Using Machine Learning Methods to study research questions in health, labor and family economics

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Philipp Kugler

Tübingen

2022

Tag der mündlichen Prüfung:	10.07.2023
Dekan:	Prof. Dr. Ansgar Thiel
1. Gutachter:	Prof. Dr. Martin Biewen
2. Gutachter:	Prof. Dr. Bernhard Boockmann

Publication Notes

Chapter 2 of this doctoral thesis is based on the paper Kugler, P. (2022). The role of wage beliefs in the decision to become a nurse. *Health Economics, 31*(1), 94-111.

A shorter version of chapter 3 of this doctoral thesis has been published as Biewen, M., & Kugler, P. (2021). Two-Stage Least Squares random forests with an application to Angrist and Evans (1998). *Economics Letters, 204*, 109893.

Chapter 4 is based on Kugler, P., Kroczek, M., & Zühlke, A. (2022). Parenting Styles, socioeconomic status and (non-)cognitive skills. *IAW-Diskussionspapiere*, Nr. 138.

Danksagung

Während meiner Dissertation haben mich zahlreiche Personen unterstützt, denen ich an dieser Stelle gerne danken möchte.

Zuallererst möchte ich Martin für seine Unterstützung als Doktorvater und Koautor danken. Vielen Dank für die unkomplizierte, angenehme Zusammenarbeit, dafür, dass du dir immer Zeit genommen hast, dein großes Wissen mit mir zu teilen und mir stets hilfreichen Input gegeben hast. Ebenso dankbar bin ich meinen anderen beiden Koautoren und Mitpromovierenden am IAW, Anne und Martin. Unsere Zusammenarbeit war nicht nur spannend und produktiv, sondern hat vor allem auch sehr viel Spaß gemacht. Ihr seid mir über die letzten Jahre gute Freunde geworden.

Zudem möchte ich mich auch herzlich bei Wilhelm Kohler und Bernhard bedanken, dass sie mir durch die Anstellung am IAW erst die Möglichkeit zu dieser Arbeit gegeben haben. Ich danke auch dem ganzen IAW Team für das gute Arbeitsklima, die vielen lustigen Mittags- und Kaffeepausen und die lebhaften Diskussionen während meiner Dissertation.

Zuletzt möchte ich mich bei meiner Frau Laura bedanken. Danke, dass du immer Recht hast und alles weißt. Diese Dissertation wäre so viel leichter gewesen, wenn ich häufiger auf dich gehört hätte.

Contents

List of Tables

List of Figures

Chapter 1

Dissertation introduction

Everyday, we find ourselves in situations where we use machines which have learned from data to perform specific tasks. We use face recognition to unlock our phones, translate texts from one language into another using Google Translator or talk to voice assistants like Amazon's Alexa or Apple's Siri. Google News clusters news articles with the same topic from a large number of websites and YouTube proposes videos that are similar to the ones we recently watched. Other machines are able to predict our preferences. Based on own movie ratings, Netflix suggests what movies we may like. Social media platforms such as Facebook or Instagram have perfected the ways to show us advertisements of products we are most likely interested in and customize our feeds to show us the content that we like.

Although all these machines are crucially different in the tasks they are performing, they are all based on a set of tools called machine learning methods. In the literature, there is no unique definition of what machine learning exactly is (see Athey, 2018, for a discussion). In a rather narrow sense, machine learning is a field in which algorithms are designed to solve different tasks of data analysis. Most tasks fall into two categories, supervised and unsupervised ones (James et al., 2013). Unsupervised machine learning algorithms refer to a set of methods that are used when there are many covariates without a label (i.e. no outcome). For example, such methods are useful when the researcher wants to identify groups of observations that have similar covariates or wants to estimate the joint distribution of many variables. They are often applied in video, image and text analysis. Supervised machine learning algorithms are used to fit a model that connects the covariates with an outcome. The goal is to build a model that is able to predict the outcome on previously unseen data. Whereas the goodness of fit is the central element in machine learning, most economists focus on the identification and estimation of causal effects. Therefore, economists often do not put much importance on the goodness of fit (see Imbens and Athey, 2021, or Mullainathan and Spiess, 2017 for an in-depth discussion on the difference between predictive and causal tasks). Maybe because of the different focus, economists have not started to integrate machine learning methods into their research until the 2010s. Since then, this branch of the literature rapidly grew. Today, a variety of machine learning methods to answer economic questions exist and an increasing number of economists apply them in their research. Athey (2018) and Athey and Imbens (2017, 2019) give reviews on these recent developments and discuss them thoroughly. The following introduction to machine learning methods and to the topics of this dissertation draws in some parts on these excellent overviews.

This doctoral thesis contributes to applied machine learning research by exploring and discussing novel methods to a number of relevant research questions. I specifically look into the question of how and when machine learning methods can be useful to answer economic questions. To this end, each chapter focuses on one specific area in which recent methodological advances have been made that are of particular interest for economists. Chapter 2 applies post-double-selection (Belloni et al., 2012, 2014a,b) to estimate average effects. Chapter 3 uses the generalized random forest framework (Athey et al., 2019) to work out the case of a Two-Stage Least Squares random forest aimed at estimating heterogeneous effects. Chapter 4 applies latent dirichlet analysis for survey data (Munro and Ng, 2022) to study the role of latent variables in a family economics application.

Chapter 2 and 3 rely on supervised machine learning methods. Supervised machine learning methods can only in some cases be used off the shelf to answer economic research questions. They are valuable if causal inference is not important, for example if the question is predictive at its core. One could use conventional supervised machine learning methods to predict asset prices (Grammig et al., 2020), loan repayment (Björkegren and Grissen, 2017) or demand curves (Bajari et al., 2015). Kleinberg et al. (2015) point out that supervised machine learning methods also can be used off the shelf to look into a wide range of policy decision problems. For example, one could analyze the characteristics of teachers that will add the most value or one could predict the length of an unemployment episode to help individuals to determine their savings rate and job search strategy. However, many research questions in economics and social sciences are of a causal nature. In these settings, supervised machine learning methods cannot be directly applied. First, the identification of causal effects requires some kind of assumption or structure. Second, the ground truth is not observable in settings where the goal is to estimate an effect, whereas in prediction settings the truth can always be observed. Therefore, it is (relatively) straightforward to optimize and evaluate a prediction model by computing the mean squared error in an independent test set. To apply machine learning methods in settings where the goal is to estimate a causal effect, the objective function has to be adjusted. Moreover, statistical theory is more important than it is in predictive settings to evaluate how well the estimated effect approximates the truth. To address these issues, a recent but rapidly developing literature combines the strengths of supervised machine learning methods and conventional econometric methods in order to estimate causal effects.

A large part of the conventional econometric literature on causal inference is about estimating average treatment effects under the assumption of unconfoundedness which assumes that the treatment assignment is as good as random after controlling for observed characteristics. The assumption requires that all characteristics of a unit that affect the treatment assignment and the potential outcomes, i.e. the unit's outcome in an alternative treatment state, are observed (Imbens and Wooldridge, 2009, provide a review on this literature). Starting in the 1990s, various semi-parametric estimators have been proposed (e.g., Robinson, 1988; Hahn, 1998; Heckman et al., 1998; Abadie and Imbens, 2006). All these methods estimate a low dimensional parameter (average treatment effect) by flexibly modeling the way how a small number of covariates relative to sample size affect the outcome. Machine learning methods to estimate average effects build on this line of research. They provide new approaches to estimate semi-parametric models, when the researcher observes many covariates relative to sample size.

Building on Robinson's (1988) partially linear model, Belloni et al. (2012, 2014b) propose a procedure that uses the least absolute shrinkage and selection operator (lasso, Tibshirani, 1996) to choose controls. Lasso estimates a regression model using an objective function that penalizes the absolute sum of coefficients to prevent the model from overfitting. The penalty draws all regression coefficients towards zero, many of them exactly to zero effectively dropping them from the model. Belloni et al. (2012, 2014b) point out that choosing controls only based on a lasso model of the outcome equation would lead to biased average effects. Lasso might drop covariates that are weakly correlated with the outcome but strongly correlated with the treatment indicator, since the algorithm's only purpose is to predict the outcome as precisely as possible. The reverse would be the case if controls were chosen using a lasso model of the treatment indicator. To overcome this problem, they propose post-double-selection, a procedure that selects covariates by estimating two lasso models. Using all covariates, one lasso model of the outcome and one lasso model of the treatment indicator is estimated. Then, all covariates with non-zero coefficients in either of the lasso models are used as controls. In chapter 2, post-double-selection is used to analyze the role of the beliefs about a nurse's wage in the decision to become one. In a more recent paper, Chernozhukov et al. (2017) propose a general procedure, they call 'double machine learning'. The method uses score functions that satisfy the Neyman orthogonality condition, for example doubly robust scores in the sense of Robins et al. (1994, 1995), and applies sample splitting to achieve good statistical properties. The potentially high-dimensional nuisance parameters of the score function (parameters which are not of primary interest but have to be estimated) are estimated with conventional machine learning methods and are used as plug-in estimators for their population equivalent. For the case of Robinson's (1988) partially linear model, the procedure complements post-double-selection. Similar to the method proposed by Robinson (1988), double machine learning yields a semi-parametric residual-on-residual regression. First, a non-parametric regression of the outcome on the covariates and a non-parametric regression of the treatment indicator on the covariates is conducted. Then, the residuals from the first regression are regressed on the residuals from the second regression. Robinson (1988) uses a kernel regression to run the non-parametric regressions. Using the approach proposed by Chernozhukov et al. (2017), any machine learning method can be used for this task.

Another line of the literature concerned with the estimation of average effects designs methods that mimic randomized experiments more closely. The methods reweight the observations such that covariates are balanced between treatment and control group. To balance covariates, propensity score matching has long been used (Heckman et al., 1998). Early work proposed to use machine learning methods instead of logit or probit models to predict the propensity score (e.g., McCaffrey et al., 2004; Wyss et al., 2014). However, as for example Belloni et al. (2014b) point out, such an approach often has poor properties since machine learning methods do not necessarily choose the most important confounders, i.e. covariates that are correlated with both treatment and outcome. Instead covariates that are most predictive of the propensity score are selected. Recent methods replace the propensity score to balance the covariate distribution with weights that are designed such that balance between treatment and control group can be directly achieved (e.g., Hainmueller, 2012; Iacus et al., 2012; Zubizarreta, 2015; Imai and Ratkovic, 2014). Athey et al. (2018b) propose 'residual balancing', an estimator combining balancing weights with regression adjustment, i.e. closely related to doubly robust methods for estimating average treatment effects. However, their approach can deal with many covariates because the conditional mean of the outcomes is estimated using a regularized linear model (lasso or elastic net). Assuming that the outcome model is linear, residual balancing does not need to impose any structure on the propensity score model other than overlap. Note that using weights that balance selected covariates is implicitly equivalent to using regression adjustments (Robins et al., 2007). Therefore, residual balancing closely relates to post-double-selection, which however performs worse in situations where the assignment model is complex (Athey et al., 2018b).

In addition to the analysis of average effects, the analysis of effect heterogeneity is important to gain a basic scientific understanding. Moreover, it can give important insights for policy makers by identifying individuals that benefit the most from a policy or for deciding which groups should get a certain treatment (see Athey and Imbens, 2017, for further discussion). Often, researchers conduct a subgroup analysis by either including interactions into the model or running the model on subsamples and compare the results between them. Some approaches systematically search for heterogeneity over many dimensions and account for multiple hypotheses testing (e.g., Chernozhukov et al., 2018; List et al., 2019). Machine learning methods offer an alternative to identify groups that differ the most in their effect size. For this task, decision trees are a natural choice (Breiman et al., 1984). Conventional decision trees successively split a given sample into subgroups and assign the same prediction (e.g., the mean of the outcome) to each observation in the same group. The sample is split such that the objective, in case of a regression tree the mean squared error, is minimized in each step. Due to the iterative partitioning of the data, the model can be represented as a tree and final partitions are called leafs. Su et al. (2009) and Athey and Imbens (2016) propose using a tree to partition the covariate space and then to estimate the treatment effect in each partition of the data. To account for the fact that the mean squared error cannot be computed because the true effect is unknown, they suggest several alternative objective functions. The preferred estimator of Athey and Imbens (2016) is the causal tree, which splits the covariate space in order to maximize the heterogeneity. Moreover, Athey and Imbens (2016) propose sample splitting to avoid bias that occurs when the same data is used to discover heterogeneity and to estimate the effects. They split the sample into two parts, building the tree on one sample and estimating the effects using the other sample. Zeileis et al. (2008) and Asher et al. (2016) propose more general frameworks in which trees are used to reveal heterogeneity. Zeileis et al. (2008) use trees to automatically find heterogeneity in parametric models such as maximum likelihood, whereas Asher et al. (2016) pair the generalized method of moments with classification trees to analyze heterogeneity. Asher et al. (2016) use the sample splitting procedure proposed by Athey and Imbens (2016) to provide an asymptotic theory of their estimator.

Often, researchers and policy-makers are not only interested in differences between subgroups but want to obtain a smooth estimation of effect heterogeneity, e.g., when personalized decisions have to be made. This could be achieved by using kernel or matching estimators (Athey and Imbens, 2019). However, these methods do not work well when there are many covariates. Imai and Ratkovic (2013) view the problem to estimate heterogeneous effects in high dimensions as a variable selection problem and use support vector machines to solve it. Using double machine learning (Chernozhukov et al., 2017), any machine learning method can be used to estimate heterogeneous effects if there is a proper score function (see Knaus 2020 for an application under the assumption of unconfoundedness). Nie and Wager (2021) propose the R-learner to estimate heterogeneous treatment effects in observational studies in two steps. First, two nuisance parameters (conditional mean outcome and propensity score) are estimated using flexible machine learning methods. To recover the heterogeneous effect, the predictions are plugged into a loss function based on Robinson (1988).

A natural choice to non-parametrically estimate the treatment effect as a function of a high-dimensional vector of covariates is to use a random forest. A random forest averages many regression or classification trees that are repeatedly built on random subsamples of the data. Random forests have been adapted to solve specific tasks. For example, Meinshausen and Ridgeway (2006) propose a quantile forest to infer conditional quantiles instead of conditional means and Ishwaran et al. (2008) design a survival forest to analyze right-censored survival data. Wager and Athey (2018) propose the causal forest, a forest composed of many causal trees (Athey and Imbens, 2016), to estimate heterogeneous treatment effects. Using sample splitting, Wager and Athey (2018) show that the estimates of a causal forest are consistent and asymptotically normal. They further provide a valid estimator of the variance to build confidence intervals. More recently, Athey et al. (2019) propose a general framework that utilizes random forests to estimate conditional moment conditions. In a nutshell, the method consists of two steps. In the first step, the method assigns one weight to each observation depending on its similarity to a given test point (covariate vector at which the effect is estimated). To this end, the random forest is viewed as a nearest neighbor metric (e.g., Lin and Jeon, 2006). Observations that are close to the test point in terms of the random forest, i.e. more often fall into the same leaf as the test point, obtain higher weights. Observations that only seldom fall into the same leaf as the test point, obtain lower weights. In the second step, these weights are used to solve weighted moment conditions, similar to local maximum likelihood where weights come from a kernel (Tibshirani and Hastie, 1987). The authors establish consistency and asymptotic normality of their estimator and provide a procedure to compute valid confidence intervals. Athey et al. (2019) apply their method to quantile regression and to an instrumental variable model with one instrument. In chapter 3, we will use Athey et al.'s (2019) framework to work out the case of a Two-Stage Least Squares random forest that can be used to estimate heterogeneous causal effects, when there is more than one valid instrument. Hartford et al. (2017) proposes an alternative method that uses a deep neural network to estimate heterogeneous effects in instrumental variable settings. Bayesian additive regression trees (BART, Chipman et al. 2010) are closely related to random forests. Hill (2011) and Green and Kern (2012) use BART to estimate heterogeneous effects. Although asymptotic properties are unknown, BART seems to work well in practice.

Personalized effects are often used to determine how to allocate a policy. In the machine learning literature, this problem has been discussed for the case where a causal effect can be identified under the assumption of unconfoundedness (e.g., Strehl et al., 2010; Dudík et al., 2014, 2011; Li et al., 2015; Swaminathan and Joachims, 2015). For a binary and exogenous treatment with known treatment probability, Kitagawa and Tetenov (2018) propose an algorithm based on inverse probability weighting. Building on their work, Athey and Wager (2021) develop a family of algorithms that can be applied in a variety of settings. They consider the case of a binary and continuous exogenous treatment (given observables) and the case of a binary endogenous treatment, when the researcher has a valid instrument. The authors show that the problem of optimal policy assignment can be reframed as a classification problem which can be solved using any machine learning method. Due to their interpretability, Athey and Wager (2021) propose to use decision trees. Zhou et al. (2022) extend the algorithm to the case with many potential treatments.

