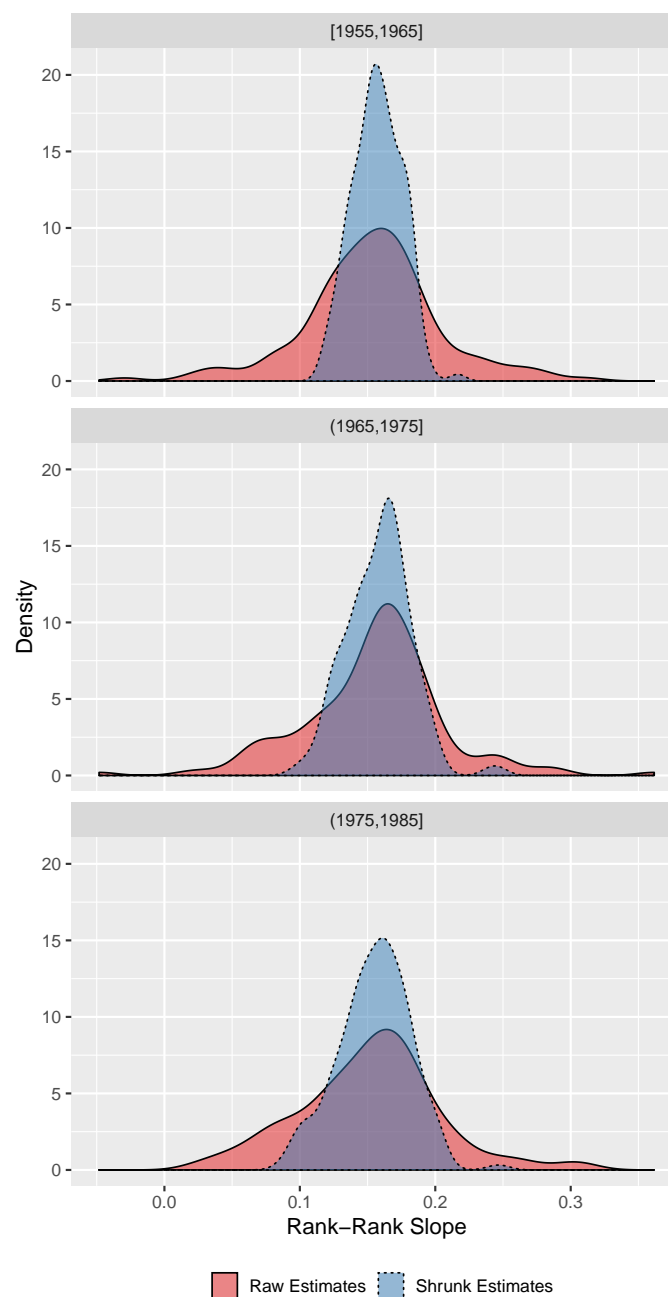
Figure 3.45: Raw Versus Shrunk Rank-Rank Slopes
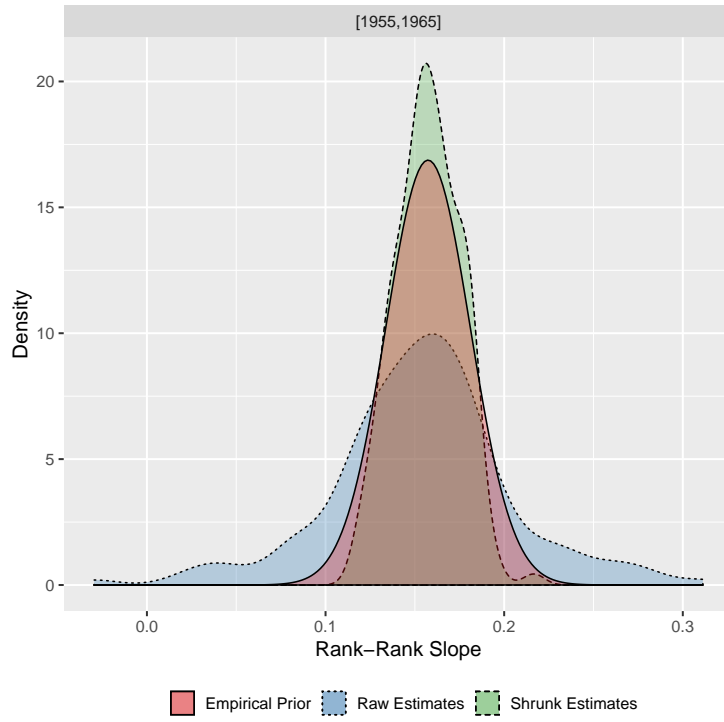
Figure 3.46: Raw Versus Shrunk Rank-Rank Slopes
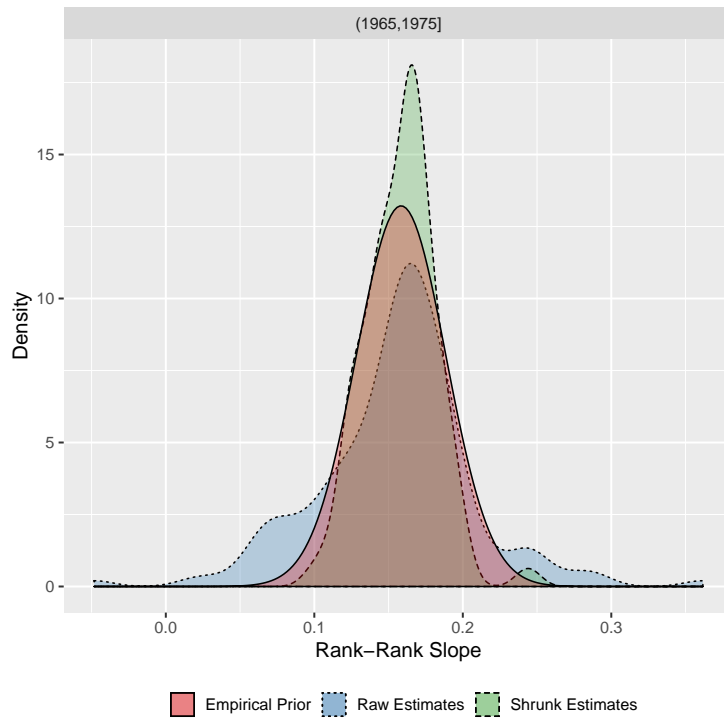
Figure 3.47: Raw Versus Shrunk Rank-Rank Slopes
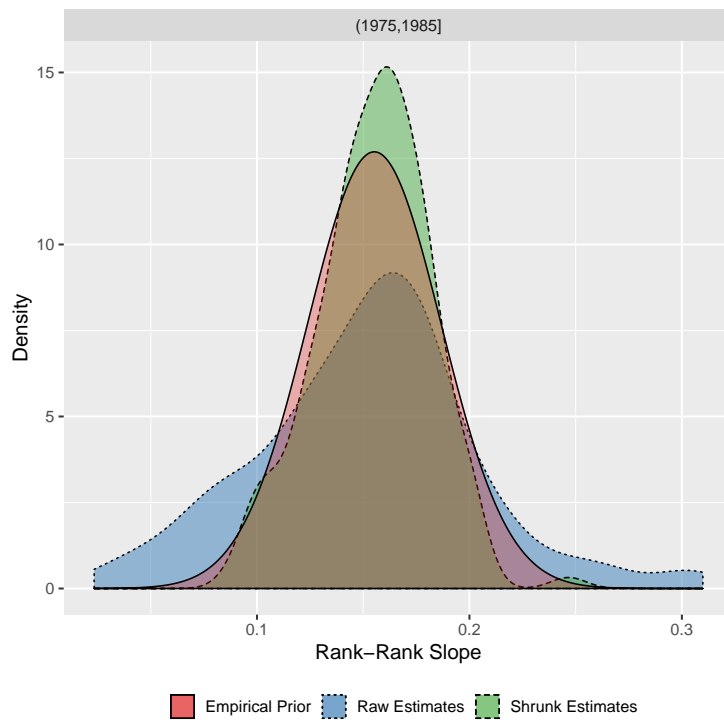