In contrast to chapters 2 and 3, chapter 4 applies an unsupervised machine learning method to infer latent classes. Unlike supervised machine learning methods, unsupervised methods can often be directly applied to answer questions that are of interest for economists. Some unsupervised machine learning algorithms, for example generative adversarial networks (GANs) proposed by Goodfellow et al. (2014), can be used to estimate the joint distribution given a large set of covariates. GANs search a model that is able to generate data that look like the sample. The fit of the model is assessed by comparing the generated data with the original data, similar to a Turing test (i.e. testing whether one is being able to tell whether the data was generated by the model or whether it is the original data). To this end, GANs are formulated as the solution of a minimax problem between two models, a generator and discriminator. The generator generates synthetic data and the discriminator classifies whether the data is synthetic or not. Whereas the discriminator maximizes the accuracy of its classification, the generator minimizes it. In economics, GANs can be used to simulate artificial data that closely mimic real datasets. Athey et al. (2021b) propose to use them to conduct Monte Carlos Simulations to evaluate newly developed methods. This limits the freedom of the researcher in simulating data and increases the trust in Monte Carlos simulations. Moreover, Kaji et al. (2020) propose an adversarial approach to estimate the parameters of complex structural models. A generator is used to generate synthetic observations using the structural model and a discriminator has to tell whether the observation is synthetic or not.

Most unsupervised methods are used to reduce the dimensionality of the data when there are many covariates. Some methods partition the data into subsamples, each containing individuals with similar characteristics. One famous algorithm for this task is K-Means Clustering (Hartigan and Wong, 1979). The algorithm iterates between assigning observations to clusters by minimizing the distance of each observation to the centroid (center of the clusters) and accordingly updating the centroids until convergence is achieved. Other methods search for a low dimensional representation of the data. For example, matrix factorization finds two low-dimensional matrices whose product approximates a larger matrix (the data). In practice, this approach is often used to solve matrix completion problems, e.g., recommender systems to suggest movies. There, individuals give ratings to some movies, but not to many others and the goal is to make the best prediction for the missing entries. Athey et al. (2021a) transfer the perspective of a matrix completion problem to conduct causal inference in a panel-data setting. They propose a method to predict (not observable) counterfactual outcomes of treated individuals.

Matrix factorization can be interpreted as representing a matrix as a vector of latent characteristics for each row and column (i.e. a vector for each individual and each movie or a vector for each unit and each time period). The prediction of a cell in the matrix is the inner product of both vectors. Economists frequently rely on latent variable models, in particular in cases where the researcher wants to understand an underlying concept to make better sense of the data. For example, Carneiro et al. (2003) and Cunha et al. (2010) use latent variable models to estimate the technology of cognitive and non-cognitive skills. Stock and Watson

(2011) give an extensive review on dynamic factor models, often used in macroeconomics to summarize the variation of important indicators such as output, employment or prices. In the machine learning literature, latent variable models called topic models are of often used for unsupervised text analysis. Topic models assume that documents consist of a mixture of topics, latent characteristics, and each topic is a collection of words (see Blei and Lafferty 2009 for a review on topic models). An early method to estimate such a model was latent semantic indexing (LSI, Deerwester et al. 1990). LSI applies singular value decomposition to retrieve the latent semantic structures. More recent topic models are based on latent dirichlet analysis (LDA, Blei et al. (2003), a hierarchical bayesian model which can be solved using MCMC methods. LDA makes use of the dirichlet distribution, which is a distribution over discrete distributions. The topic proportions are assumed to be dirichlet random variables over topics, and the topics are assumed to be dirichlet random variables over words.

Munro and Ng (2022) develop LDA for survey data (LDA-S), a bayesian hierarchical latent class model closely related to conventional LDA. LDA-S has several features that help economists to analyze hidden structure in the data. The model connects unobserved heterogeneity with observed characteristics and survey responses and explicitly acknowledges that survey responses are categorical. Munro and Ng (2022) show that the statistical model corresponds to a structural model of utility maximization, which guides the interpretation and estimation of the model parameters. LDA-S is used in chapter 4 of this thesis to construct parenting styles and study their effect on cognitive and non-cognitive skills.

Moreover, tools applied in the machine learning literature to solve Bayesian models can be helpful for economists to estimate structural model at a larger scale. For example, Ruiz et al. (2020) propose a hierarchical model of consumer choice and consider thousands of products simultaneously. To reduce the dimension of the data, they apply hierarchical Poisson factorization (Gopalan et al., 2015) to represent each item as a vector of latent attributes. Instead of letting the consumer simultaneously consider all possible bundles, they assume that the consumer sequentially adds items to the shopping basket. Imposing such human computational constraints into the structural models is both reasonable from a theoretical perspective and makes the model computationally efficient. To solve their model, they use variational inference to approximate the posterior distribution and apply gradient descent to find the parameters of the model. Building on Ruiz et al. (2020), Athey et al. (2018a) analyze the consumer choice over restaurants using data on the morning location and lunch time restaurant choice. Similar to the model by Ruiz et al. (2020) the dimension of the data is reduced by latent variable models. In their model they allow both the users' willingness to travel and the users' utility for each restaurant to vary across user-item pairs.

To conclude this introduction, I briefly outline the three studies that represent the main body of this dissertation. Each chapter applies one recent machine learning method to study one substantive economic research question.

Average Effects: The role of wage beliefs and information in the decision to become a nurse

The focus of chapter 2 lies on the estimation of average effects. Using post-double-selection (Belloni et al., 2012, 2014a,b), I investigate the policy relevant question of whether and how wage beliefs and information influence the decision to become a nurse. This question takes an important place in the recent literature. Due to demographic change and technological progress in medicine, the demand for skilled nurses has increased in industrialized countries over the past decades. The existing literature discusses a series of factors that might alleviate the lack of skilled workers. Among these, wages are the most controversial. Some authors identify the wage as a very important factor influencing labor supply decisions of nurses. However, others suggest that the labor supply of nurses is relatively inelastic in terms of wages. The focus of this chapter is on beliefs about wages and how they influence the decision to become a nurse. Such beliefs may affect educational choices and could be easily changed by policy-makers – at least compared to other factors such as preferences.

The effect of the wage beliefs on the career choice can only be interpreted as causal, if all factors that affect both the career choice and the wage beliefs are observed. In order to justify this assumption the data has to contain extensive background information measured before the career choice took place. I use data of 14- to 15-year-olds who are about to obtain a lower secondary degree. The data covers not only educational and parental background but also measures for personality, competencies, interests and attitudes. Overall, the data contains more than 150 potential control variables. To handle such a large number of potential controls, I use post-double-selection. The data-driven procedure chooses the most important confounders and flexibly accounts for non-linear confounding. However, omitted variable bias may still be likely due to the complexity of occupational choice and the formation of wage beliefs. To analyze the impact of potential unobserved confounding, I follow a novel approach by Cinelli and Hazlett (2020). The method assesses the minimal strength that unobserved confounding needs to have in order to change the conclusion by using the impact of known, observed, strong confounders as benchmark.

The results show that, contrary to common perceptions, the wage beliefs play a positive and statistically significant role in the decision to become a nurse. I show that this effect is driven by individuals who do not become a nurse and understate a nurse's wage. The empirical results lead to two important policy implications. First, increasing the wage may help to overcome the shortage observed in many countries. Second, providing information on the (relative) wage may be a successful strategy to attract more individuals into this profession. The results of the sensitivity analysis show that potential unobserved confounders would have to be strong to overrule these conclusions.

Heterogeneous effects: Two-Stage Least Squares random forests with an application to Angrist and Evans (1998)

Chapter 3 is concerned with the estimation of heterogeneous effects. Recently, Athey et al. (2019) have generalized the concept of random forests to a general class of estimation methods that solve conditional moment conditions. They apply their method to the estimation of conditional average partial effects under exogeneity and conditional instrumental variable estimation based on the classic one-instrument formula (Wald's formula, e.g. Angrist and Pischke, 2008). In this chapter, we extend the one-instrument random forest to the case with multiple instruments, the 2SLS random forest. We work out all the expressions for estimation, sample splitting and variance estimation, and address the problem of choosing the optimal tuning parameters of an instrumental variable forest. Finally, we provide an implementation in R and C++. In the second part of the chapter, we use the 2SLS random forest to revisit the classic application of instrumental variables in Angrist and Evans (1998, Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size). They use sibling-sex composition instruments in order to investigate the effect of family size on parental labor supply. Including coarse group categories in their 2SLS regressions, they also provide a basic analysis of heterogeneity of these effects across characteristics such as mother's education or husband's earnings (e.g., low/middle/high father's income, or low/middle/high education). We revisit this question using our 2SLS random forests. Comparing the results with the estimates in Angrist and Evans (1998), we find that the general magnitude of the effects as well as basic qualitative patterns generally coincide well. However, the random forest shows in a much more detailed way, and simultaneously in more than one dimension, the exact geometry of effect heterogeneity. For example, for women with high husband's income, the loss in labor supply is small and not very sensitive to own education. In contrast, for women in poorer households, the loss in labor supply strongly depends on own education and is much larger for low levels of education compared to high levels of education. This reflects the opportunity costs of highly educated women in poorer households (in terms of foregone household income) if they do not participate in the labor market.

Latent variable modeling: Parenting styles, socioeconomic status and (non-)cognitive skills

In the last chapter, we use a latent variable model to infer latent parenting styles, the broad strategy of how parents interact with their children. Various studies analyze the role of parental investment in the development of the child. Most often, these studies focus on time and monetary investments. The choice and effects of parenting styles, another dimension of parental investment, is a rather novel topic in the economic literature.

This chapter contributes to the literature in many ways. First, we apply latent dirichlet analysis for survey data (LDA-S, Munro and Ng, 2022), a novel method which can handle a large set of measures on parent-child interactions. Among other measures, we take into account how parents monitor their child, how parents enforce their will, and how inconsistent parents are in their parenting. Therefore, we are able to describe parenting styles in more detail than previous studies. This allows us to separate styles that differ only in terms of a few, but important, dimensions. Second, the theoretical framework of LDA-S provides an economically interpretable link between parent-child interactions and parents' socioeconomic environment. Third, the model results in latent classes which easily refer to theoretical models proposed by sociologists. In this way, the data driven approach can be embedded into theoretical frameworks. Fourth, we fill the gap on the link between parenting styles and household composition. Fifth, rich data on children's (non-)cognitive skills allow us to explore the association between parenting styles and children's skills.

Applying LDA-S results in four parenting styles. Two styles closely resemble an authoritative and authoritarian style as defined by sociologists. The two other styles can be interpreted as variations of these two styles. The latent variable model shows that the choice of the parenting style is strongly associated with household income, parents' education and whether the child is an only child. In the last part of the chapter, we analyze how parenting styles contribute to the skill gap between children from different socioeconomic environments. We find that styles associated with having more than one child and having a low household income are linked with lower skills. Interestingly, parents' education is not systematically connected to parenting styles which are related to more favorable outcomes. These results give important directions for policy-makers. To reduce the skill gap one could promote styles that are associated with the most favorable outcomes. However, the effectiveness of parents in implementing certain parenting styles may depend on their personal characteristics. Our results suggest that parental skills and time resources of parents might limit the choice of the parenting style. Policy-makers could foster parents' (non-)cognitive skills or help parents to allocate their available time between children more efficiently.

Chapter 2

Average effects: The role of wage beliefs and information in the decision to become a nurse

2.1 Introduction

Due to demographic change and technological progress in medicine, the demand for skilled nurses has increased over the past decades (German Employment Agency, 2020). This trend will continue in the coming years and will further aggravate the lack of nurses. To counteract this development, it is important to analyze and to understand the occupational behavior of nurses. The existing literature discusses a series of factors that might alleviate the lack of skilled workers. These include individual preferences of (future) nurses, improving working conditions and increasing wages. I contribute to this discussion by analyzing the effect of the beliefs about a nurse's wage of young students on the probability of becoming one. This is particularly interesting for at least two reasons: First, wages are the most controversially discussed factor in the literature. Some authors identify it as a very important factor influencing labor supply decisions of nurses (Hanel et al., 2014; Doiron et al., 2014). However, others suggest that the labor supply of nurses is relatively inelastic in terms of wages. Factors such as personal attitude and working conditions seem to play a much larger role (Shields, 2004; McCabe et al., 2005).

Since there are large differences in earnings depending on the occupational choices, the economic literature on the effect of the expected wage is rich (Altonji et al., 2016). The majority of studies agree that the wage has a significant and positive effect on the career choice (e.g. Boudarbat, 2008; Montmarquette et al., 2002). Nonetheless, most studies find that preferences and interests play a larger role in career choice than the wage expectations (Beffy et al., 2012; Arcidiacono, 2004).

In line with the economic literature, the nursing literature suggests preferences and interests to be the most important factors influencing the decision to become a nurse. In particular, caring for people is identified as the key reason for choosing the profession (e.g. Wilkes et al., 2015; Petrucci et al., 2016; Matthes, 2019). Concerning the wage, several studies find that it only plays a minor role in the decision-making process (e.g. McCabe et al., 2005; Bomball et al., 2010; Cho et al., 2010). Based on these results, policy-makers might be tempted to focus on non-monetary factors to attract more young people into nursing. However, this contrasts recent work by Hanel et al. (2014) and Schweri and Hartog (2017). Schweri and Hartog (2017) examine the effect of ex-ante wage expectations on the decision to pursue a nursing degree (tertiary education) by using data on healthcare trainees (upper-secondary education) in Switzerland. Therefore, they analyze the decision on the intensive margin. Their results show that the greater ex-ante wage expectations of a nursing degree, the higher the probability to pursue such a degree later on. This indicates that higher wages may attract more students to become a high-skilled nurse. Hanel et al. (2014) estimate a model of labor supply decisions using data on individuals who hold a nursing qualification. The model accounts for the intensive and extensive margin by allowing individuals to enter and to exit occupations. As a result, they find a considerable high wage elasticity. This differs fundamentally from other work that detect very small elasticities (Shields, 2004; Andreassen et al., 2017). These differences can be fully explained by the frequent neglect of the extensive margin and the exclusive analysis of the intensive margin. Although Hanel et al. (2014) do not account for the choice of becoming a nurse, their results suggest that wages may heavily drive the career choice, i.e. a decision on the extensive margin.

The focus of this paper is on beliefs about wages. Such beliefs may affect educational choices and could be easily changed by policy makers - at least compared to other factors such as preferences. For example, Jensen (2010) analyzes preceived returns to secondary schooling of students in the Dominican Republic. He finds that the expected returns are underestimated. By providing information, students completed more years of education. At the same time, Dante et al. (2013) find that students who do not become a nurse basically know nothing about it (e.g. initial wages).

I use extensive panel data of former German 9th graders. It contains information on the wage that young students think a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician earns. This information enables me to estimate the effect of the beliefs about a nurse's wage on the probability to become one. Moreover, I estimate the effect of other factors (e.g. social orientation) on the probability of choosing the profession of a nurse. This allows to assess the magnitude of the impact of the beliefs about a nurse's wage and to fit my results into the recent literature.

In addition, the data contains extensive background information on the individuals measured in 9th grade, i.e. before their occupational decision took place. This covers not only educational and parental background but also measures for personality, competencies, interests and attitudes. Overall, the data allows to observe over 150 characteristics. By applying the lasso proposed by Tibshirani (1996), a method that draws coefficients towards zero or exactly to zero, I am able to select the relevant controls and to model non-linearities in confounding. However, the lasso is tailored to choose variables such that an outcome is precisely predicted. Therefore, it cannot be applied directly for variable selection, when the aim is to estimate a partial effect. As a solution, Belloni et al. (2012, 2014b) propose the *post-double-selection*, which is a two-step procedure to identify relevant controls and their functional form. To interpret the estimated effect as causal, I need to assume that no factors affecting the dependent variable and the variable of interest remain unobserved (unconfoundedness). Despite the ability to condition on a rich set of controls and flexibly model their functional form, this assumption is very strict and likely to be violated in some way. To mitigate the concerns about omitted variable bias and to get an idea about its consequences, I follow a novel approach by Cinelli and Hazlett (2020). For linear models, they propose to assess the minimal strength that unobserved confounding needs to have on the wage beliefs and on the career choice in order to change the conclusion. To this end, Cinelli and Hazlett (2020) propose a procedure for benchmarking based on observed covariates. The knowledge about main predictors for career choice or the wage beliefs is the crucial premise for the benchmarking to be valuable. Fortunately, literature on determinants of wage expectations and factors driving young people into nursing is rich. Thus, credible benchmarking on observed covariates is possible.

This is by far not the only approach to assess the sensitivity of results. Several approaches exist. For example, in an influential paper, Oster (2019) proposes a method for computing the relative degree of selection on observed and unobserved variables to match a given treatment effect (which is zero, for example). However, the degree of relative selection is hard to grasp and interpret. Moreover, the computation requires the specification of the unknown maximum explanatory power that can be achieved by a regression of the outcome on both observed and unobserved controls. By contrast, the method by Cinelli and Hazlett (2020) only relies on quantities that are easy to understand and interpret.