Figure 3.48: Raw Versus Shrunk Rank-Rank Slopes
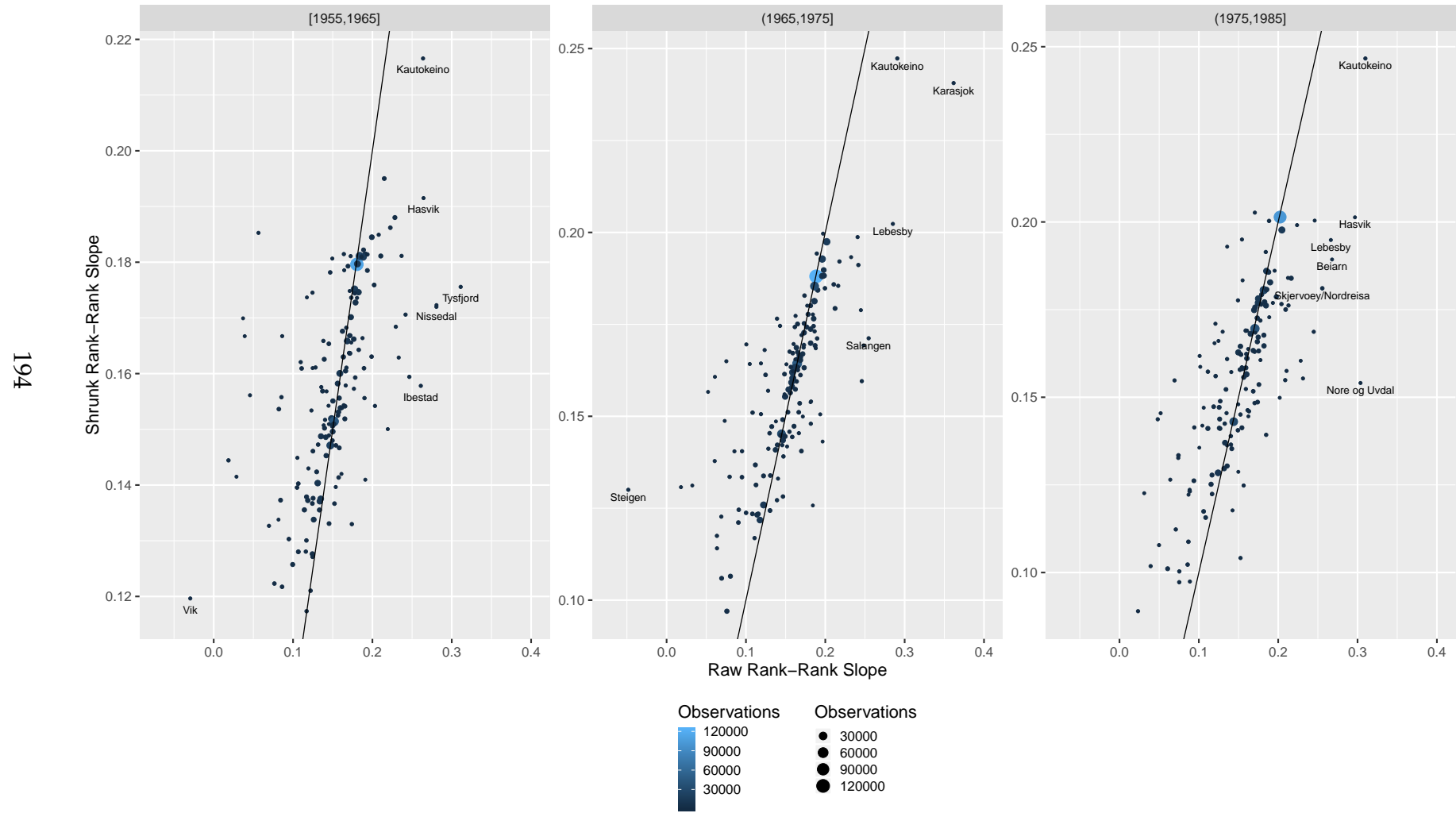
Figure 3.49: Map of Shrunk Rank-Rank Slopes
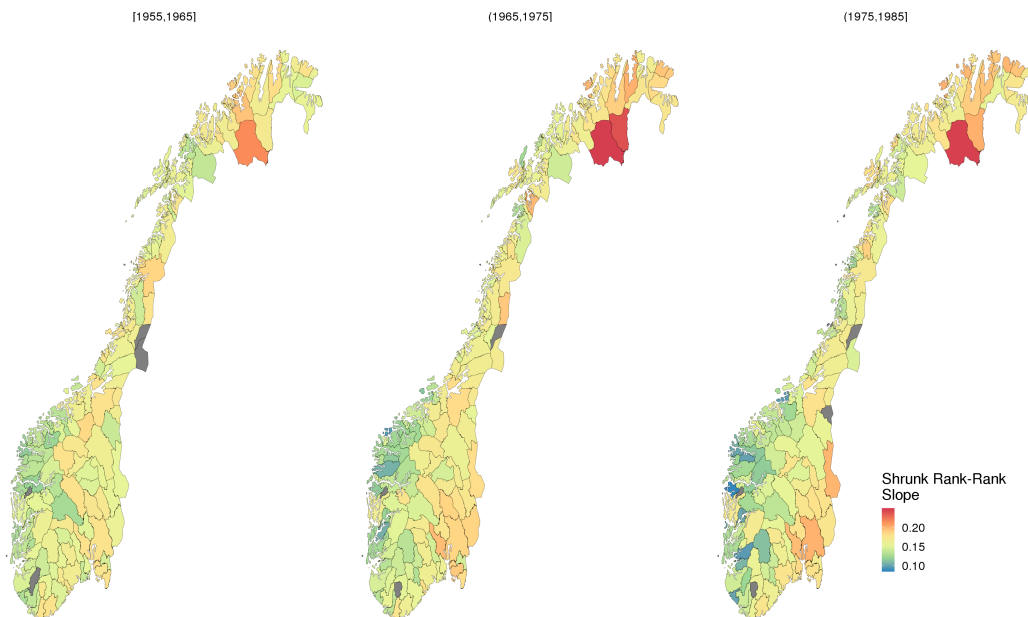
Figure 3.50: Map of Shrunk Rank-Rank Slopes
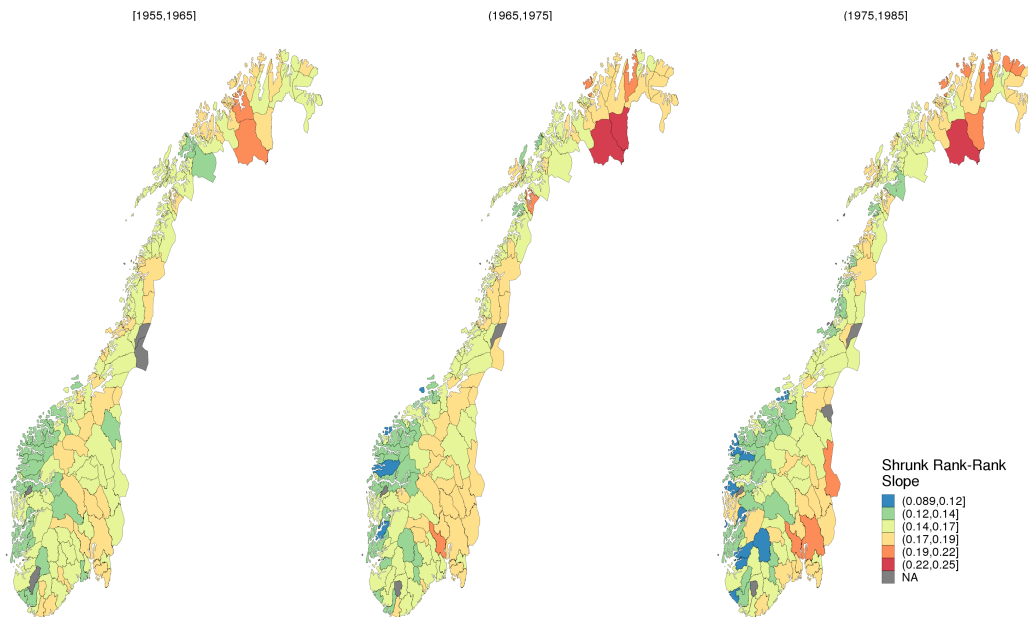
Appendix