My results show that higher beliefs about a nurse's wage increase the probability to become a nurse. In line with recent literature, individual preferences play a larger role than the beliefs. Since the career choice is a decision on the extensive margin, my results are also consistent with those of Hanel et al. (2014). The importance of the extensive margin is further underlined by the result that effects are driven by young people who do not become a nurse and underestimate the wage. This means that the public perception of wages in nursing is too low. Therefore, nursing is less attractive than other occupations for which wages are not systematically understated. To combat the lack of skilled nurses, policy-makers can make the profession more attractive by increasing the beliefs about a nurse's (relative) wage.

The remaining paper is structured as follows. Section 2 outlines the methods applied in the empirical analysis and briefly describes the data, the wage belief measures as well as the control variables. In section 3, I present and discuss the main results of my analysis. Section 4 concludes.

2.2 Methods

2.2.1 Empirical strategy

Post-Double-Selection

The partial effect of the wage belief w_i on the probability to become a nurse is estimated by a partially linear model

$$y_i = \beta w_i + g(x_i) + \zeta_i, \tag{2.1}$$

where $y_i \in \{0, 1\}$ denotes the binary choice to become a nurse. The function $g(x_i)$ is unknown and potentially complicated. I approximate it by a linear combination that may include higher order polynomials and interactions

$$g(x_i) = x_i'\theta_y + r_{yi}, \tag{2.2}$$

where r_{yi} is an approximation error. The aim is to estimate β. However, it is a difficult task to define a set of variables to be included in the model and to model their functional form (i.e. what polynomials and interactions to include). Therefore, I rely on data-driven variable selection and follow the *post-double-selection* (PDS) approach proposed by Belloni et al. (2012, 2014b). The lasso is a shrinkage method that imposes a penalty on the size of the coefficients, i.e. shrinks them towards zero or exactly to zero. This prevents models with many variables that are correlated with each other from overfitting (Hastie et al., 2009). The lasso is defined as

$$\hat{\gamma}^{lasso} = \arg\min_{\gamma} \Bigg\{ \underbrace{\frac{1}{2} \sum_{i=1}^{N} \left(y_i - \gamma_0 - \sum_{j=1}^{p} x_{ij}\gamma_j \right)^2}_{\text{residual sum of squares}} + \underbrace{\lambda \sum_{j=1}^{p} |\gamma_j|}_{\text{penalty term}} \Bigg\}, \tag{2.3}$$

where $\sum_{j=1}^{p} |\gamma_j|$ imposes the penalty on the size of the coefficients and the parameter $\lambda \geq 0$ controls the magnitude of the punishment.

A naive approach to estimate β would be to apply the lasso estimator to equation (3.11) and to exclude β from the penalty term such that it is enforced to stay in the model. Afterwards one might use a least-squares regression of the outcome on w_i and controls with non-zero coefficients. However, this approach leads to biased estimates because of omitted variables. The lasso is designed to learn a forecasting rule of y_i given w_i and x_i and not to learn about the relationship between y_i and w_i given controls x_i (Belloni et al., 2014a). Therefore, lasso cannot be used off the shelf for the estimation of partial effects. As a solution, Belloni et al. (2012, 2014b) propose an intuitive and easy-to-implement procedure. First, the lasso is used to estimate a model predicting the outcome given x_i in equation (2.4) and a further model predicting the wage beliefs given x_i in equation (2.5)

$$y_i = x_i'\pi + \epsilon_i, \tag{2.4}$$

$$w_i = x_i'\theta_w + \nu_i. \tag{2.5}$$

Subsequently, all variables with non-zero coefficients in either of the two models are kept as control variables in order to estimate $\hat{\beta}$ in equation (3.11) by an ordinary least squares regression. This step is known as the "post-lasso". The crucial assumption under which PDS works is approximate sparsity. It states that the wage belief and the career choice can be approximated by equation (2.4) and (2.5) using only a small number of covariates relative to the sample size. Note, that approximate sparsity is also implicitly assumed in conventional OLS analysis where no double selection by lasso takes place. Additional variables that are considered as important for ensuring robustness, can be included (amelioration set). The condition is that the amelioration set is not substantially larger than the number of variables chosen via the lasso (Belloni et al., 2014b).

The choice of λ is of importance. With the aim of prediction, standard lasso applications choose λ by cross-validation. However, this analysis aims to estimate a partial effect. If λ is too large, only a few variables are selected and omitted variable bias may occur. If λ is too small, the number of variables is very large such that overfitting may become an issue. Therefore, I follow Urminsky et al. (2016) and use $\lambda = 1.1\sigma_R \frac{1}{\sqrt{N}} \Phi^{-1}(1 - \frac{0.1}{ln(N)2p})$, where N is the number of observations, p is the number of potential controls, Φ^{-1} denotes the inverse cumulative function of the standard normal distribution and σ_R the standard deviation of the residuals of the model. Finally, it is important to note that the chosen variables are not interpretable since selection depends on the sample (Mullainathan and Spiess, 2017).

Sensitivity

In order to interpret the partial effect $\hat{\beta}$ as causal, I need to rely on the assumption of unconfoundedness $\mathbb{E}[\zeta_i | w_i, r_{yi}, x_i] = 0$. It states that all factors that affect the choice y_i and the wage belief w_i at the same time must be contained in $g(x_i)$. Even though I have access to an extensive set of potential controls x_i, bias due to unobserved confounders may be likely. For example, covariates measuring the interests of the individuals might not fully capture all relevant aspects but only a share of it. Further, it cannot be ruled out, that some factors may remain fully unobserved. Moreover, the assumption of approximate sparsity may be violated. There may exist covariates that are not selected by lasso but affect both, the wage belief and the decision to become a nurse. To analyze the sensitivity of the results due to potentially unobserved (non-)linear confounding factors z, I make use of a procedure proposed by Cinelli and Hazlett (2020). In a nutshell, they propose to assess the sensitivity of the estimates by analyzing whether a confounder is strong enough to change the conclusion if it is as strong as a very good predictor of y or w.

Conventionally, the omitted variable bias can be written as $\widehat{bias} = \hat{\gamma}\hat{\delta}$. Hence, $\hat{\gamma}$ describes the difference in the linear expectation of the outcome if z_i changes by one unit, holding everything else constant and $\hat{\delta}$ describes the difference in linear expectation of the confounder if the variable of interest changes by one unit, holding everything else constant (Cinelli and Hazlett, 2020). Arguing that both quantities $\hat{\delta}$ and $\hat{\gamma}$ are hard to grasp, Cinelli and Hazlett (2020) write the conventional omitted variable bias formula in terms of partial R^2 measures. Those are easier to interpret and can be exploited for further analysis. Denote $\hat{\beta}_{obs}$ as the observed estimated effect and $\hat{\beta}$ as the estimated effect from a model controlling unobserved confounding factors, i.e. $\hat{\beta} = \hat{\beta}_{obs} - \widehat{bias}$. Then, they show that

$$|bias| = \hat{se}(\hat{\beta}_{obs})\sqrt{\frac{R^2_{y\sim z|w,x}R^2_{w\sim z|x}}{1-R^2_{w\sim z|x}}df}, \tag{2.6}$$

where df defines the degrees of freedom, $R^2_{y\sim z|w,x}$ stands for the partial R^2 of regressing y on z after controlling for w and x and $R^2_{w\sim z|x}$ denotes the partial R^2 of regressing w on z after controlling for x. Further, the standard error of $\hat{\beta}$ can be written as

$$\hat{se} = \hat{se}\left(\beta_{obs}\right)\sqrt{\frac{1-R^2_{y\sim z|w,x}}{1-R_{w\sim z|x}}\left(\frac{df}{df-1}\right)}, \tag{2.7}$$

and the adjusted t-statistic is defined as $t_{adj} = \hat{\beta}/\hat{se}$. Applying these definitions, $\hat{\beta}$, $\hat{se}$ and t_{adj} can be computed by substituting reasonable values for $R^2_{y\sim z|w,x}$ and $R^2_{w\sim z|x}$, i.e. the strength of confounding, into equations (2.6) and (2.7). However, actual knowledge about the absolute strength is seldom available. As a solution, Cinelli and Hazlett (2020) argue that the researcher is often able to make a statement on the relative strength of potential unobserved confounding, e.g. z cannot account for as much variation of the outcome as some observed covariate x_j. There are several ways to formalize such claims. I follow Cinelli and Hazlett (2020) and claim that I measure the key determinant of y and w such that the omitted variable cannot explain as much residual variance in y or w as this determinant. Define

$$k_w = \frac{R^2_{w\sim z|x_{-j}}}{R^2_{w\sim x_j|x_{-j}}} \tag{2.8}$$

$$k_y = \frac{R^2_{y\sim z|x_{-j},w}}{R^2_{w\sim x_j|x_{-j},w}}, \tag{2.9}$$

where x_{-j} is a vector including all variables contained in x, excluding x_j. The ratios k_w and k_y show how much of the variance in w or y is explained by z relative to the explanatory power of x_j, conditional on all other covariates. In this paper $k_w = k_y = 1$, i.e. I consider the impact of a confounder z that is as strong as x_j. Given k_w and k_y, Cinelli and Hazlett (2020) show that

$$R^2_{w\sim z|x} = k_w f^2_{w\sim x_j|x_{-j}} \quad R^2_{y\sim z|w,x} \leq k_y \eta^2 f^2_{y\sim x_j|x_{-j},w}, \tag{2.10}$$

where η is a scalar that depends on k_w, k_y, and $R^2_{w\sim x_j|x_{-j}}$. Furthermore, $f^2_{w\sim x_j|x_{-j}}$ denotes partial Cohen's f of w on x_j and $f^2_{y\sim x_j|x_{-j},w}$ denotes partial Cohen's f of y on x_j.[1] Cinelli and Hazlett (2020) have shown that these robustness results are exact for a single linear confounder and conservative for multiple, possibly nonlinear, confounding factors.

It is important to emphasize that this bounding procedure heavily relies on the choice of the benchmark variable x_j. If it is not true that x_j is a key predictor of the outcome or treatment, the bounding is pointless. Hence, domain knowledge is necessary (Cinelli and Hazlett, 2020). In the following, I choose observed covariates that are often discussed in the literature. First, bounding is based on social orientation. It is the key characteristic of those who become a nurse (e.g. Matthes, 2019), while preferences are generally a decisive factor in career choice (e.g. Arcidiacono, 2004). In addition, interests also play an important role in the formation of expected wages (Wiswall and Zafar, 2015). Second, the professions of the parents play an important role in the occupational choice (e.g. Knoll et al., 2017). Therefore, the results are bounded by an indicator that indicates whether at least one of the parents is a nurse. Moreover, parents in nursing might inform their children about the wages in nursing. Third, an indicator for gender is considered. Females become nurses much more often than males (Speer, 2020). Moreover, gender also plays a crucial role in wage expectations: females expect lower wages than males (e.g. Brunello et al., 2004; Fernandes et al., 2020). Fourth, (perceived) ability determines the expected wages (Brunello et al., 2004). Therefore, a measure for ability, namely metacognition, is used to bound the results. Note, that these variables have to be part of the model in order to use them as benchmark variables. Hence, the amelioration set contains these four variables, to ensure that they are not excluded by data-driven variable selection.

2.2.2 Data

This study uses Starting Cohort Four (SC4) of the German National Educational Panel Study (NEPS). The survey collects data on young people who attended the 9th grade in German

[1] Note that Cohen's f^2 is defined as $f^2 = \frac{R^2}{1-R^2}$.

regular schools in 2010 and has been followed since (Blossfeld and von Maurice, 2011). This includes grammar schools, middle secondary schools, lower secondary schools, comprehensive schools, and schools offering all tracks of secondary education except the grammar schooltrack. Since becoming a nurse requires a vocational training, all degrees that can be obtained at these schools are sufficient to be admitted. For several reasons the data is highly suitable for investigating the role of beliefs about a nurse's wage in the decision to become one. Since the data is available from 2010 to 2016, the transition from school to further education can be observed in great detail and no retrospective information has to be used. The following analysis is based on a cross-section of the panel and focuses on the choice of the first occupational training, which certainly has a relevant impact on the further life course. Beyond that, the individuals are asked to state their beliefs about the monthly salary of a nurse, a hairdresser, a motor vehicle mechanic, a bank clerk, a teacher and a physician: *"Now we are also interested to know how high you think the income is in certain professions. What do you think the monthly income as a [...] is?"*. Consequently, the question at hand captures knowledge about average wages, knowledge of wages according to collective agreements, but also wrong beliefs due to the lack of information or wrong perceptions of wages. In order to define a measure for the beliefs about a nurse's wage, the stated wages of all six occupations are ranked from lowest to highest. If the wage cannot be assigned unambiguously due to ties, the mean rank is assigned such that the sum of ranks is preserved. Formally, I define the $i-th$ individual's rank of a nurse's wage as

$$\text{rank}_i^{\text{nurse}} = 1 + \sum_{w_i \in \{S_i \setminus w_i^{\text{nurse}}\}} \mathbb{1}\left(w_i < w_i^{\text{nurse}}\right) + 0.5 \times \mathbb{1}\left(w_i = w_i^{\text{nurse}}\right), \tag{2.11}$$

where $\mathbb{1}(\cdot)$ denotes an indicator function that takes the value 1 if the expression in the parentheses is true, S_i is the set of surveyed wage beliefs and w_i^{nurse} is the belief about a nurse's wage. Two further measures are defined as the ratio between individual's i beliefs about a nurse's wage and maximum as well as minimum stated wage

$$\text{relwage}_i^{\text{nurse, max}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{max}}}, \tag{2.12}$$

$$\text{relwage}_i^{\text{nurse, min}} = \frac{w_i^{\text{nurse}}}{w_i^{\text{min}}}. \tag{2.13}$$

In addition, I use the belief of a nurse's absolute wage.

Based on the ranking measure in equation (2.11), I can easily assess how close the relative wage beliefs are to reality by computing the deviation from the true ranks. The median wages reported by German Employment Agency (2018) provide the basis for the true rank. According to this source of information, the following *true* ranking from lowest to the highest wage was established: (1) hairdresser, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher and (6) physician. The ranking is utilized to construct a measure that captures the knowledge about relative wages by adding the absolute deviations of the stated rank of each occupation

$$\begin{aligned} \text{rank}_i^{\text{abs. dev.}} =& |\text{rank}_i^{\text{hairdresser}} - 1| + |\text{rank}_i^{\text{mechanic}} - 2| + |\text{rank}_i^{\text{nurse}} - 3| + \\ & |\text{rank}_i^{\text{bank clerk}} - 4| + |\text{rank}_i^{\text{teacher}} - 5| + |\text{rank}_i^{\text{physician}} - 6|, \end{aligned} \tag{2.14}$$

where the ranks of each occupation are computed the same way as the rank of a nurse's wage. Additionally, I can construct indicators that show whether the wage rank of a nurse is overestimated, correctly estimated or underestimated

$$\text{rank}_i^{\text{nurse, over}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} > 3\right), \tag{2.15}$$

$$\text{rank}_i^{\text{nurse, correct}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} = 3\right), \tag{2.16}$$

$$\text{rank}_i^{\text{nurse, under}} = \mathbb{1}\left(\text{rank}_i^{\text{nurse}} < 3\right). \tag{2.17}$$

Besides information on wage beliefs, there are other potentially important factors available that may drive young people into or out of nursing (see Wohlkinger et al., 2011). This enables me to assess the importance of the wage belief by comparing the effect with other effects estimated in the literature. A large share of recent work finds that those who become nurses do not rate the importance of economic factors as important as those who choose another profession. Therefore, I use a measure of the importance of economic factors (i.e. risk of unemployment and financial aspects) in choosing a career. Moreover, helping others is considered to be one of the main driving forces in choosing a nursing profession. Hence, a measure that quantifies the amount of social interests is used. Finally, I estimate a model that uses an indicator of self-assessed importance of comfort (i.e. physical working conditions).

Additionally, extensive information about the background, personal characteristics and the (social) environment of the individuals are measured before they have decided on a career. All potential controls are summarized in table A1. The exclusion of observations with at least one missing value would lead to a substantial loss of observations. Therefore, I impute missing values by chained equations (van Buuren and Groothuis-Oudshoorn, 2011). To estimate the impact of the wage belief via equation (3.11), I account for non-linearities in confounding by interacting all variables with each other and by additionally including fifth-order polynomials of non-binary covariates. As a result, 13.878 potential controls are available.

After excluding individuals with extreme beliefs, missing values in variables of interests[2] or with too many missing observations in general[3], I observe 7089 individuals that transition from school to occupational training, of whom 238 chose nursing.