3.10   Appendix B: Parental income to other child's outcomes

3.10.1   Child cognitive ability

IQ is measured is scaled using the Stanine method of scaling the test scores on a nine-point standard scale with a mean of 5 and a standard deviation of 2.

Overall, there is a decline in strength of association between father's earning percentile and child IQ. While in the 60's and early 70's a 10 percentile increase in father's earnings was associated with around a 9% of a standard deviation increase in IQ score, by mid-1980's this increase was closer to 5%.

There is a persistent gap in mobility measured in this way between the Western and Eastern regions. Throughout the period a 10 percentile point increase in father's earnings is associated with around a 0.04 of a point (or 2% of a standard deviation) larger increase in IQ in the Eastern region compared to the Western region.

3.10.2   Child BMI

3.10.3   Child Obesity

Figure 3.51: Top Ten Elastic Net Coefficients - Rank-Rank
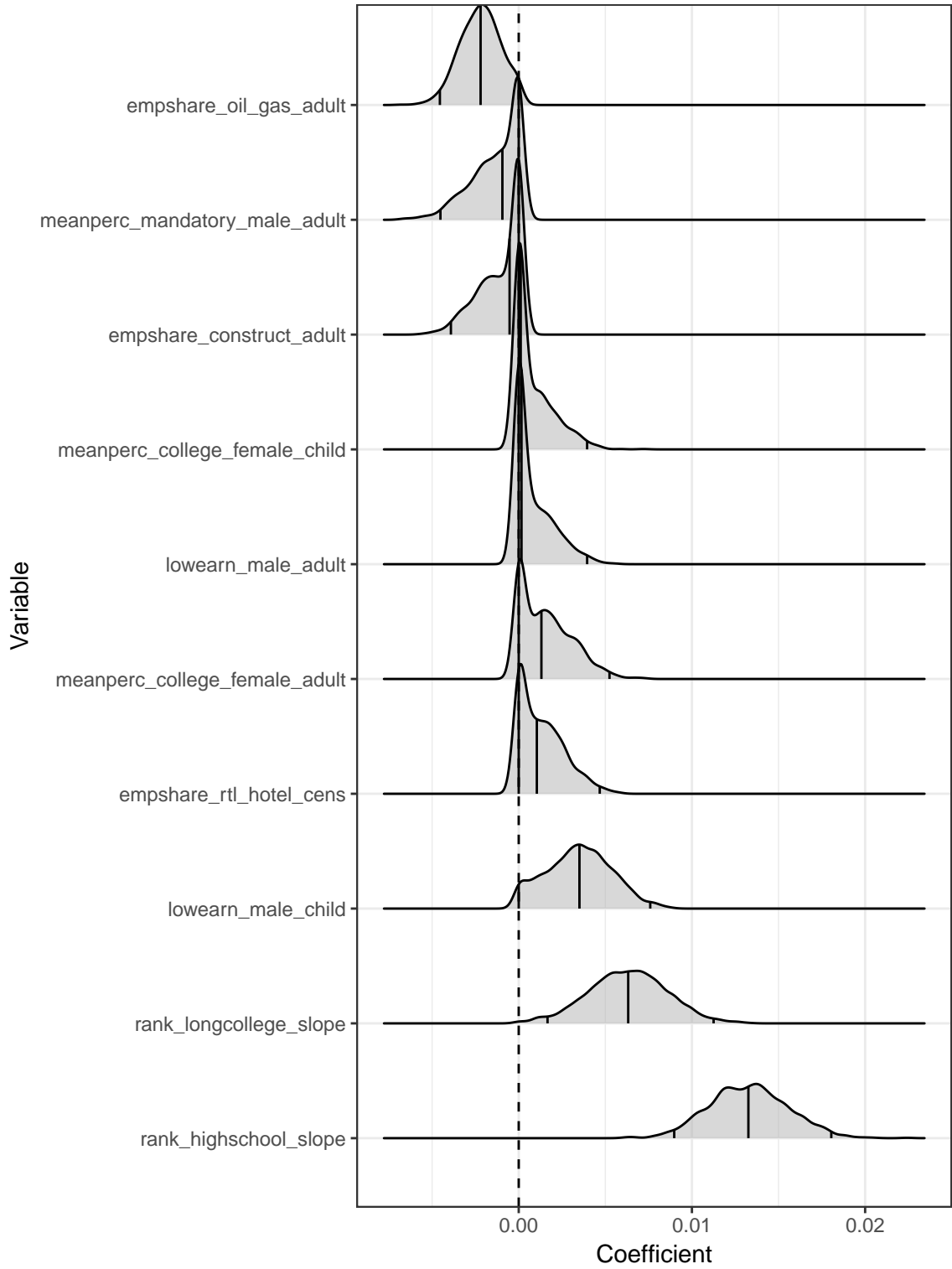
Figure 3.52: Region Fixed Effect Estimates - Rank-Rank