2.3 Results

2.3.1 The role of wage beliefs

Descriptive evidence

First, I provide some insights about the univariate relationship between the wage beliefs and the choice whether or not to become a nurse. Table 2.1 depicts the shares of the nurse's wage ranks reported by nurses and others. The ranks of both groups seem to follow a similar general

	Rank of a nurse's wage					
	1	2	3	4	5	6
nurses	5.88	39.07	36.13	13.87	4.62	0.42
others	14.86	48.65	27.63	6.38	1.96	0.53
all	14,55	48,32	27,92	6.63	2.05	0.52

Table 2.1 – Distribution of the wage rank

The table depicts the share of the beliefs about a nurse's wage rank by nurses and others. For the sake of clarity, in the case of ties, the lower rank is reported.

pattern. Both most often believe nursing to be the second and the third rank. Further, both rarely believe that nurses have the lowest earnings or state a rank higher than three. However, the specific patterns strongly differ. The respective shares of future nurses who state a rank larger than two exceed the shares of the others. Furthermore, the shares for the first and second rank are smaller among nurses. To gain further insights into the differences in wage ranks, table 2.2 presents the mean differences of the wage rank by nurses and others. Table 2.2 reveals that, on average, nurses state a lower rank of the wage earned by hairdressers or

[2] That is wage belief, economic and social orientation and importance of comfort.

[3] Precisely, I drop observations with over 18 % missing values - that is the 90 % quantile of the share of missing entries

mechanics than others, but state a larger rank of a nurse's wage. Interestingly, the average ranks of a bank clerk, teacher and physician are not significantly different between nurses and others. This means, that beliefs only differ for occupations with lower wages.

	Mean of others	Mean of nurses	Difference	P-values of test for differences in means		
Stated rank of a...				H_0 : diff. < 0	H_0 : diff. $= 0$	H_0 : diff. > 0
... hairdresser	1.29	1.16	0.12	1.00	0.00	0.00
... mechanic	2.74	2.56	0.18	1.00	0.00	0.00
... nurse	2.46	2.84	-0.39	0.00	0.00	1.00
... bank clerk	4.53	4.49	0.04	0.73	0.54	0.27
... teacher	4.41	4.35	0.06	0.83	0.33	0.17
... physician	5.56	5.58	-0.14	0.40	0.80	0.60

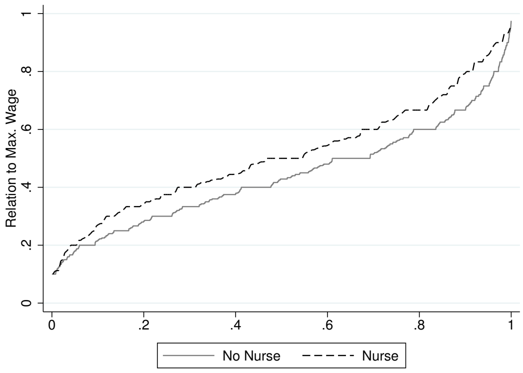
Table 2.2 – Differences in mean wage ranks by nurses and others

The table depicts the means of the wage ranks by future nurses and others together with their differences. Further, to assess if differences are statistically significant, t-tests are conducted.

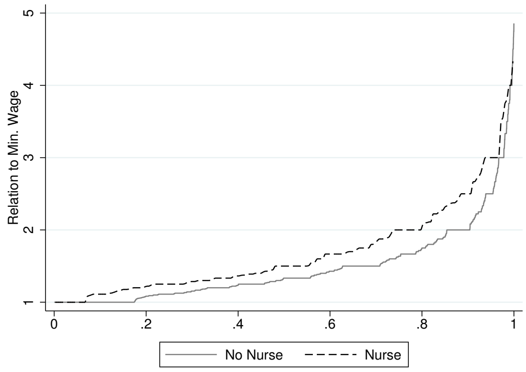
In figures 2.1a-2.1c the remaining measures of wage beliefs are depicted in quantile plots. Figure 2.1a shows the ratio of the belief of a nurse's wage and the highest stated wage belief as defined in equation (2.12). Differences in relative wage beliefs between those who become a nurse and those who do not, are very clear. Except for the lower and upper end, future nurses state higher wages. Similarities in lower and upper ends indicate that extreme beliefs do not differ systematically between groups. In figure 2.1b, the distribution of the ratio between the belief of a nurse's wage and the lowest stated wage belief is shown. For non-future nurses the extreme value at the upper end of the distribution is prominent. However, this appears to be an outlier. In the remaining distribution, the relative wage belief is larger for future nurses. Differences increase in higher quantiles. In summary, descriptive evidence consistently suggests that future nurses state a higher relative wage than those who do not become a nurse. As figure 2.1c reveals, not only the beliefs about the nurse's relative wage is higher for future nurses. At least in quantiles in the middle, the absolute wage belief is also slightly higher.

Results of PDS

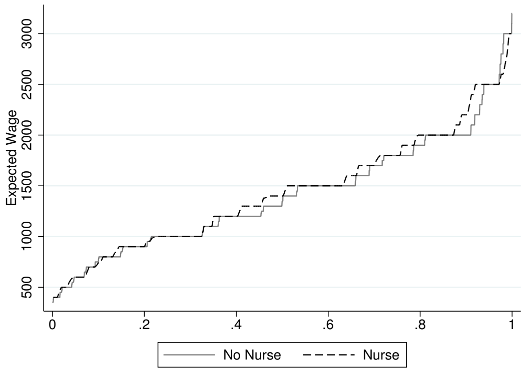
The observed descriptive differences may be caused by confounding. For example, those who have no interests in becoming a nurse may state rather low wage beliefs (e.g. Wiswall and Zafar, 2015). However, to draw more valuable policy implications, it is useful to analyze the



(a) Ratio of belief of a nurse's wage and maximum stated wage



(b) Ratio of belief of a nurse's wage and minimum stated wage



(c) Beliefs of the absolute wage of a nurse

Figure 2.1 – Continuous wage belief measures

Each panel depicts a quantile plot of one wage measure. The ratio of the belief of a nurse's wage and maximum stated wage belief is defined as $\text{relwage}_i^{\text{nurse, max}} = w_i^{\text{nurse}}/w_i^{\text{max}}$ and ratio of the belief about a nurse's wage and minimum stated wage belief is defined as $\text{relwage}_i^{\text{nurse, min}} = w_i^{\text{nurse}}/w_i^{\text{min}}$

role of beliefs in becoming a nurse, given the characteristics of the individuals (i.e. equally interested, same background, same skills, etc.). As described in section 2, I tackle this issue by using PDS to estimate the partial effect of the wage belief on the probability to become a nurse. The results are depicted in table 2.3. Each of the three columns shows the results of an unconditional model (single OLS) and the post-lasso (a conditional OLS model with controls chosen by double-selection). In the first column, future nurses are compared to all remaining young people. However, this neglects the heterogeneity of the effect of (relative) wage beliefs. Individuals who are interested in becoming a nurse, e.g. chose a similar occupation, may be more responsive to wage beliefs compared to those who have no interest in nursing at all. Hence, in the remaining columns the sample is restricted regarding the career choices. I compare future nurses to (2) young people who opted for vocational training and (3) individuals who chose a social field.[4] Each panel of the table depicts the results of one of the four measures described above. The measures are standardized to have mean 0 and standard deviation 1 such that the results can easily be compared with other factors in the subsequent section.[5]

	(1)		(2)		(3)	
	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Rank of nurse's wage						
	0.014***	0.010***	0.015***	0.012***	0.047***	0.035***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
p	-	41	-	25	-	13
Nurse's wage/highest wage						
	0.011***	0.006***	0.009***	0.006*	0.036***	0.023***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
p	-	38	-	16	-	20
Nurse's wage/lowest wage						
	0.009***	0.008***	0.011***	0.011***	0.025***	0.022***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.007)
p	-	24	-	19	-	9
Nurse's absolute wage						
	0.002	0.008***	0.006**	0.011***	0.021**	0.028***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.009)
p	-	29	-	23	-	13
N	7098		4452		1616	

Table 2.3 – Beliefs about a nurse's wage

The table depicts the results of the effect of the wage belief on the decision to become a nurse. The rank of nurse's wage is defined in equation (2.11) and the ratio of the nurse's wage and highest/lowest stated wage is defined in equation (2.12) and (2.13) respectively. Standard errors are shown in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

The first panel shows the estimated effects of a nurse's rank on the probability to become a nurse. Column one compares future nurses to all the remaining individuals in the data. As expected from the descriptive results in tables 2.1 and 2.2, results of the unconditional model show that an increase of the rank by one standard deviation is associated with a statistically significant increase in the probability to become a nurse by 1.4 percentage points. When

[4]Note, comparing future-nurses with youths who chose a vocational training, i.e. did not choose to visit a university, is motivated by the German education system. Education after school is divided into academic and vocational training, whereas nursing belongs to the latter kind.

[5]A discussion of the magnitude of the estimates will be given in the next section.

relevant controls are added, I still observe a statistically significant change by 1 percentage point. The slightly smaller coefficients of the post-lasso compared to the single OLS model shows that those, who are prone to become a nurse (e.g. having parents that are nurses and young people that have a social attitude) have higher relative wage beliefs. The results change only slightly with regard to the comparison group in column two. Comparing future-nurses with individuals who chose a social field, I find a much larger effect in both the unconditional and conditional model. These results hint to heterogeneity in the effect, i.e. larger effects for those who chose a more similar field. Interestingly, the number of included controls is much smaller when the comparison group only consists of individuals who chose a vocational training or a social field. However, this is expected since the sample size is much smaller. Thus, λ becomes larger and draws the coefficients of the lasso models stronger towards zero. Further, the sample in column 1 is more heterogeneous than the ones in column 2 and 3. Therefore, fewer variables may be required to explain differences.[6]

The second panel contains the results of the effect of the ratio of the beliefs about a nurse's wage and the highest stated wage. Results for the entire sample in the first column show that even after controlling for relevant confounders, I find a statistically significant and positive effect on the probability to become a nurse. Similar to results of the wage rank, the coefficient in the post-lasso model is smaller than in the unconditional model. The effect stays positive and significantly different from zero when the comparison group is changed. For those who chose a social field in column 3, the effect is again much larger. This suggests effect-heterogeneity.

The following panel shows the impact of the ratio of the beliefs about the wage of a nurse and the lowest stated wage on the decision to become a nurse. Comparing those who become nurses with all other individuals, results of the unconditional model show that an increase in the relative wage increases the probability to become a nurse by 0.9 percentage points. When relevant controls are included, the probability increases by 0.8 percentage points. Just like the estimates in the first and second panel, the results also indicate heterogeneity. The effects become even larger when the comparison group consists of young people who chose a social field. Figure 2.1b revealed that some non-future nurses have an outlying high ratio between a nurse's wage and the lowest stated wage. These outliers do not qualitatively alter the results. Their exclusion, if anything were to change, would cause even larger effects.

[6] Note that regarding the choice of variables, mainly interactions are chosen. This hints to strong non-linearities in confounding and stresses the importance of flexible choice of controls.

The last panel of table 2.3 contains the results of the impact of the absolute wage belief on the probability to become a nurse. After conditioning on relevant controls, I find a statistically significant effect of the absolute wage on the probability to become a nurse that stays significant when the composition of the comparison group is changed. Interestingly, in all columns, the coefficient in the model with no controls is smaller than in the models including controls. Whereas those who are prone to become a nurse expect higher relative wages, they have a lower absolute wage belief. As observed for the relative wage measures, the effect becomes larger when the comparison group is restricted to individuals who chose a similar occupation.

In summary, results in table 2.3 show that even after conditioning on an extensive set of relevant controls and accounting for non-linearities in confounding, the beliefs about a nurse's wage affects the probability to become a nurse. This holds true for both the relative and absolute wage. Moreover, I find evidence that effects are heterogeneous. These effects are stronger for individuals who are more prone to choose a social occupation.

Sensitivity regarding omitted variable bias

It is natural to ask whether the positive association between the choice to become a nurse and the wage belief can be interpreted as a causal one. Despite the large set of potential controls, the presence of omitted variable bias may be likely due to the complexity of occupational choice and the formation of wage beliefs. To analyze the impact of potential unobserved confounding, I conduct a sensitivity analysis as described in section 2. The results are shown in table 2.4. Each panel depicts the results of one wage measure. As discussed above, I follow Cinelli and Hazlett (2020) and make use of observed covariates that are strong predictors of the occupational choice or the wage belief to analyze the consequence of potential omitted variables. Columns 2 to 5 display the adjusted estimate $\hat{\beta}$ and t-statistic t_{adj}. They are obtained when an unobserved confounder, that explains as much variance in y and w as predefined benchmark variables, is additionally controlled for. As mentioned above, the variables used to bound the consequences of omitted variable bias are (1) gender, (2) parents' occupation, (3) social interests and (4) metacognition. The first column indicates whether $\hat{\beta}$ and t_{adj} are computed using only one variable or whether it is based on all transformations in the model involving the variable. For example, the adjusted estimate and t-statistic with no transformations are computed under the assumption that an unobserved confounder that is as strong as gender exists. In contrast, $\hat{\beta}$ and t_{adj} including transformations are computed by assuming that an unobserved confounder exists, that is as strong as gender and all interactions that are included in the model and where gender is involved in (e.g. interaction between gender and math-skills, gender and social interests, etc.). I expect benchmarks that account for transformations to have a much larger impact than benchmarks of single covariates, because many transformations are chosen by the lasso. The last column shows the results that would have been obtained if an omitted variable that explains as much as all four variables together had been controlled for.

	(1)	(2)		(3)		(4)		(5)		(6)	
	Including transformations	Gender: female		Parents nurses		Social interests		Metacognition		All	
		$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}	$\hat{\beta}$	t_{adj}
Rank of nurse's wage											
nurse vs. all	No	0.009	4.43	0.009	4.39	0.009	4.38	0.010	4.45	0.009	4.28
	Yes	0.008	3.92	0.008	3.87	0.008	3.97	0.009	4.40	0.005	2.38
nurse vs. vocational training	No	0.012	3.75	0.012	3.78	0.012	3.75	0.012	3.79	0.011	3.62
	Yes	0.010	3.19	0.010	3.19	0.010	3.26	0.011	3.57	0.004	1.20
nurse vs. social field	No	0.035	3.94	0.035	3.97	0.035	3.97	0.035	3.97	0.034	3.85
	Yes	0.035	3.94	0.031	3.59	0.035	3.97	0.035	3.96	0.030	3.43
Nurse's wage/highest wage											
nurse vs. all	No	0.006	2.77	0.006	2.79	0.006	2.74	0.006	2.79	0.006	2.62
	Yes	0.005	2.44	0.005	2.30	0.005	2.29	0.006	2.79	0.002	1.12
nurse vs. vocational training	No	0.006	1.85	0.006	1.90	0.006	1.87	0.006	1.90	0.005	1.70
	Yes	0.005	1.59	0.005	1.51	0.004	1.40	0.006	1.88	0.001	0.30
nurse vs. social field	No	0.023	2.56	0.023	2.57	0.022	2.44	0.023	2.58	0.021	2.40
	Yes	0.023	2.56	0.020	2.27	0.022	2.44	0.023	2.58	0.018	2.06
Nurse's wage/lowest wage											
nurse vs. all	No	0.008	3.57	0.007	3.48	0.008	3.57	0.008	3.58	0.007	3.43
	Yes	0.007	3.50	0.007	3.37	0.007	3.15	0.007	3.54	0.005	2.62
nurse vs. vocational training	No	0.010	3.26	0.010	3.18	0.010	3.27	0.010	3.27	0.010	3.10
	Yes	0.010	3.22	0.010	3.04	0.009	2.84	0.010	3.23	0.008	2.46
nurse vs. social field	No	0.021	2.90	0.022	3.01	0.020	2.77	0.022	2.99	0.019	2.61
	Yes	0.021	2.90	0.021	2.95	0.020	2.77	0.022	2.99	0.017	2.38
Nurse's wage											
nurse vs. all	No	0.008	3.52	0.007	3.49	0.008	3.51	0.008	3.52	0.007	3.43
	Yes	0.006	2.67	0.007	3.21	0.007	3.18	0.007	3.52	0.003	1.20
nurse vs. vocational training	No	0.011	3.43	0.011	3.46	0.011	3.42	0.011	3.43	0.011	3.26
	Yes	0.008	2.54	0.010	3.18	0.010	3.10	0.011	3.42	0.003	0.81
nurse vs. social field	No	0.028	3.12	0.027	3.07	0.028	3.12	0.028	3.12	0.026	2.96
	Yes	0.025	2.84	0.025	2.85	0.028	3.12	0.028	3.12	0.019	2.14

Table 2.4 – Sensitivity due to omitted variables

The table depicts the results on the sensitivity of the effect of the wage beliefs on the decision to become a nurse. The adjusted t-statistic is based on the adjusted estimate $\hat{\beta}$ and adjusted standard errors $\hat{se}$. $R^2_{w\sim z|x}$ and $R^2_{y\sim z|w,x}$ are computed as defined in equation (2.10) setting $k_w = k_y = 1$, i.e. unobserved confounders that are as strong as the considered benchmark variables.