Region FE
Estimates

	(−0.077,−0.0614]
	(−0.0614,−0.046]
	(−0.0306,−0.0153]
	(−0.0153,0.000139]
	(0.000139,0.0155]
	(0.0155,0.0309]
	(0.0309,0.0463]
	(0.0463,0.0619]
	NA
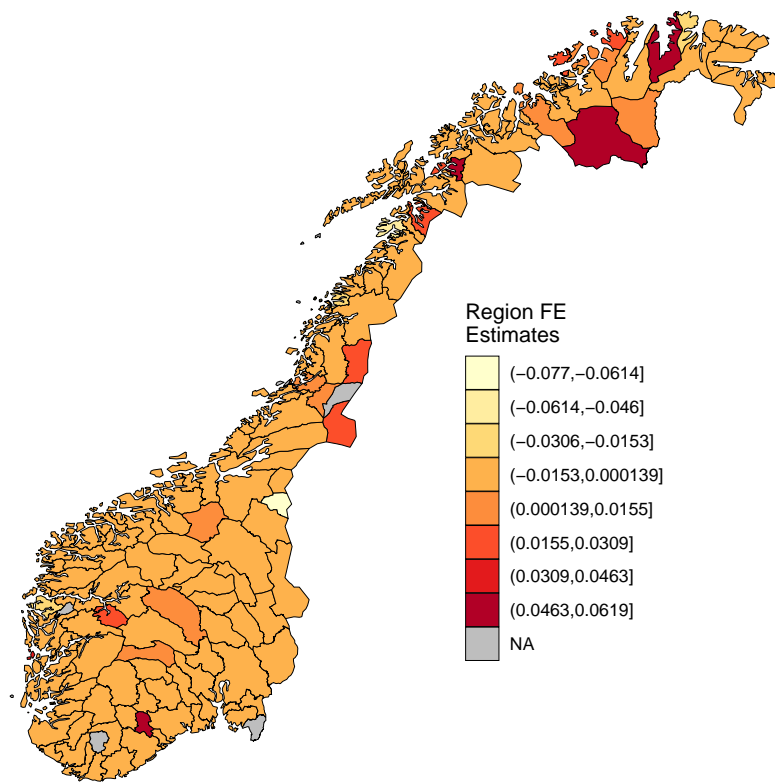
Figure 3.53: Top Ten Elastic Net Coefficients - Bottom to top
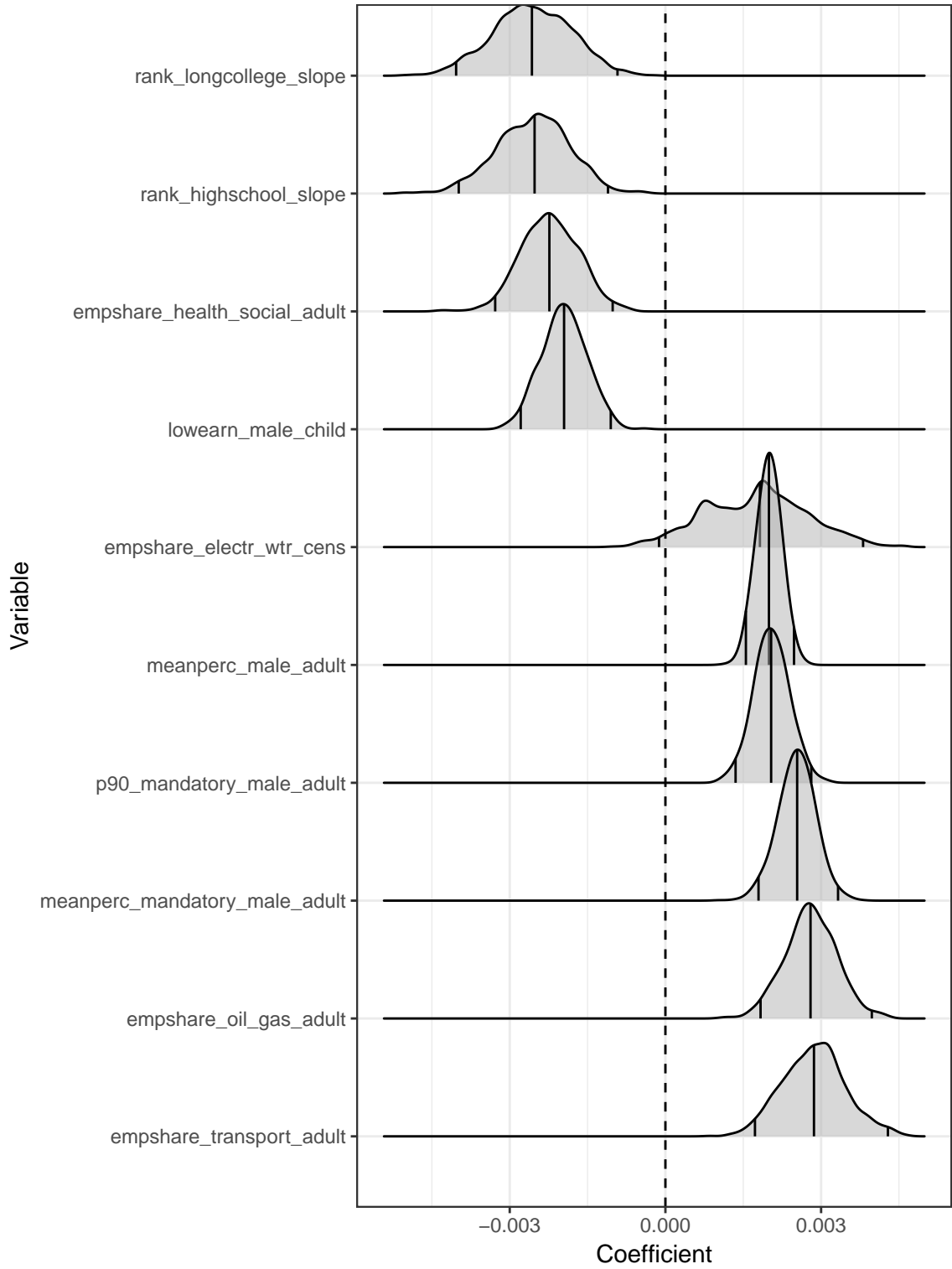
Figure 3.54: Region Fixed Effect Estimates - Bottom to top