The first panel depicts the sensitivity of the results regarding the rank of a nurse's wage. The adjusted estimate only decreases slightly and equals 0.009, provided that there exists an unobserved confounder as strong as gender for which is additionally controlled. The change in the adjusted t-statistic is very small such that results stay significant at a 1% significance level. Confounders as strong as parent's occupation, social interests and metacognition only lead to minor changes. Even if I additionally control for a confounder that is as strong as all four benchmark variables combined, the conclusion that the rank of a nurse's wage significantly affects the choice to become a nurse is still valid. As expected, adjusted estimates $\hat{\beta}$ are drawn to zero by a larger amount when transformations are included. Nonetheless, these changes are small. The effect decreases to 0.004 if I control for a confounder that is as strong as all four benchmark variables together and includes all their transformations. It remains significant at the 5%-level. A change in the comparison group leads to similar robust results. The only noteworthy change in the conclusion occurs when the comparison group consists only of those who chose a vocational training. It is caused by a confounder that is as strong as all four benchmark variables including their transformations. The adjusted t-statistic shows that if such a confounder exists, there is no statistically significant effect anymore.

The second panel displays the sensitivity of the results on the ratio between the beliefs about a nurse's wage and the highest wage. The results for the entire sample show that only controlling for a confounder that is as strong as all four benchmark variables and their respective transformations has an impact that is large enough to change the conclusion. The effect decreases to 0.002 and is not statistically significant. The sensitivity analysis reveals that the estimated effect is sensitive when the comparison group consists of individuals undergoing a vocational training. A confounder as strong as single variables is not strong enough to change the conclusion. However, a cofounder as strong as gender, parental occupation or interests together with their respective transformations leads to an effect that is not statistically significant different from zero. It is evident that a confounder, as strong as all four benchmark variables combined and including their transformation, leads to an insignificant effect too. The result of comparisons between nurses and individuals in a social field are not sensitive to any of the considered strengths of confounding.

The third panel depicts the sensitivity of the ratio between a nurse's wage and the lowest stated wage. The results show that no confounder as strong as the considered benchmark variables is strong enough to change the conclusion. Even a confounder as strong as all four benchmark variables together including their transformation does not lead to remarkable changes in the estimated effect. Similar sensitivity results can also be observed when the comparison group is changed.

The last panel shows how a confounder changes the estimated effect of the beliefs about a nurse's absolute wage. A confounder as strong as a single variable does not have an impact on the estimated effect. Even a confounder as strong as all four benchmark variables together does not change the estimated effect. However, the impact of a confounder as strong as all four benchmark variables including their transformations is considerable. The effect decreases to 0.003 and is not significantly different from zero. The results are similar when the comparison group is changed. Comparing nurses to individuals who chose a vocational training, the impact of a confounder as strong as all four variables combined including their transformations is strong enough to change the conclusion. The effect substantially decreases to 0.003 and is not significantly different from zero. Choosing individuals in a social field as comparison group, none of the considered confounders is strong enough to change the conclusion.

Taking into account that the gender, the parents and the interests are key drivers of occupational choice, it can be concluded that results of the wage rank, the ratio between a nurse's wage and the lowest wage and the absolute wage are only sensitive to a confounder that is strong. Similarly, the results on the effect of the ratio between the nurse's wage and the highest wage are only sensitive regarding a strong confounder when comparing nurses to all other individuals or to those who chose a social field. However, the results are sensitive when nurses are compared to those who chose a vocational training. If a confounder with a certain strength exists, only subgroups are affected by the ratio.

2.3.2 How much do other factors matter?

In order to assess the size of the effects of the wage beliefs and to obtain some reassurance about the validity of the data, I compare the effect to other factors discussed in the recent literature. More precisely, I estimate three further PDS models using the self-assessed importance of economic factors, social interests and self-assessed importance of comfort aspects instead of the wage belief. The results are depicted in table 2.5. To compare the size of the effects with the effect of the wage belief in table 2.3, measures are standardized to have mean 0 and standard deviation 1.

I examine the impact of the importance of economic factors on young people's involvement in nursing or perhaps even their withdrawal from nursing. The results are depicted in the first panel of table 2.5. Independent of the composition of the comparison group, I cannot conclude that the importance of economic factors plays a role in the decision to become a nurse. This result replicates findings of recent research: Nurses do not give much weight to

	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Importance of economic factors						
	-0.002	0.000	-0.006*	0.000	0.008	0.015
	(0.002)	(0.003)	(0.003)	(0.004)	(0.008)	(0.010)
p	–	52	–	31	–	23
Social interests						
	0.028***	0.028***	0.042***	0.041***	0.042***	0.052***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.010)	(0.011)
p	–	79	–	58	–	25
Importance of comfort aspects						
	-0.007***	-0.009***	-0.010***	-0.012***	-0.026***	-0.036***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.009)	(0.010)
p	–	34	–	23	–	14
N	7098		4452		1616	

Table 2.5 – Relevance of other factors

The table depicts the results of the effect of other factors than the nurse's wage on the decision to become one. Standard errors are depicted in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

economic factors. However, it is noticeable that future nurses do not weight economic factors lower than other individuals.

The next panel presents the results on the role of social interests in the decision to become a nurse. As expected, the results suggest that social interests play an important role in the decision to become a nurse. This holds true when the comparison group only consists of those who chose a social field. Compared to the effect of wage beliefs, the effect of social interests is considerably larger (more than twice as large). The finding perfectly fits into both the nursing and the economic literature. It is often shown that preferences matter the most in the choice of training (e.g. Arcidiacono, 2004; Wiswall and Zafar, 2015). Therefore, the result provides some additional reassurance and further supports the results in table 2.3.

Nursing is generally known for its rather exhausting tasks and inflexible working hours. To investigate the effect of this reputation, I analyze the role of the importance comfort aspects on the probability of becoming a nurse. The results are presented in the last panel of table 2.5. They suggest that the larger the importance of comfort aspects, the lower the likelihood of becoming a nurse. Interestingly, compared to the coefficients in an unconditional model, the absolute size of the coefficients is larger in conditional models. That is, individuals that are more prone to become a nurse, put less emphasis on comfort aspects in their occupation. In summary, I find that the size of the effect of the wage belief is smaller than the role of individual interests and has about the same size as the importance of comfort aspects.

2.3.3 Assessing wage information

In general, there may be three reasons for finding a positive effect of the wage belief. Results can be driven by future nurses who overestimate wages or by non-future nurses who underestimate wages. Moreover, both can occur simultaneously. To this end, I estimate further PDS models. Instead of the wage beliefs, I use measures that capture information relative to actual wages. As discussed above, changing the rank measure to information measures defined in equations (2.14)-(2.17) is straightforward. The results of the analysis are given in table 2.6.

	nurse vs. all		nurse vs. vocational training		nurse vs. social field	
	Single OLS	Post-Lasso	Single OLS	Post-Lasso	Single OLS	Post-Lasso
Cumulative absolute deviation to true ranks						
	-0.004**	-0.006***	-0.012***	-0.008**	-0.017**	-0.023**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)
p	-	45	-	35	-	17
Nurse's wage rank overestimated						
	0.042***	0.033***	0.041***	0.038***	0.144***	0.111***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.024)	(0.025)
p	-	40	-	17	-	18
Nurse's wage rank correctly estimated						
	0.007	0.002	0.010	0.002	0.008	0.008
	(0.005)	(0.005)	(0.008)	(0.008)	(0.020)	(0.019)
p	-	15	-	10	-	9
Nurse's wage rank underestimated						
	-0.026***	-0.018***	-0.032***	-0.024***	-0.082***	-0.061***
	(0.004)	(0.004)	(0.008)	(0.007)	(0.018)	(0.017)
p	-	27	-	13	-	9
N	7098		4452		1616	

Table 2.6 – Information about nurse's wages

The table depicts the results of the effect of information about nurse's wage on the decision to become one. The measures are defined in equations (2.14)-(2.17). The true ranking is: (1) hairdresser, (2) motor vehicle mechanic, (3) nurse, (4) bank clerk, (5) teacher and (6) physician. The cumulative absolute deviation to true ranks is standardized. Standard errors are reported in parentheses. Significance of the coefficient at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, * respectively. N indicates the number of observations and p the number of chosen controls.

In the first panel, I consider a measure that captures the general level of information defined in equation (2.14). The larger the measure, the higher the deviations from the actual relative wage and consequently the lower the level of information. The coefficient is standardized in order to assess its size. The result shows that an increase in the absolute cumulative deviation by one standard deviation decreases the probability to become a nurse by 0.6 percentage points. The effect becomes even larger when nurses are compared to those who chose a more similar occupation. This means that those who become a nurse can rank surveyed wages more precisely than those who do not become a nurse. Hence, I conclude that future nurses are well informed about relative wages.

In the remaining panels the effect of overestimation in equation (2.15), correct estimation in equation (2.16) and underestimation in equation (2.17) on the probability to become a nurse is analyzed. The results show that overestimation of the nurse's rank increases and underestimation of the nurse's rank decreases the probability to become a nurse. Correct estimation does not affect the probability of becoming a nurse. These results remain statistically significant after controlling for an extensive set of confounders chosen by double-lasso-selection (e.g. general interests).

On the one hand, the results indicate that future nurses more often overestimate and less often underestimate the wage of a nurse. On the other hand, future nurses rank wages more in accordance with the true wages. Descriptive results in table 2.1 give further insights that are crucial for the interpretation of these seemingly contradicting results. The share of individuals who think that nurses earn the least among all six surveyed wages is much higher among non-future nurses than among future nurses (15% vs. 6%). In general, the share of individuals that underestimate the wage of a nurse is large (63%). Even a significant share of future nurses underestimate the wage (45%). In contrast, the share of those who overestimate a nurse's wage rank is low (13%). Furthermore, the comparison of mean wage ranks between future nurses and others in table 2.2 shows that there is only a significant difference in the rank of the three occupations with the lowest wages, i.e hairdresser, mechanic and nurse. There are no significant differences concerning occupations with higher wages, i.e. the wage rank of a bank clerk, teacher and physician. Therefore, I conclude that future nurses do not have exceptionally high wage beliefs, but individuals who do not become a nurse have beliefs that are too low. Even future nurses often underestimate the relative wages. In summary, the analysis suggests that the perception of a low wage in nursing among young people may be an obstacle to attract more individuals to nursing.

2.4 Discussion

This paper investigates the policy relevant question of whether and how wage beliefs and information influences individual career choices to become a nurse. To this end, I used state-of-the-art methods for causal machine learning (post-double-selection, Belloni et al., 2014a) and sensitivity analysis (Cinelli and Hazlett, 2020). My analysis does not use retrospective information that is potentially plagued by reverse causation, but longitudinal data following 9-th graders up to their decision whether or not to enter nursing training. I report two sets of substantive findings. First, contrary to common perceptions, individuals' beliefs about the wages in nursing do influence the probability of taking up nursing. The size of the effect is smaller than the effect of individual preferences but similar to other factors such as comfort aspects. Second, I show that understating the true rank of wages in nursing decreases the likelihood of starting a nursing career. These results suggest two important policy implications. First, boosting wages in nursing may help to overcome the shortage observed in many countries. Second, providing more accurate information about actual (relative) wages in nursing would also help to attract more individuals into this profession. The study has some limitations which have to be kept in mind. First, the occupational choice and the formation of wage beliefs are complex processes. Despite the fact that I have access to a rich set of controls and carefully choose them using data-driven variable selection, the assumption of unconfoundedness may be violated. To mitigate this concern, I conduct a sensitivity analysis. Although the existence of unobserved confounders cannot be ruled out, results show that potential unobserved confounders would have to be strong to overturn the conclusions. Second, I have access to wage beliefs on six occupations from a wide range of fields. These occupations are well known and are often chosen by young students. However, one could argue that it might be difficult to draw conclusions on the relative wage beliefs that can be generalized since wage beliefs on only six occupations are available. Yet, the absolute wage beliefs also increase the probability to become a nurse. To address these concerns, further research - preferably (quasi-)experimental studies - on the effect of wage beliefs on the probability to become a nurse is needed.

Chapter 3

Heterogeneous effects: Two-Stage Least Squares random forests with an application to Angrist and Evans (1998)

3.1 Introduction

Random forests (Breiman, 2001) are a successful and increasingly popular method for fitting flexible regression models based on statistical learning. The method consists in successively splitting a given sample into heterogeneous subgroups (yielding regression trees), and on repeating this procedure on random variations of the data (leading to random forests). Athey et al. (2019) have generalized the concept of random forests to a general class of estimation methods that solve conditional moment conditions. The applications presented in Athey et al. (2019) include the estimation of conditional average partial effects under exogeneity and conditional instrumental variable estimation based on the classic one-instrument formula (Wald's formula, e.g. Angrist and Pischke, 2008). Unfortunately, this formula does not easily extend to the case with multiple instruments, which is the case often encountered by practitioners.

This paper has two goals. The first one is to develop a conditional instrumental variable estimator based on multiple instruments in the general framework introduced by Athey et al. (2019), and to work out the expressions for estimation, sample splitting and variance estimation needed for implementation in software. This contributes to completing the toolbox of machine learning techniques for classical econometric problems and to verifying the validity of Athey et al.'s (2019) general framework. We also address the problem of tuning an instrumental variables forest which, to our best knowledge, has not been considered in the literature before. Finally, we provide an implementation in R and C$++$, extending previous codes contributed by Athey et al. (2019). Our second goal is to use this estimator to revisit a classic application of instrumental variables by Angrist and Evans (1998), who used sibling-sex composition instruments in order to investigate the effect of family size on parental labor supply. Including coarse group categories in their 2SLS regressions, they also provided a basic analysis of heterogeneity of these effects across characteristics such as mother's education or husband's earnings. We revisit this question using instrumental variables random forests, which allow one to plot detailed maps of heterogeneous effects across multiple dimensions which is not possible using standard regression techniques. This yields deeper insights into the nature of heterogeneity in these effects, going beyond the analysis in Angrist and Evans (1998).

The rest of the paper is structured as follows. Section 3.2 describes the extension of instrumental variables forests to the case with multiple instruments (two-stage least squares random forests). Section 3.3 presents our empirical application. Section 3.4 concludes.

3.2 Two-Stage Least Squares (2SLS) random forests

3.2.1 Generalized random forests

Athey et al. (2019) develop a general framework for building random forests for the estimation of local (i.e. conditional) effects $\theta(x)$ that are the solutions to moment conditions

$$\mathbb{E}\left[\psi_{\theta(x),\nu(x)}(O_i)|X_i = x\right] = 0 \text{ for all } x \in \mathcal{X}, \tag{3.1}$$

where $\nu(x)$ are nuisance parameters and $O_i, i = 1, \ldots n$ are i.i.d. sample data.

The generalized random forests estimates $\hat{\theta}(x), \hat{\nu}(x)$ are obtained as

$$\left(\hat{\theta}(x), \hat{\nu}(x)\right) \in \underset{\theta,\nu}{\arg\min} \left\| \sum_{i=1}^{n} \alpha_i(x)\psi_{\theta,\nu}(O_i) \right\|_2, \tag{3.2}$$

with

$$\alpha_{bi}(x) = \frac{1(\{X_i \in L_b(x)\})}{|L_b(x)|},\ \ \alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x), \tag{3.3}$$

where $L_b(x)$ is the set of observations falling into the same leaf as the test point x in tree b. The weights $\alpha_i(x)$ count how often observation X_i was in the same leaf as x across all fitted trees $b = 1, \ldots, B$. They thus determine the relevance of different observations i for fitting $\hat{\theta}(x)$ in the estimating equation (3.2) (i.e. a local weight).

As described in Athey et al. (2019), the tree-building algorithm proceeds by producing successive splits that maximize heterogeneity. Let $P \subset \mathcal{X}$ be a parent node which is to be split into two children $C_1, C_2 \subset \mathcal{X}$. Athey et al. (2019) show that this can be done by first generating pseudo-outcomes

$$\rho_i = -\xi' A_P^{-1} \psi_{\hat{\theta}_P, \hat{\nu}_P}(O_i), \tag{3.4}$$

where

$$A_p = \frac{1}{|\{i : X_i \in P\}|} \sum_{i:X_i \in P} \nabla \psi_{\hat{\theta}_P, \hat{\nu}_P}(O_i), \tag{3.5}$$

and ξ is a vector that picks out the θ coordinate from the (θ, ν) vector. The parameters $\hat{\theta}_P, \hat{\nu}_P$ are the estimators solving the empirical estimation equation in the parent node, i.e.

$$\left(\hat{\theta}_P, \hat{\nu}_P\right) \in \underset{\theta,\nu}{\arg\min} \left\| \sum_{i:X_i \in P} \psi_{\theta,\nu}(O_i) \right\|_2. \tag{3.6}$$

The splitting is then done on the pseudo-outcomes, i.e. P is split into C_1, C_2 by maximizing

$$\tilde{\Delta}(C_1, C_2) = \sum_{j=1}^{2} \frac{1}{|\{i : X_i \in C_j\}|} \left(\sum_{i:X_i \in C_j} \rho_i \right)^2. \tag{3.7}$$

In order to achieve consistency, Athey et al. (2019) use in addition a subsample splitting technique (‘honesty’), which divides subsamples of the data in order to grow trees on one part of the data and to solve the estimating equation (3.2) on another part of the data.