Region FE
Estimates

(−0.046,−0.0359]
(−0.0259,−0.0158]
(−0.0158,−0.00578]
(−0.00578,0.00426]
(0.00426,0.0143]
(0.0143,0.0243]
(0.0243,0.0344]
(0.0344,0.0445]
NA
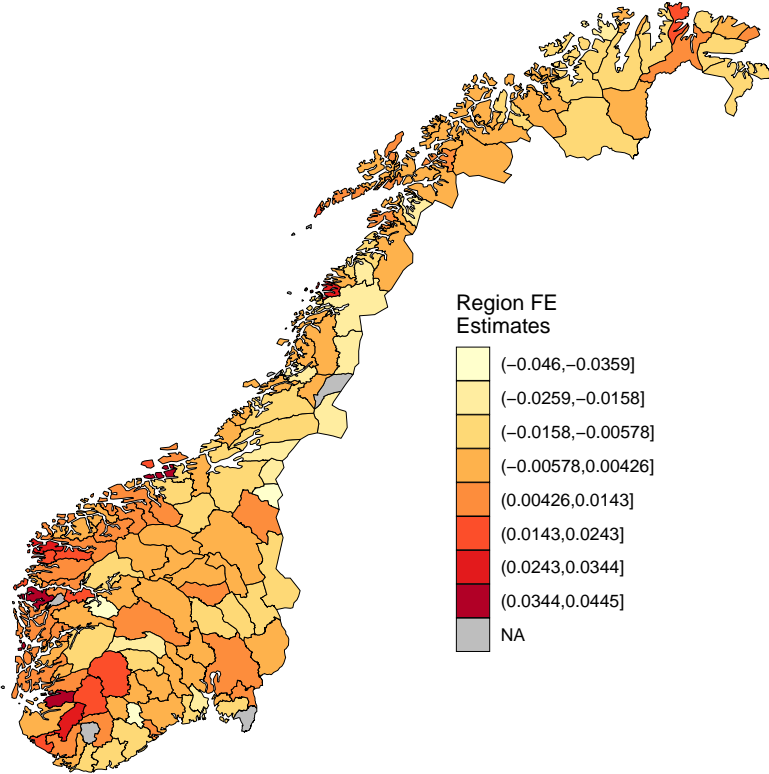
Figure 3.55: Comparison of Rank-Rank Coefficients
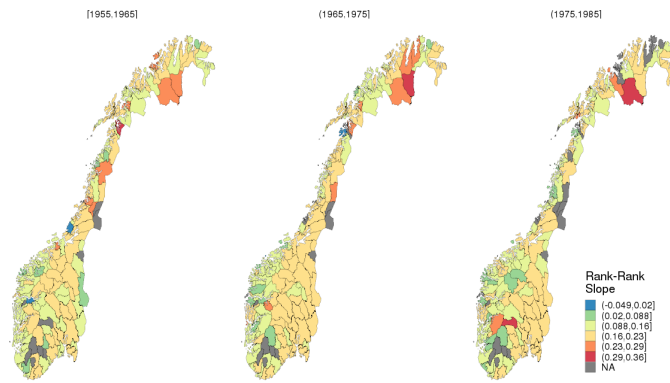
Current Level of Aggregation
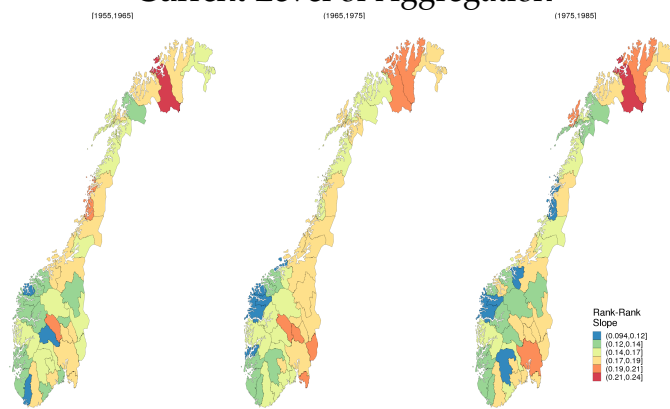
Larger Regions

Figure 3.56: Correlation between father's income rank and children's IQ
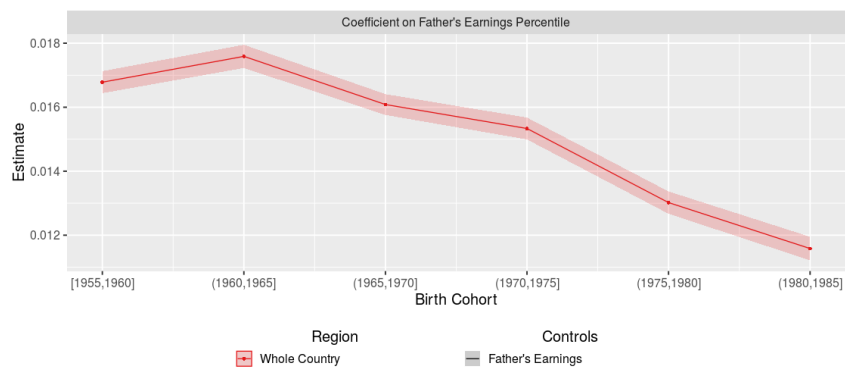
Figure 3.57: Children's IQ in top and bottom quintiles of parental income distribution
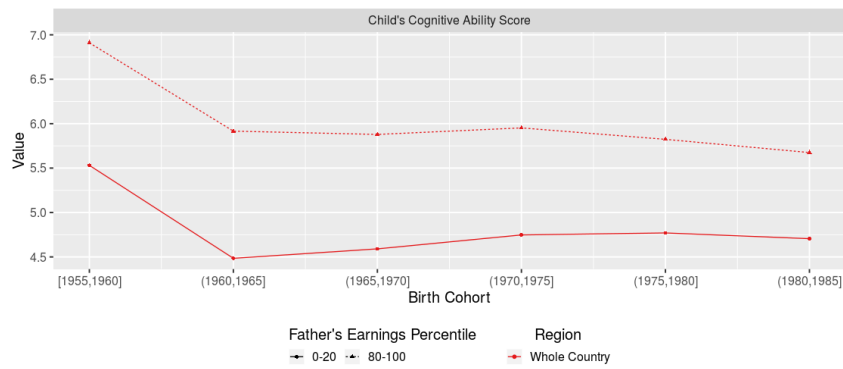
Figure 3.58: Correlation between father's income rank and children's IQ (regional)
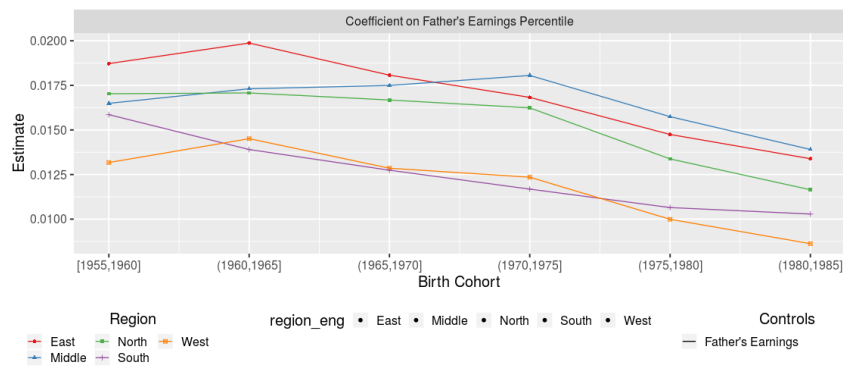
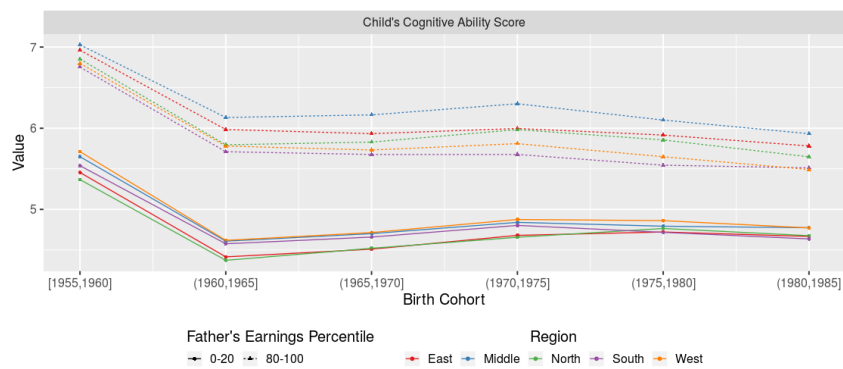**Figure 3.59:** Children's IQ in top and bottom quintiles of parental income distribution (regional)

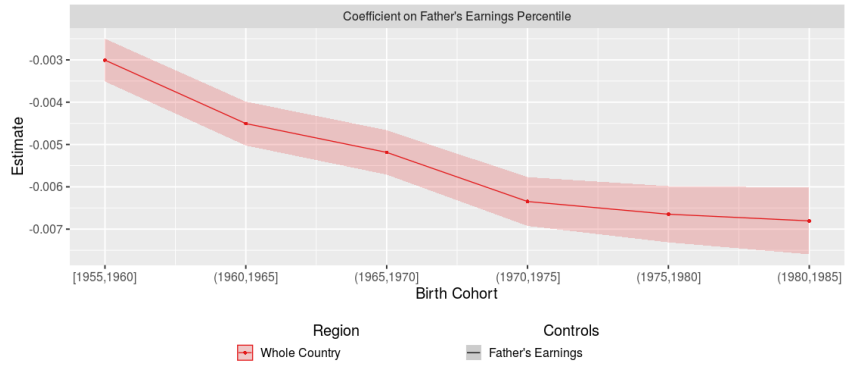**Figure 3.60:** Correlation between father's income rank and children's BMI

Figure 3.61: Children's BMI in top and bottom parental income quintile
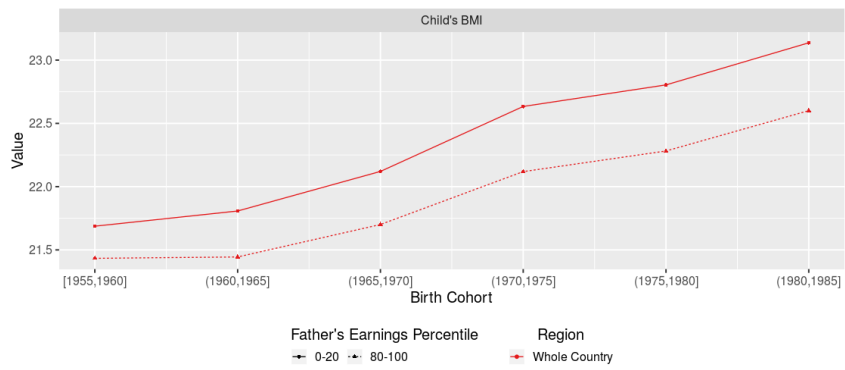

Figure 3.62: Correlation between father's income rank and children's BMI
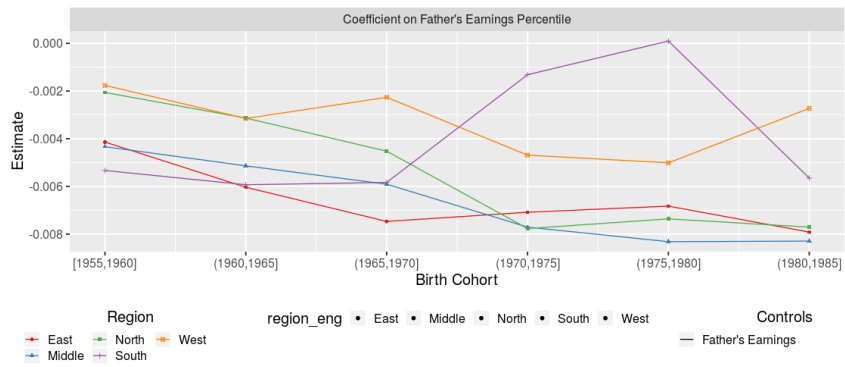
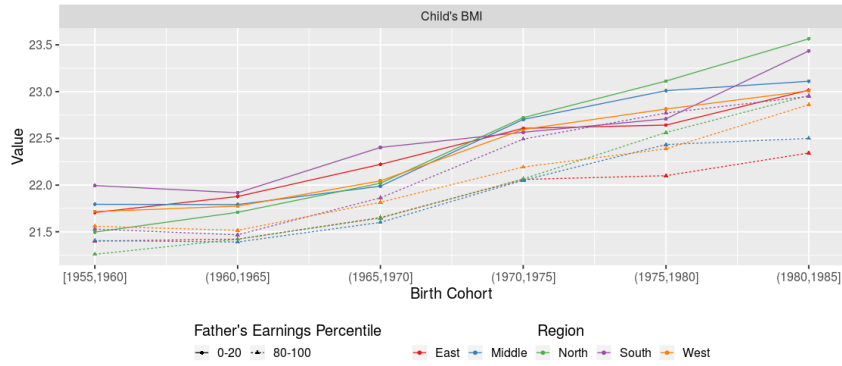**Figure 3.63:** Children's BMI in top and bottom parental income quintile (regional)

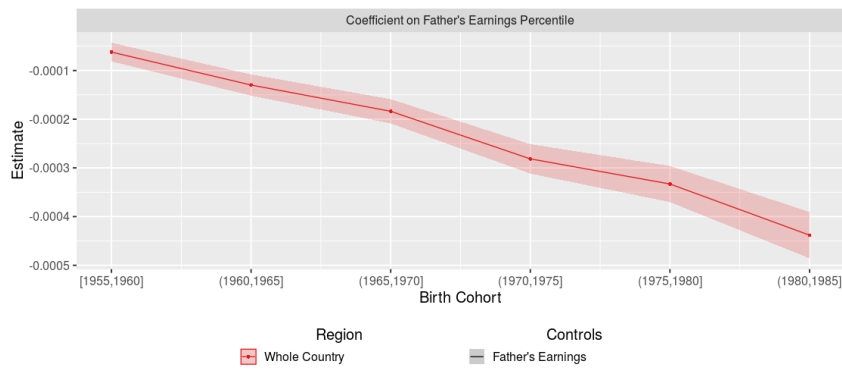**Figure 3.64:** Correlation between father's income rank and children's obesity

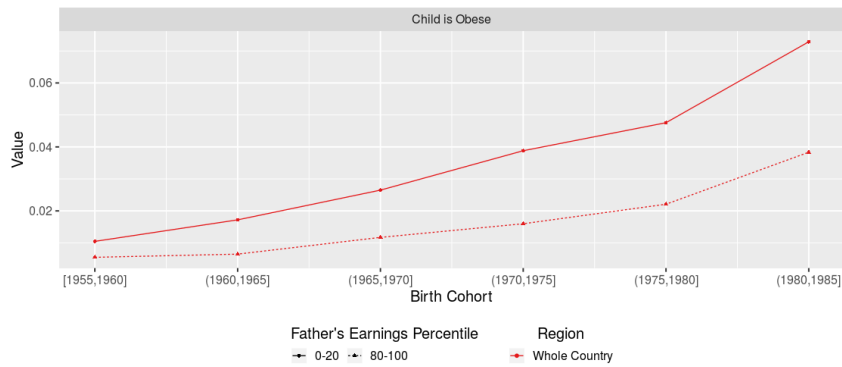**Figure 3.65:** Obesity in top and bottom parental income quintiles