For statistical inference, Athey et al. (2019) show that the variance of $\hat{\theta}(x)$ can be consistently estimated as

$$\hat{\sigma}_n^2(x) = \xi' \hat{V}_n(x)^{-1} \hat{H}_n(x) \hat{V}_n(x)^{-1'} \xi \tag{3.8}$$

with

$$H_n(x) = var\left(\sum_{i=1}^{n} \alpha_i(x) \psi_{\theta,\nu}(O_i) \right) \tag{3.9}$$

and $V_n(x)$ a consistent estimator of

$$V(x) = \frac{\partial}{\partial(\theta, \nu)} \mathbb{E}\left(\psi_{\theta,\nu}(O_i) | X_i = x\right) \Big|_{\theta(x),\nu(x)}. \tag{3.10}$$

3.2.2 Application to instrumental variables estimation

We now describe the application of this framework to instrumental variables estimation. As one of their applications, Athey et al. (2019) consider the structural model

$$Y_i = \mu(X_i) + \tau(X_i)W_i + \epsilon_i, \tag{3.11}$$

where Y_i is the outcome, and W_i is a treatment variable that is potentially correlated with ϵ_i. In order to estimate $\tau(X_i)$, they consider the case of a scalar instrumental variable Z_i assumed to be independent of ϵ_i conditional on X_i. This yields the moment condition $\mathbb{E}(Z_i(Y_i - \mu(X_i) - \tau(X_i)W_i)|X_i = x) = 0$, which, along with the moment condition defining the intercept $\mu(X_i)$, defines the function $\psi(\cdot)_{\tau(x),\mu(x)}$ in (3.2) (Athey et al., 2019, p. 1172).

If there is more than one instrumental variable (i.e. if Z_i is a $M \times 1$ vector), one might be tempted to use the (vector) moment condition $\mathbb{E}(Z_i(Y_i - \mu(X_i) - \tau(X_i)W_i)|X_i = x) = 0$, but this is not possible because the framework described in Athey et al. (2019) is tailored to the just-identified case with as many moment conditions as estimated parameters (this is evident from A_P^{-1} in (3.4) and $\hat{V}_n(x)^{-1}$ in (3.8), showing that the function $\psi(\cdot)$ has as many arguments as it has dimensions). In order to arrive at a just-identified representation, we use a local variant of the two-stage least squares estimator (2SLS) to which we apply the technique of stacking moment conditions to form two-step estimation procedures (Wooldridge, 2010, p. 529).

This yields the representation

$$\psi_{\substack{\tau(x),\mu(x),\\ \gamma_1(x),\gamma_0(x)}}(Y_i, W_i, Z_i) = \begin{pmatrix} \widetilde{W}_i(Y_i - \mu(x) - \tau(x)\widetilde{W}_i) \\ Y_i - \mu(x) - \tau(x)\widetilde{W}_i \\ Z_i\left(W_i - \gamma_0(x) - Z_i'\gamma_1(x)\right) \\ W_i - \gamma_0(x) - Z_i'\gamma_1(x) \end{pmatrix}, \tag{3.12}$$

where $\widetilde{W}_i$ is the abbreviation of $\gamma_0(x) + Z_i'\gamma_1(x)$. Defining $W_i^c = (1\ W_i)'$, $Z_i^c = (1\ Z_i)'$, $\widehat{W}_i = \hat{\gamma}_0(x) + Z_i'\hat{\gamma}_1(x)$ and $\widehat{W}_i^c = (1\ \widehat{W}_i)'$, the $M + 3$ resulting moment conditions in (3.2) can be solved analytically yielding

$$
\begin{aligned}
\begin{pmatrix} \widehat{\gamma}_0(x) \\ \widehat{\gamma}_1(x) \end{pmatrix} &= \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c Z_i^{c'} \right)^{-1} \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c W_i \right) \\
\begin{pmatrix} \hat{\mu}(x) \\ \hat{\tau}(x) \end{pmatrix} &= \left(\sum_{i=1}^{n} \alpha_i(x) \widehat{W}_i^c \widehat{W}_i^{c'} \right)^{-1} \left(\sum_{i=1}^{n} \alpha_i(x) \widehat{W}_i^c Y_i \right) \\
&= \left(\left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c W_i^{c'} \right)' \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c Z_i^{c'} \right)^{-1} \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c W_i^{c'} \right) \right)^{-1} \\
&\quad \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c W_i^{c'} \right)' \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c Z_i^{c'} \right)^{-1} \left(\sum_{i=1}^{n} \alpha_i(x) Z_i^c Y_i \right).
\end{aligned}
\tag{3.13}
$$

Note that the score function $\psi(\tau, \mu, \gamma_1, \gamma_0)$ as defined in (3.12) is the negative gradient of a convex function $.5\, \psi(\tau, \mu, \gamma_1, \gamma_0)' \psi(\tau, \mu, \gamma_1, \gamma_0)$ which is required for consistency of the random forest (assumption 6 in Athey et al., 2019). The score function is very well-behaved and satisfies all other regularity assumptions listed in Athey et al. (2019).

The partial derivatives of the score function that are needed to compute the matrix A_P and the pseudo-outcomes ρ_i are given by

$$
\nabla\psi_{\substack{\tau(x),\mu(x),\\ \gamma_1(x),\gamma_0(x)}}(Y_i, W_i, Z_i) = \begin{pmatrix}
\widetilde{W}_{Pi}\widetilde{W}_{Pi} & \widetilde{W}_{Pi} & \multicolumn{4}{c}{-Z_i' \left(Y_i - \mu_P(x) - \tau_P(x)\widetilde{W}_{Pi} \right) + \tau_P(x) Z_i' \widetilde{W}_{Pi}} & -\left(Y_i - \mu_P(x) - \tau_P(x)\widetilde{W}_{Pi} \right) + \tau_P(x)\widetilde{W}_{Pi} \\
\widetilde{W}_{Pi} & 1 & \multicolumn{4}{c}{\tau_P(x) Z_i'} & \tau_P(x) \\
0 & 0 & Z_{i1}Z_{i1} & Z_{i1}Z_{i2} & ..... & Z_{i1}Z_{iM} & Z_{i1} \\
. & . & Z_{i2}Z_{i1} & Z_{i2}Z_{i2} & ..... & Z_{i2}Z_{iM} & Z_{i2} \\
. & . & ..... & ..... & ..... & ..... & . \\
. & . & Z_{iM}Z_{i1} & Z_{iM}Z_{i2} & ..... & Z_{iM}Z_{iM} & Z_{iM} \\
0 & 0 & Z_{i1} & Z_{i2} & ..... & Z_{iM} & 1
\end{pmatrix}
$$

where $\tau_P(x)$ and $\mu_P(x)$ denote estimates in the parent node and $\widetilde{W}_{Pi} = \gamma_{P0}(x) + Z_i' \gamma_{P1}(x)$ (for simplicity, we have changed the sign of $\psi(\cdot)$ before taking the derivative).

Defining $\gamma(x) = (\gamma_0(x), \gamma_1(x))'$, the matrix $V(x)$ needed for variance estimation is given by

$$
V(x) = \begin{pmatrix}
\mathbb{E}\left[\widetilde{W}_i\widetilde{W}_i|X=x\right] & \mathbb{E}\left[\widetilde{W}_i|X=x\right] & \mathbb{E}\left[-Z_i'\left(Y_i-\mu(x)-\tau(x)\widetilde{W}_i\right)+\tau(x)Z_i'\widetilde{W}_i|X=x\right] & & & \mathbb{E}\left[-\left(Y_i-\mu(x)-\tau(x)\widetilde{W}_i\right)+\tau(x)\widetilde{W}_i|X=x\right] \\
\mathbb{E}\left[\widetilde{W}_i|X=x\right] & 1 & \mathbb{E}\left[\tau(x)Z_i'|X=x\right] & & & \mathbb{E}\left[\tau(x)|X=x\right] \\
0 & 0 & \mathbb{E}\left[Z_{i1}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{i1}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{i1}|X=x\right] \\
. & . & \mathbb{E}\left[Z_{i2}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{i2}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{i2}|X=x\right] \\
. & . & ..... & ..... & ..... & . \\
. & . & \mathbb{E}\left[Z_{iM}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{iM}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{iM}|X=x\right] \\
0 & 0 & \mathbb{E}\left[Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{iM}|X=x\right] & 1
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\gamma(x)'\mathbb{E}[Z_i^c Z_i^{c'}|X]\gamma(x) & \gamma(x)'\mathbb{E}[Z_i^c|X] & -\mathbb{E}[Y_iZ_i'|X]+\mu(x)\mathbb{E}[Z_i'|X]+2\tau(x)\gamma(x)'\mathbb{E}[Z_i^cZ_i'|X] & & & -\mathbb{E}[Y_i|X]+\mu(x)+2\tau(x)\gamma(x)'\mathbb{E}[Z_i^c|X] \\
\gamma(x)'\mathbb{E}[Z_i^{c'}|X] & 1 & \tau(x)\mathbb{E}[Z_i'|X] & & & \tau(x) \\
0 & 0 & \mathbb{E}\left[Z_{i1}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{i1}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{i1}|X=x\right] \\
. & . & \mathbb{E}\left[Z_{i2}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{i2}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{i2}|X=x\right] \\
. & . & ..... & ..... & ..... & . \\
. & . & \mathbb{E}\left[Z_{iM}Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{iM}Z_{iM}|X=x\right] & \mathbb{E}\left[Z_{iM}|X=x\right] \\
0 & 0 & \mathbb{E}\left[Z_{i1}|X=x\right] & ..... & \mathbb{E}\left[Z_{iM}|X=x\right] & 1
\end{pmatrix}
$$

whose entries we estimate as in Athey et al. (2019) by the honest regression forests produced by the underlying estimation problem.

As described in Athey et al. (2019), the matrix $H_n(x)$ can be estimated by a bootstrap of little bag technique (Sexton and Laake, 2009), for which the overall number of trees $b = 1, \ldots, B$ is partitioned into $g = 1, \ldots G$ bags, where all the $l = B/G$ trees in one bag are estimated on the same half sample. Let tree b be the dth tree in bag g (i.e. tree gd), and denote $\Psi_b = \Psi_{gd} = \sum_{i=1}^n \alpha_{bi}\psi(O_i)$ (i.e. using only data from tree $b = gd$), an estimate $\hat{H}_n(x)$ is then obtained as the solution to

$$
\begin{aligned}
&\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{l}\sum_{d=1}^{l}\Psi_{gd}\right)\left(\frac{1}{l}\sum_{d=1}^{l}\Psi_{gd}\right)' \\
&= \hat{H}_n(x) + \frac{1}{l-1}\frac{1}{G}\sum_{g=1}^{G}\left[\frac{1}{l}\sum_{d=1}^{l}\left(\Psi_{gd}-\frac{1}{l}\sum_{d=1}^{l}\Psi_{gd}\right)\left(\Psi_{gd}-\frac{1}{l}\sum_{d=1}^{l}\Psi_{gd}\right)'\right].
\end{aligned}
\tag{3.14}
$$

We follow Athey et al. (2019) who recommend to carry out all of the above computations not for the original outcomes $\{Y_i, W_i, Z_i\}$, but for conditionally centered outcomes $\{Y_i^*, W_i^*, Z_i^*\}$. Let $y(x) = E(Y_i|X = x), w(x) = E(W_i|X = x)$ and $z(x) = E(Z_i|X = x)$, then $Y_i^* = Y_i - \hat{y}^{(-i)}(X_i)$, $W_i^* = W_i - \hat{w}^{(-i)}(X_i)$ and $Z_i^* = Z_i - \hat{z}^{(-i)}(X_i)$, where $\hat{y}^{(-i)}, \hat{w}^{(-i)}$ and $\hat{z}^{(-i)}$ are estimated using separate regression forests not using information based on observation i.

3.2.3 Tuning

Growing random forests requires the choice of basic tuning parameters such as the minimal node size, the subsample fraction, the number of variables used for splitting and parameters that control the imbalance of splits. The optimal choice of such tuning parameters is an open research topic. A common practical approach is to minimize a suitable loss function based on out-of-bag predictions. If W_i in (3.11) was exogenous, then a possible loss function would be the so-called R-learner

$$R = \frac{1}{n}\sum_{i=1}^{n}\Big(\big[Y_i - \hat{y}^{(-i)}(X_i)\big] - \hat{\tau}^{(-i)}(X_i)\,\big[W_i - \hat{w}^{(-i)}(X_i)\big] \Big)^2 \tag{3.15}$$

(Nie and Wager, 2021).

This would lead to a spurious fit, however, because W_i is endogenous. The endogeneity of W_i makes the identification of a suitable loss function more difficult than in the case of unconfoundedness. For parameter tuning, we therefore resort to an idea based on Chernozhukov and Hansen (2008) who argue that the reduced form of an instrumental variables problem is a representation that is always valid and informative about the relationship studied.[1] We, therefore, define our loss function as

$$L = \sum_{i=1}^{n}\Big(\big[Y_i - \hat{y}^{(-i)}(X_i)\big] - \hat{\rho}^{(-i)}(X_i)'\,\big[Z_i - \hat{z}^{(-i)}(X_i)\big] \Big)^2 \tag{3.16}$$

with $\hat{\rho}^{(-i)}(X_i)$ being the out-of-bag version of

$$\hat{\rho}(X_i) = \Big(\sum_{i=1}^{n}\alpha_i(x)(Z_i-\hat{z}^{(-i)}(X_i))(Z_i-\hat{z}^{(-i)}(X_i))'\Big)^{-1}\sum_{i=1}^{n}\alpha_i(x)(Z_i-\hat{z}^{(-i)}(X_i))(Y_i-\hat{y}^{(-i)}(X_i))'. \tag{3.17}$$

3.3 Revisiting Angrist and Evans (1998) using Two-Stage Least Squares random forests

We apply the above estimator to the estimation of the effect of the number of children on the labor supply of married women as described in more detail in Angrist and Evans (1998). Based on over 250.000 observations of married women aged between 21 and 35 years from

[1] We thank Stefan Wager for pointing this out to us.

the 1980 US Census, we follow as closely as possible the specifications in Angrist and Evans (1998), but use the alternative method of two-stage least squares random forests. The main difference between the 2SLS regression models used in Angrist and Evans (1998) and the random forests used here is that the latter allow us to estimate local effects, i.e. the effect of additional children on female labor supply for narrow subgroups of women defined by their observed characteristics.

The variables used to measure female labor supply ($= Y_i$) are either $Worked\ for\ pay$ (indicating whether the woman reported to work for pay in the given year) or $Weeks\ worked$ (representing the number of weeks worked in the given year). The treatment variable ($= W_i$) is whether the woman had more than two children, which is instrumented by the two instrumental variables $Two\ boys$ ($= Z_1$) and $Two\ girls$ ($= Z_2$) indicating whether the first two children were either boys or girls. As argued by Angrist and Evans (1998), these instruments are credibly random but influence the likelihood of having more than two children.[2] Following the construction of the instrument, the estimation sample is restricted to women who had at least two children. The covariates of the analysis ($= X_i$) are dummies for race ($= Black,\ Hispanic,\ Other\ race$), schooling of the woman in years ($= Mother's\ schooling$), her age ($= Age$), age at first birth ($= Age\ at\ first\ birth$), whether the first child was a boy ($= Boy\ 1st$) and father's income ($= Father's\ income$).

Before we present our empirical results, figures 3.1 and 3.2 show more details on our tuning procedure.

[2]Our setup requires that these instruments are valid *conditional on observed covariates* X_i. Given the nature of the instruments, this is apriori plausible. In addition, Farbmacher et al. (2022) have shown that the validity of these instruments cannot be rejected even conditional on observable characteristics.