Figure 3.66: Correlation between father's income rank and children's obesity
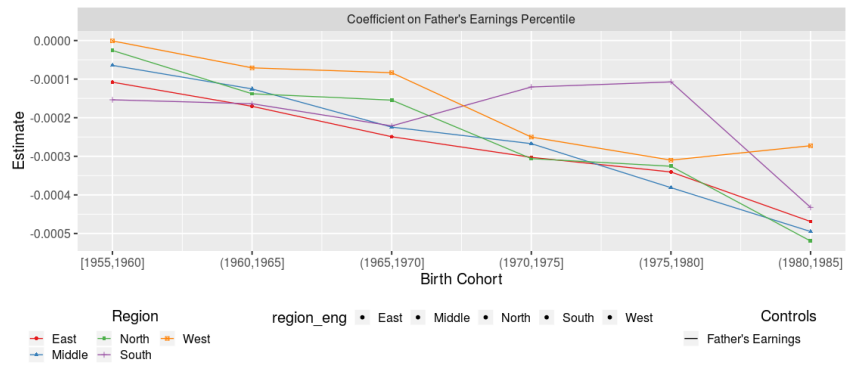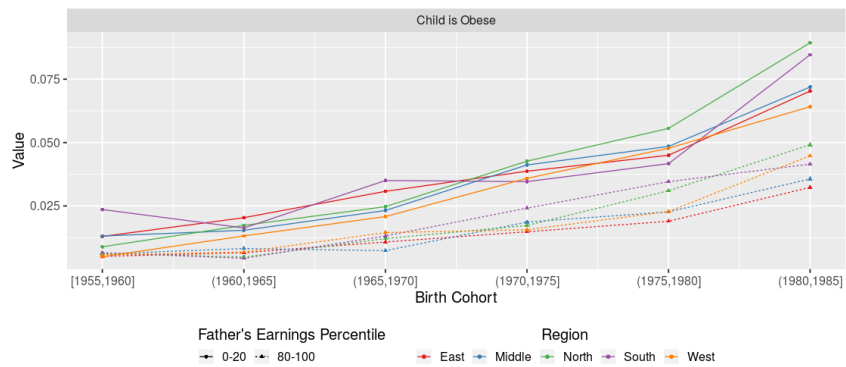

Figure 3.67: Obesity in top and bottom parental income quintiles

References

Adermon, Adrian, Mikael Lindahl, and Marten Palme (2019). *Dynastic Human Capital, Inequality and Intergenerational Mobility*. SSRN Scholarly Paper ID 3384306. Rochester, NY: Social Science Research Network.

Athey, Susan and Guido W. Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation". In: *Journal of Economic Perspectives* 31.2, pp. 3–32.

Barrabés, Nuria and Greta Kjølstad Østli (2016). *Norsk standard for utdanningsgruppering 2016*. Statistics Norway.

Becker, Gary S. and Nigel Tomes (1979). "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility". In: *Journal of Political Economy* 87.6, pp. 1153–1189.

Becker, Gary S. and Nigel Tomes (1986). "Human Capital and the Rise and Fall of Families". In: *Journal of Labor Economics* 4.3, S1–S39.

Belley, Philippe and Lance Lochner (2007). "The Changing Role of Family Income and Ability in Determining Educational Achievement". In: *Journal of Human Capital* 1.1, pp. 37–89.

Bertrand, Marianne and Emir Kamenica (2018). "Coming apart? Cultural distances in the United States over time". In: *Working Paper*, p. 81.

Bhuller, Manudeep S. (2009). *Inndeling av Norge i arbeidsmarkedsregioner*. 2009/24. Statistics Norway.

Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes (2017). "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return". In: *Journal of Labor Economics* 35.4, pp. 993–1030.

Björklund, Anders and Kjell G. Salvanes (2011). "Chapter 3 - Education and Family Background: Mechanisms and Policies". In: *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 3. Supplement C vols. Elsevier, pp. 201–247.

Black, Sandra E. and Paul J. Devereux (2010). *Recent Developments in Intergenerational Mobility*. Working Paper 15889. National Bureau of Economic Research.

Black, Sandra E. and Paul J. Devereux (2011). "Chapter 16 - Recent Developments in Intergenerational Mobility". In: *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4. Elsevier, pp. 1487–1541.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2005). "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital". In: *American Economic Review* 95.1, pp. 437–449.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2009). "Like father, like son? A note on the intergenerational transmission of IQ scores". In: *Economics Letters* 105.1, pp. 138–140.

Blanden, Joanne, Ellen Greaves, Paul Gregg, Lindsey Macmillan, and Luke Sibieta (2015). "Understanding the improved performance of disadvantaged pupils in London". In: *Social Policy in a Cold Climate*, pp. 2–47.

Blau, Peter M. and Otis Dudley Duncan (1967). *The American Occupational Structure*. John Wiley & Sons.

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage (2017). "A Comparison of Intergen-

erational Mobility Curves in Germany, Norway, Sweden, and the US". In: *The Scandinavian Journal of Economics* 119.1, pp. 72–101.

Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage (2005). "Intergenerational Earnings Mobility in Norway: Levels and Trends*". In: *Scandinavian Journal of Economics* 107.3, pp. 419–435.

Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage (2007). "Trends in Intergenerational Mobility across Offspring's Earnings Distribution in Norway". In: *Industrial Relations: A Journal of Economy and Society* 46.1, pp. 112–129.

Braun, Sebastian Till and Jan Stuhler (2018). "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany". In: *The Economic Journal* 128.609, pp. 576–611.

Breiman, Leo (2017). *Classification and Regression Trees*. Routledge.

Brunori, Paolo, Paul Hufe, and Daniel Gerszon Mahler (2018). *The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees*. Policy Research Working Papers. The World Bank.

Butikofer, Aline, Antonio Dalla Zuanna, and Kjell G. Salvanes (2018). *Breaking the Links: Natural Resource Booms and Intergenerational Mobility*. SSRN Scholarly Paper ID 3246895. Rochester, NY: Social Science Research Network.

Bütikofer, Aline, Erling Risa, and Kjell G. Salvanes (2019). "Status Traps and Human Capital Investment". In: *Unpublished*.

Carneiro, Pedro Manuel, Kai Liu, and Kjell G. Salvanes (2018). *The Supply of Skill and Endogenous Technical Change: Evidence from a College Expansion Reform*. SSRN Scholarly Paper ID 3214563. Rochester, NY: Social Science Research Network.

Casella, George and Roger L. Berger (2002). *Statistical Inference*. Second Edition. Duxbury Thomson Learning.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2017). *Mobility report cards: The role of colleges in intergenerational mobility*. National Bureau of Economic Research.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan (2019). "Income Segregation and Intergenerational Mobility Across Colleges in the United States". In: *Unpublished*.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang (2017). "The fading American dream: Trends in absolute income mobility since 1940". In: *Science* 356.6336, pp. 398–406.

Chetty, Raj and Nathaniel Hendren (2018a). "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects". In: *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

Chetty, Raj and Nathaniel Hendren (2018b). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates". In: *The 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: *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner (2014). "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility". In: *American Economic Review* 104.5, pp. 141–147.

Clark, Gregory (2014). *The Son Also Rises*. Princeton University Press.

Connolly, Marie, Miles Corak, and Catherine Haeck (2019). "Intergenerational Mobility Between and Within Canada and the United States". In: *Journal of Labor Economics* 37 (S2), S595–S641.

Corak, Miles (2013). "Income Inequality, Equality of Opportunity, and Intergenerational Mobility". In: *Journal of Economic Perspectives* 27.3, pp. 79–102.

Corak, Miles, Matthew J. Lindquist, and Bhashkar Mazumder (2014). "A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States". In: *Labour Economics* 30, pp. 185–200.