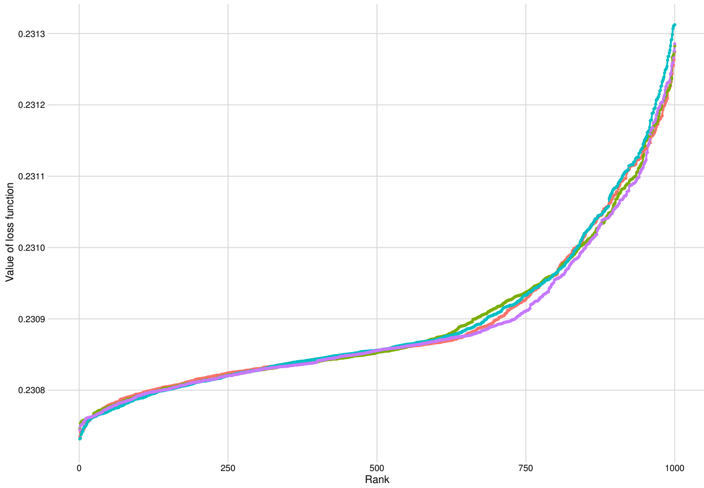


Figure 3.1 – Worked for pay: ordered values of loss function for different candidate tuning parameters

Tuning parameters optimized: minimal node size, subsample fraction, number of splitting variables, split balance parameter α, imbalance penalty. The different colours show the results for four separate Kriging runs (Roustant et al., 2012). Four each Kriging run, 200 random points from the space of tuning parameters are drawn. These are complemented by Kriging interpolations to generate 1000 points of the loss function surface.

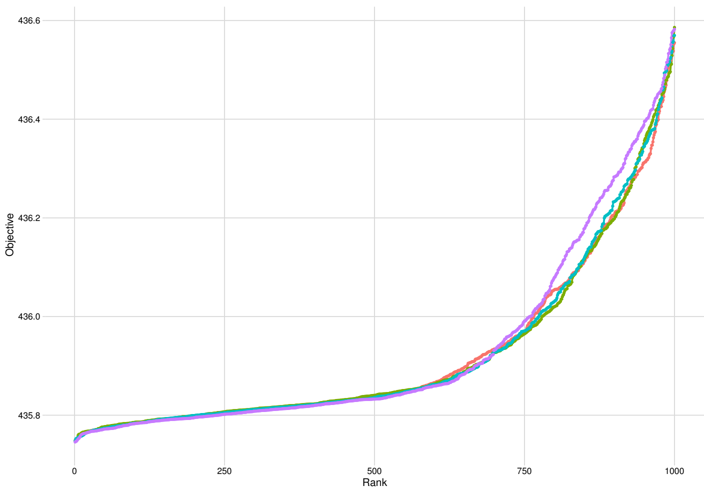


Figure 3.2 – Weeks worked: ordered values of loss function for different candidate parameters

Tuning parameters optimized: minimal node size, subsample fraction, number of splitting variables, split balance parameter α, imbalance penalty. The different colours show the results for four separate Kriging runs (Roustant et al., 2012). Four each Kriging run, 200 random points from the space of tuning parameters are drawn. These are complemented by Kriging interpolations to generate 1000 points of the loss function surface.

Following the implementation in Athey et al. (2019), we use a Kriging procedure (Roustant et al., 2012) to approximate the loss function surface and then choose the tuning parameters that correspond to the minimal value on the approximated loss function surface. For this purpose, we draw 200 random points from the space of tuning parameters and complement them by Kriging interpolations to generate 1000 points of the loss function surface. We carry out this procedure four times in order to minimize the risk of unrepresentative random draws. As in Athey et al. (2019)'s implementation, these computations use forests with a smaller number of trees than our final forests to save computation time. We optimize the following tuning parameters: minimal node size, subsample fraction, number of splitting variables, split balance parameter α, imbalance penalty.[3]

Figures 3.1 and 3.2 show that the minimal values and the shape of the loss functions are very similar across the four Kriging runs, making us confident that they are representative examples of the loss function surface. The final minimizing values for the tuning parameters are given in table 3.1. These were obtained as the smallest minimum out of the four Kriging runs. In general, our random forest results were quite robust to changes of the tuning parameters in a neighborhood of the loss function minimizing values, and only moderately sensitive to larger deviations from them.

	Worked for pay	*Weeks worked*
Minimal node size	1066	601
Subsample fraction	0.101144	.1299106
# Splitting variables	6	4
Balance parameter α	2.831253e-03	.0118385
Imbalance penalty	3.672790e-01	1.0620941
Minimal loss function	0.2307317	435.7457

Table 3.1 – Loss function minimizing tuning parameters

We now present our 2SLS random forest results. Figures 3.3 to 3.14 show the estimated treatment effects of having more than two children on whether the woman reported to work for pay and for the number of weeks worked per year along different covariate dimensions. Covariates not shown in a graph were set to median values. All random forests are based on 100,000 trees.

Figure 3.3 plots treatment effects of having more than two children on working for pay along the different values of the observed variables father's income and mother's education.

[3] For the definition of these parameters, see the software implementation of Athey et al. (2019). When splitting a parent node, the size of each child node must not be smaller than α times the size of the parent node. The imbalance parameter penalizes size differences between children of a parent node.

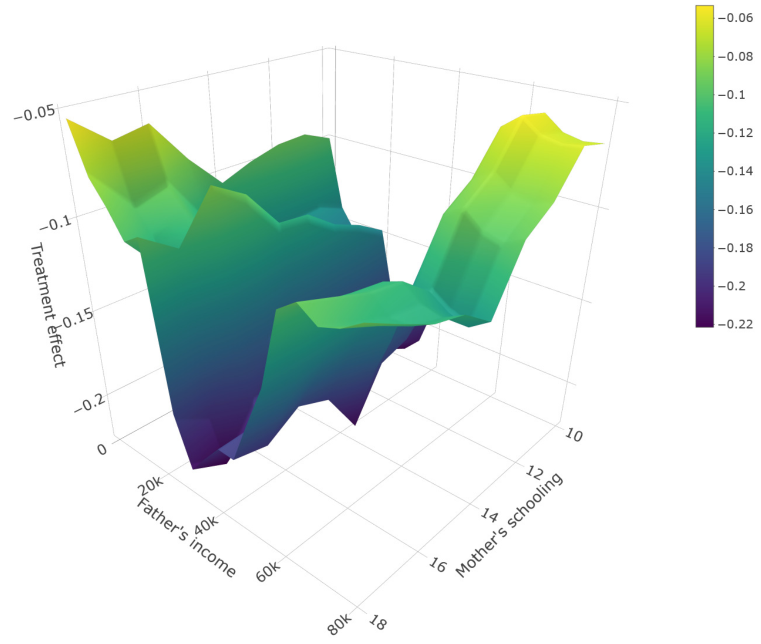


Figure 3.3 – Treatment effects of more than two children on worked for pay along the dimensions father's income and mother's education (2SLS random forest)

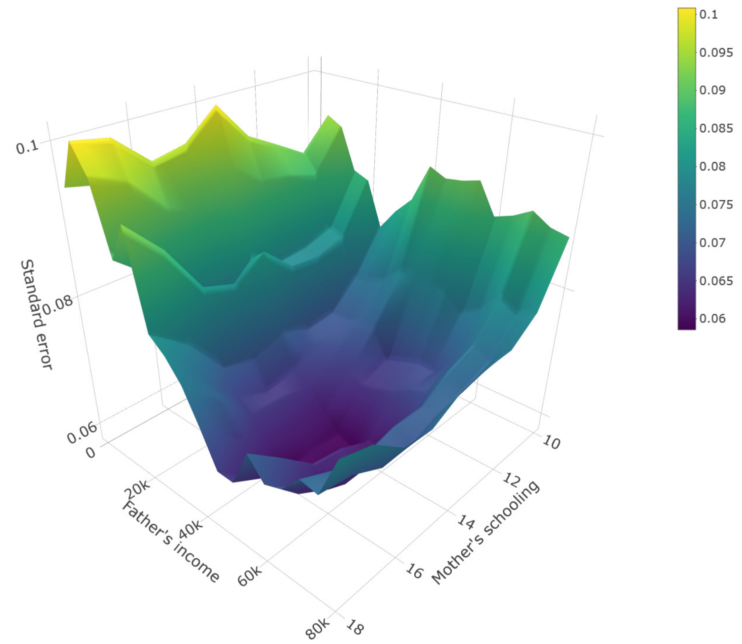


Figure 3.4 – Standard errors worked for pay along the dimensions father's income vs. mother's education (2SLS random forest)

The estimates along the dimension of father's income correspond quite well to the ones in Angrist and Evans for fathers income in the bottom third (-.122), the middle third (-.165) and the upper third (-.078)(Angrist and Evans, 1998, p. 468, table 9, panel A, column 5) if one considers mothers with a high school degree (12 years of education). The estimates for different values of mother's education in Angrist and Evans (1998) (table 9, panel B, less than high school: -.121, high school: -.147, more than high school: -.082) are also similar to the corresponding averaged values in figure 3.3, but the simple categorization into three groups misses the complex interaction effects uncovered by figure 3.3: for mother's with low husband's income, there is a strictly positive education gradient (higher education leads to a lower loss in labor supply due to children), while the effects become V-shaped for mother's with higher husband's income. If one looks more specifically at the effects of mother's education for mothers whose husband's income is in the middle third (the median is around 36,000 dollars), then the labor supply effects become even more negative (-.2 or lower). This is also the case in Angrist and Evans, although the estimates there tend to become rather imprecise when smaller subgroups are being considered (Angrist and Evans, 1998, table 9, panel C).

Figure 3.4 shows that most of the effects in figure 3.3 are reasonably precisely estimated, with most estimated standard errors ranging between .06 and .1.[4] The plot also nicely reflects the density of observations along the two dimensions (low standard errors in the center and rising standard errors towards areas with few observations).[5]

Figure 3.5 displays the corresponding estimates for weeks worked per year. Again, the estimates in Angrist and Evans for father's income (table 9, panel A, bottom third: -7.55 weeks, middle third: -7.11 weeks, top third: -3.17 weeks) are quite similar to the ones in the graph for mothers with a high school degree (12 years of education). In the direction of mother's education, the Angrist and Evans' estimates (less than high school: -7.12, high school: -6.42, more than high school: -2.93) correspond well to the estimates shown in the graph for a range of father's income between 40,000 and 60,000 dollars, but the random forest estimates suggest that the effects depend a lot on the exact value of father's income. For median father's income (around 36,000 dollars) and low values of mother's education, labor supply effects again get more negative (-10 weeks per year), both in the graph and in (Angrist and Evans, 1998, table 9, panel C, last column).

[4]These standard errors were computed according to equations (3.10) and (3.14) using a bag size of 200.

[5]The axes typically cover around 90 percent of the sample observations in each dimension (each axis approximately ranges from the 5th to the 95th percentile of the corresponding variable).

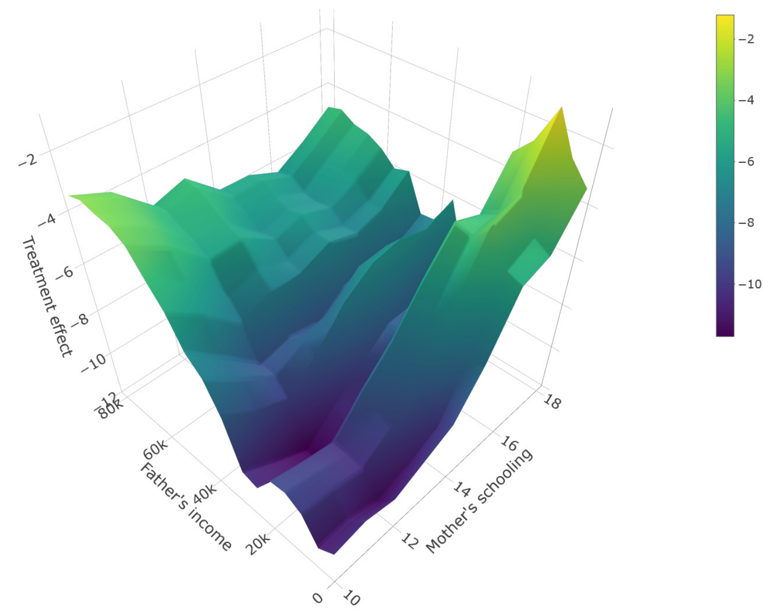


Figure 3.5 – Treatment effects of more than two children on weeks worked along the dimensions father's income and mother's education (2SLS random forest)

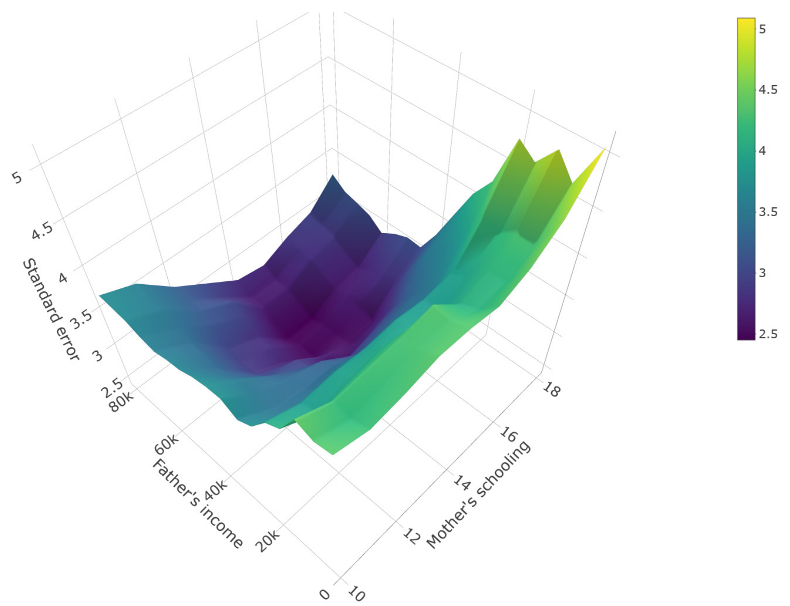


Figure 3.6 – Standard errors weeks worked along the dimensions father's income vs. mother's education (2SLS random forest)

The patterns revealed by figure 3.5 make much sense: for women with high husband's income, the loss in labor supply is small and not very sensitive to own education, while highly educated women in poorer households face high opportunity costs (in terms of foregone household income) if they do not participate in the labor market. The corresponding standard errors shown in figure 3.6 suggest that most effects are reasonably precisely estimated, although there are always areas in which this is not the case (areas with sparse density, e.g., few husbands have earnings close to zero).

Summing up our comparison of the instrumental variables random forests with the estimates in Angrist and Evans (1998) based on including basic group categories into 2SLS regressions (e.g. low/middle/high father's income, or low/middle/high education), we find that the general magnitude of the effects as well as basic qualitative patterns generally coincide well, but that the random forest shows in a much more detailed way, and simultaneously in more than one dimension, the exact geometry of effect heterogeneity.

A strength of the random forest methodology is that it models interaction effects in a fully unrestricted and automatic way. Figure 3.7 presents another example of such interaction effects, plotting the effects of having more than two children on the number of weeks worked per year across the dimensions father's income and mother's age. While the labor supply effect for low to medium values of father's income is U-shaped (young and old mothers do not reduce labor supply as much as middle aged mothers), it more and more transforms into a monotonically falling pattern for higher values of father's income. A possible explanation is that women in low income households face the necessity to return to the labor market once the children have reached a certain age, while those in high income households do not. Figure 3.9 shows the corresponding graph for the extensive margin (i.e. whether the woman worked for pay). The overall pattern is similar, but the interaction effect is less strong. Another example for heterogeneous effects with interactions is given in figure 3.11, showing the labor supply effects at the extensive margin along the dimensions of mother's age and mother's education. For older mothers, higher schooling is related to a lower loss in labor supply in the presence of children, while for younger mothers, the gradient is inversely U-shaped. Again, an explanation may be that mothers with high levels of education have a large incentive to return to work once their children have reached a certain age. Figure 3.13 suggests that this interaction effect does not apply to the number of weeks worked per year, where the effects of higher schooling is the same at all ages (i.e. more schooling leads to a smaller loss in labor supply due to children).