Dahl, Molly and Thomas DeLeire (2009). "The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data." In: *Institute for Research on Poverty, University of Wisconsin-Madison*, p. 44.

Durlauf, Steven N., Andros Kourtellos, and Chih Ming Tan (2017). "Status Traps". In: *Journal of Business & Economic Statistics* 35.2, pp. 265–287.

Freund, Yoav and Robert E Schapire (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *J. Comput. Syst. Sci.* 55.1, pp. 119–139.

Friedman, Jerome H. (2001). "Greedy function approximation: A gradient boosting machine." In: *The Annals of Statistics* 29.5, pp. 1189–1232.

Galton, Francis (1877). "Typical Laws of Heredity". In: *Proceedings of the Royal Institution*.

Gathergood, John, Neale Mahoney, Neil Stewart, and Jörg Weber (2018). "How Do Individuals Repay Their Debt? The Balance-Matching Heuristic". In: *Becker Friedman Institute for Research in Economics Working Paper No. 2018-7*, p. 80.

Gentzkow, Matthew, Jesse Shapiro, and Matt Taddy (2016). *Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech*. w22423. Cambridge, MA: National Bureau of Economic Research.

Güell, Maia, José V. Rodríguez Mora, and Gary Solon (2018). "New Directions in Measuring Intergenerational Mobility: Introduction". In: *The Economic Journal* 128.612, F335–F339.

Güell, Maia, José V. Rodríguez Mora, and Christopher I. Telmer (2015). "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating". In: *The Review of Economic Studies* 82.2, pp. 693–735.

Gundersen, Frants and Dag Juvkam (2013). *Inndelinger i senterstruktur, sentralitet og BA-regioner*. 1. Norsk institutt for by- og regionforskning.

Hastie, Trevor and Junyang Qian (2014). *Glmnet Vignette*. URL: `https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html` (visited on 08/10/2019).

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. Second. Springer.

Hertz, Tom (2007). "Trends in the Intergenerational Elasticity of Family Income in the United States". In: *Industrial Relations: A Journal of Economy and Society* 46.1, pp. 22–50.

Hothorn, Torsten and Achim Zeileis (2016). *partykit: A Toolkit for Recursive Partytioning*. URL: `https://cran.r-project.org/package=partykit` (visited on 10/19/2017).

Jæger, Mads Meier (2012). "The Extended Family and Children's Educational Success". In: *American Sociological Review* 77.6, pp. 903–922.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad (2016). "Field of Study, Earnings, and Self-Selection". In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Kleinberg, Jon, Annie Liang, and Sendhil Mullainathan (2018). "The Theory is Predictive, but is it Complete? An Application to Human Perception of Randomness". In: *Working Paper*, p. 23.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015). "Prediction Policy Problems". In: *American Economic Review* 105.5, pp. 491–495.

Kourtellos, Andros, Christa Marr, and Chih Ming Tan (2018). *Local Intergenerational Mobility*. SSRN Scholarly Paper ID 2496826. Rochester, NY: Social Science Research Network.

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.

Liang, Annie and Drew Fudenberg (2018). "Predicting and Understanding Initial Play". In: *Working Paper*.

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren (2015). "Long-Term Intergenerational Persistence of Human Capital An Empirical Analysis of Four Generations". In: *Journal of Human Resources* 50.1, pp. 1–33.

Long, Jason and Joseph Ferrie (2018). "Grandfathers Matter(ed): Occupational Mobility Across Three Generations in the US and Britain, 1850–1911". In: *The Economic Journal* 128.612, F422–F445.

Mare, Robert D. (2011). "A Multigenerational View of Inequality". In: *Demography* 48.1, pp. 1–23.

Mullainathan, Sendhil and Ziad Obermeyer (2017). "Does Machine Learning Automate Moral Hazard and Error?" In: *American Economic Review* 107.5, pp. 476–480.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine Learning: An Applied Econometric Approach". In: *Journal of Economic Perspectives* 31.2, pp. 87–106.

Nicoletti, Cheti and Birgitta Rabe (2013). "Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter?" In: *Economica* 80, pp. 197–218.

Nybom, Martin and Jan Stuhler (2013). *Interpreting Trends in Intergenerational Income Mobility*. SSRN Scholarly Paper ID 2314809. Rochester, NY: Social Science Research Network.

OECD (2014). *Education at a Glance 2014*.

Page, Marianne E. and Gary Solon (2003). "Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban". In: *Journal of Labor Economics* 21.4, pp. 831–855.

Pekkarinen, Tuomas, Kjell G. Salvanes, and Matti Sarvimäki (2017). "The Evolution of Social Mobility: Norway during the Twentieth Century". In: *The Scandinavian Journal of Economics* 119.1, pp. 5–33.

Raaum, Oddbjørn, Kjell G. Salvanes, and Erik Ø. Sørensen (2006). "The Neighbourhood is Not What it Used to be*". In: *The Economic Journal* 116.508, pp. 200–222.

Roemer, John E and Alain Trannoy (1998). *Equality of Opportunity*. Harvard University Press.

Rothstein, Jesse (2019). "Inequality of Educational Opportunity? Schools as Mediators of the Intergenerational Transmission of Income". In: *Journal of Labor Economics* 37 (S1), S85–S123.

Solon, Gary (1999). "Chapter 29 - Intergenerational Mobility in the Labor Market". In: *Handbook of Labor Economics*. Ed. by Orley C. Ashenfelter and David Card. Vol. 3. Part A vols. Elsevier, pp. 1761–1800.

Solon, Gary (2018). "What Do We Know So Far about Multigenerational Mobility?" In: *The Economic Journal* 128.612, F340–F352.

Solon, Gary, Marianne E. Page, and Greg J. Duncan (2000). "Correlations between Neighboring Children in Their Subsequent Educational Attainment". In: *The Review of Economics and Statistics* 82.3, pp. 383–392.

Sundet, Jon Martin, Dag G. Barlaug, and Tore M. Torjussen (2004). "The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century". In: *Intelligence* 32.4, pp. 349–362.

Sundet, Jon Martin, Kristian Tambs, Jennifer R. Harris, Per Magnus, and Tore M. Torjussen (2005). "Resolving the genetic and environmental sources of the correlation between height and intelligence: a study of nearly 2600 Norwegian male twin pairs". In: *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies* 8.4, pp. 307–311.

Thrane, Vidkunn Coucheron (1977). *Evneprøving av Utskrivingspliktige i Norge 1950-53*. Arbeidsrapport nr. 26. INAS.

Vaart, Aad W. van der, Sandrine Dudoit, and Mark J. van der Laan (2006). "Oracle Inequalities for multi-fold cross validation". In: *Statistics & Decisions* 24, pp. 351–371.

Vosters, Kelly (2017). "Is the simple Law of Mobility really a law? Testing Clark's hypothesis". In: *The Economic Journal* Forthcoming.

Vosters, Kelly and Martin Nybom (2016). "Intergenerational Persistence in Latent Socioeconomic Status: Evidence from Sweden and the United States". In: *Journal of Labor Economics* 35.3, pp. 869–901.

Zeileis, Achim, Torsten Hothorn, and Kurt Hornik (2008). "Model-Based Recursive Partitioning". In: *Journal of Computational and Graphical Statistics* 17.2, pp. 492–514.

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.