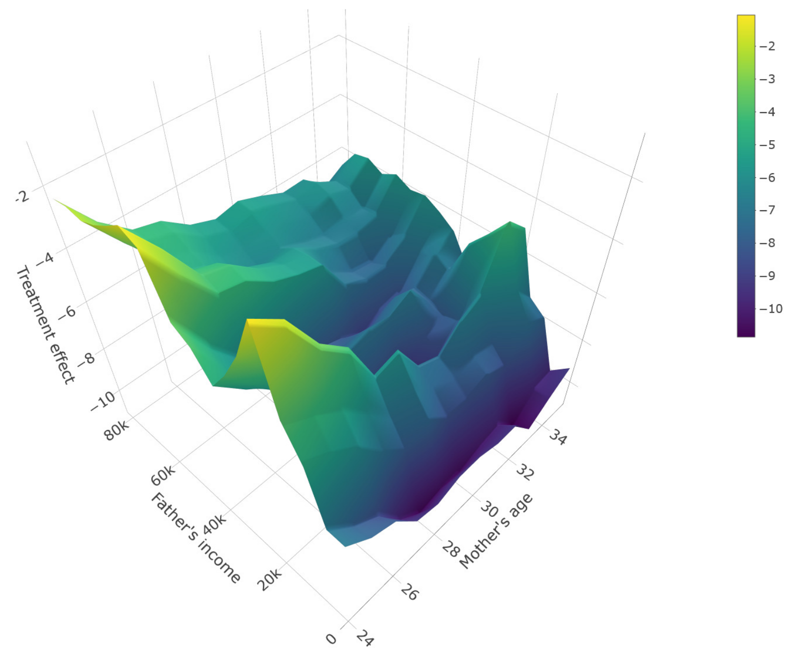


Figure 3.7 – Treatment effects of more than two children on weeks worked along the dimensions father's income and mother's age (2SLS random forest)

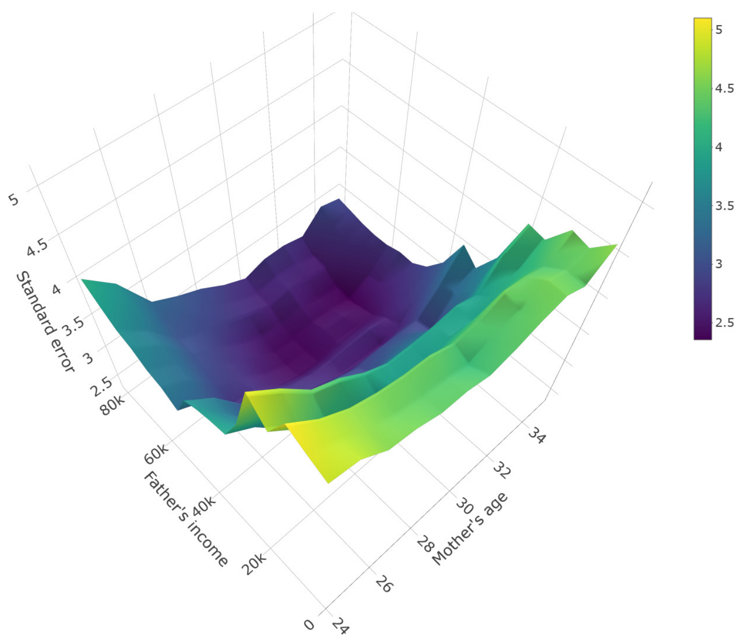


Figure 3.8 – Standard errors weeks worked along the dimensions father's income vs. mother's age (2SLS random forest)

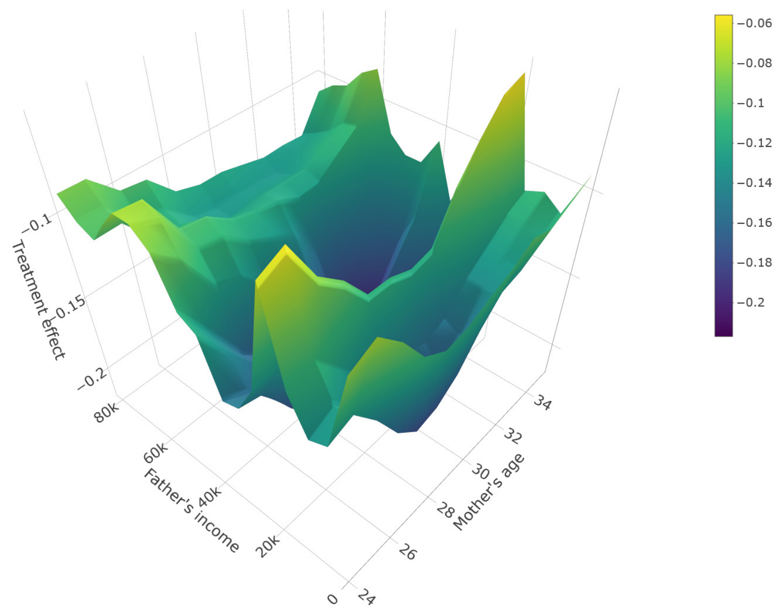


Figure 3.9 – Treatment effects of more than two children on worked for pay along the dimensions father's income and mother's age (2SLS random forest)

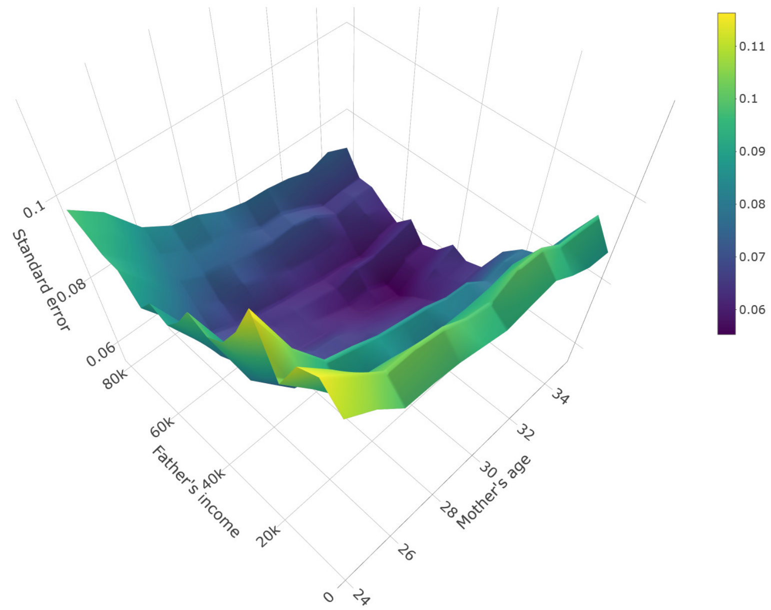


Figure 3.10 – Standard errors worked for pay along the dimensions father's income vs. mother's age (2SLS random forest)

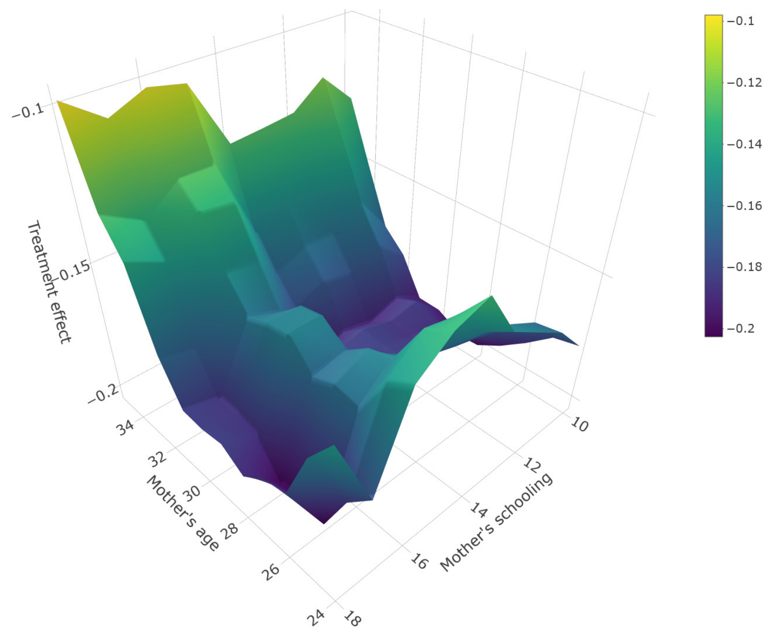


Figure 3.11 – Treatment effects of more than two children on worked for pay along the dimensions mother's education and mother's age (2SLS random forest)

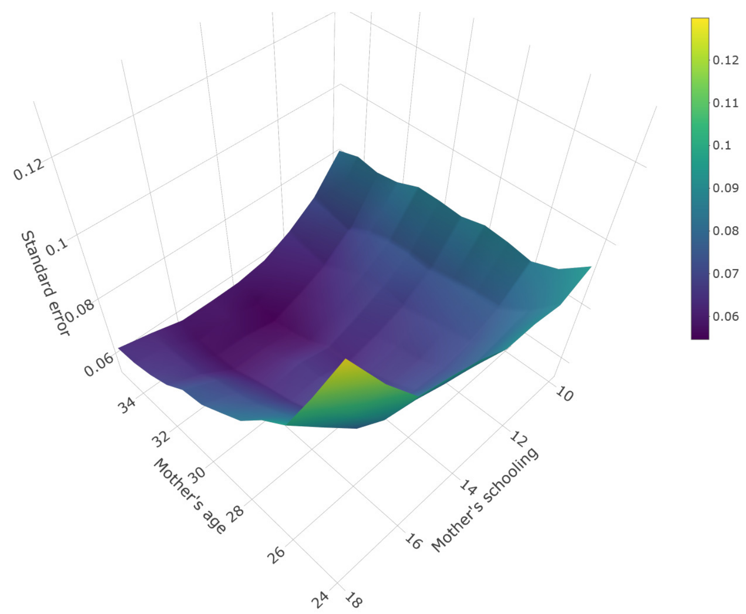


Figure 3.12 – Standard errors worked for pay along the dimensions mother's education vs. mother's age (2SLS random forest)

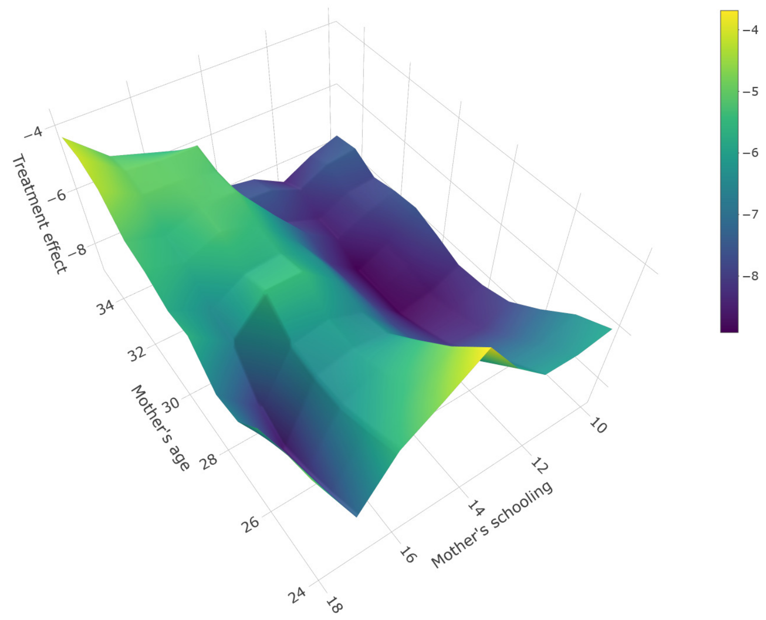


Figure 3.13 – Treatment effects of more than two children on weeks worked along the dimensions mother's education and mother's age (2SLS random forest)

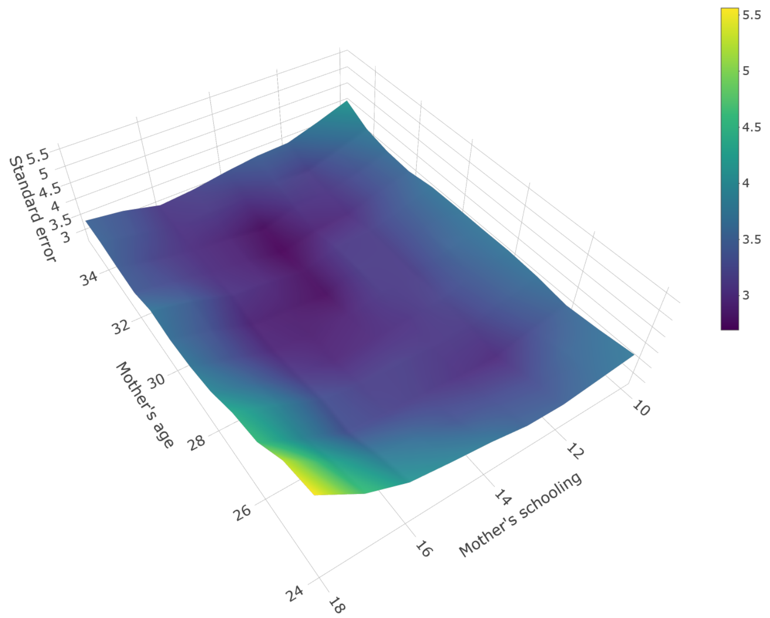


Figure 3.14 – Standard errors weeks worked along the dimensions mother's education income vs. mother's age (2SLS random forest)

We now present a summary of the effect heterogeneity as detected by our two-stage least squares random forests.[6] Figures 3.15 and 3.16 display the mean values of each covariate at different points of the treatment effect distribution. To fit all results on one scale, we standardize the covariates by dividing them by the difference of their maximal and minimal values (for dummy variables, this yields the fraction of cases at a particular point in the distribution).

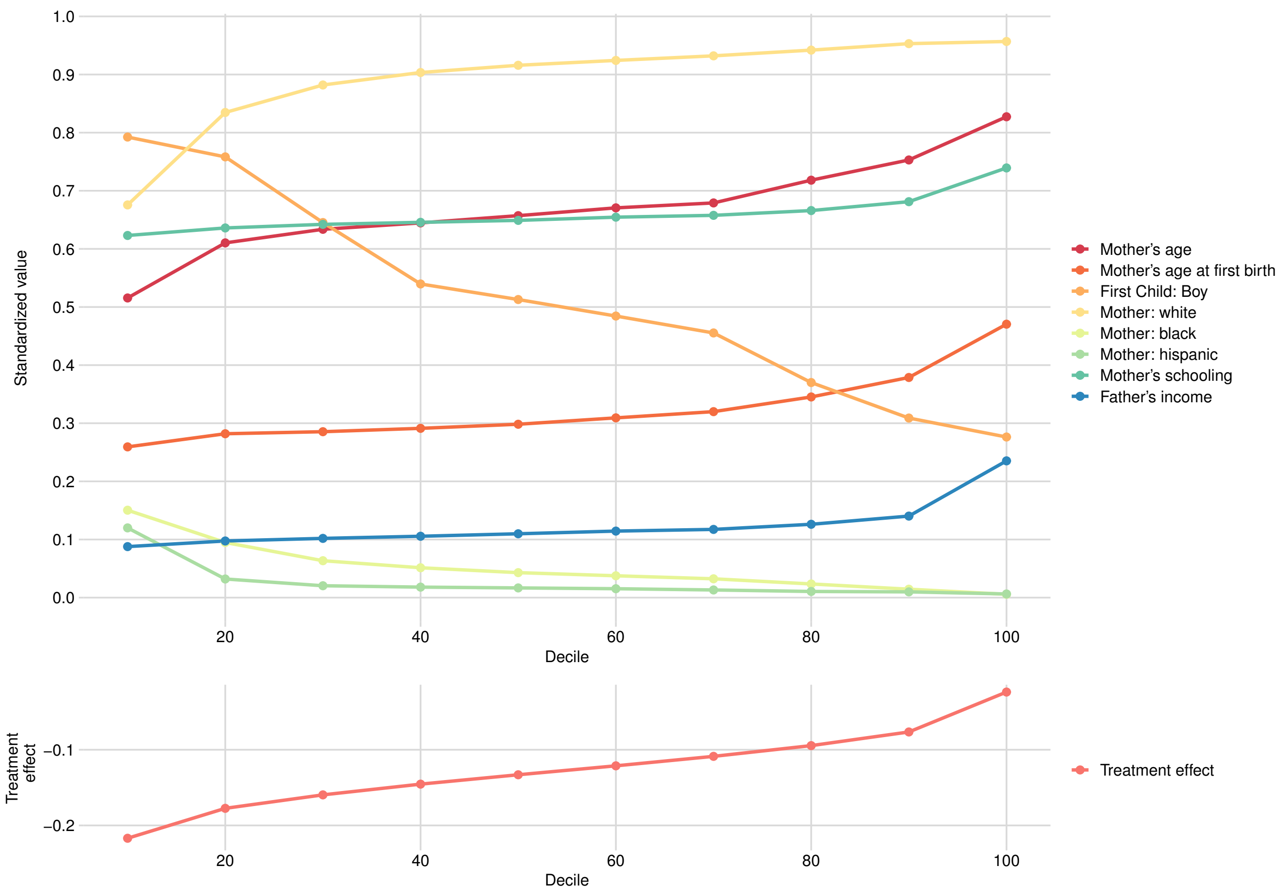


Figure 3.15 – Worked for pay: average values of covariates across different points of the treatment effect distribution

The y-axis shows the mean standardized values of the given covariate for a given decile of the distribution of treatment effects. The standardized values are obtained by dividing the value of each covariate by the difference between the maximal and the minimal value. In the case of dummy variables, this shows the fraction of cases.

The two graphs provide interesting insights into the structure of labor supply effects of children across different covariate dimensions. For example, it turns out that ethnic minorities and young mothers are very much overrepresented among large reductions in participation due to

[6] To our best knowledge, this way of summarizing effect heterogeneity was first proposed by Athey et al. (2020).

children (see lower deciles in figure 3.15). On the other hand, older women, women with more years of education, women who were older at the time of their first birth as well as women with high husband's earnings are much overrepresented among the cases with relatively mild reductions in labor supply due to children (see upper deciles in figure 3.15).

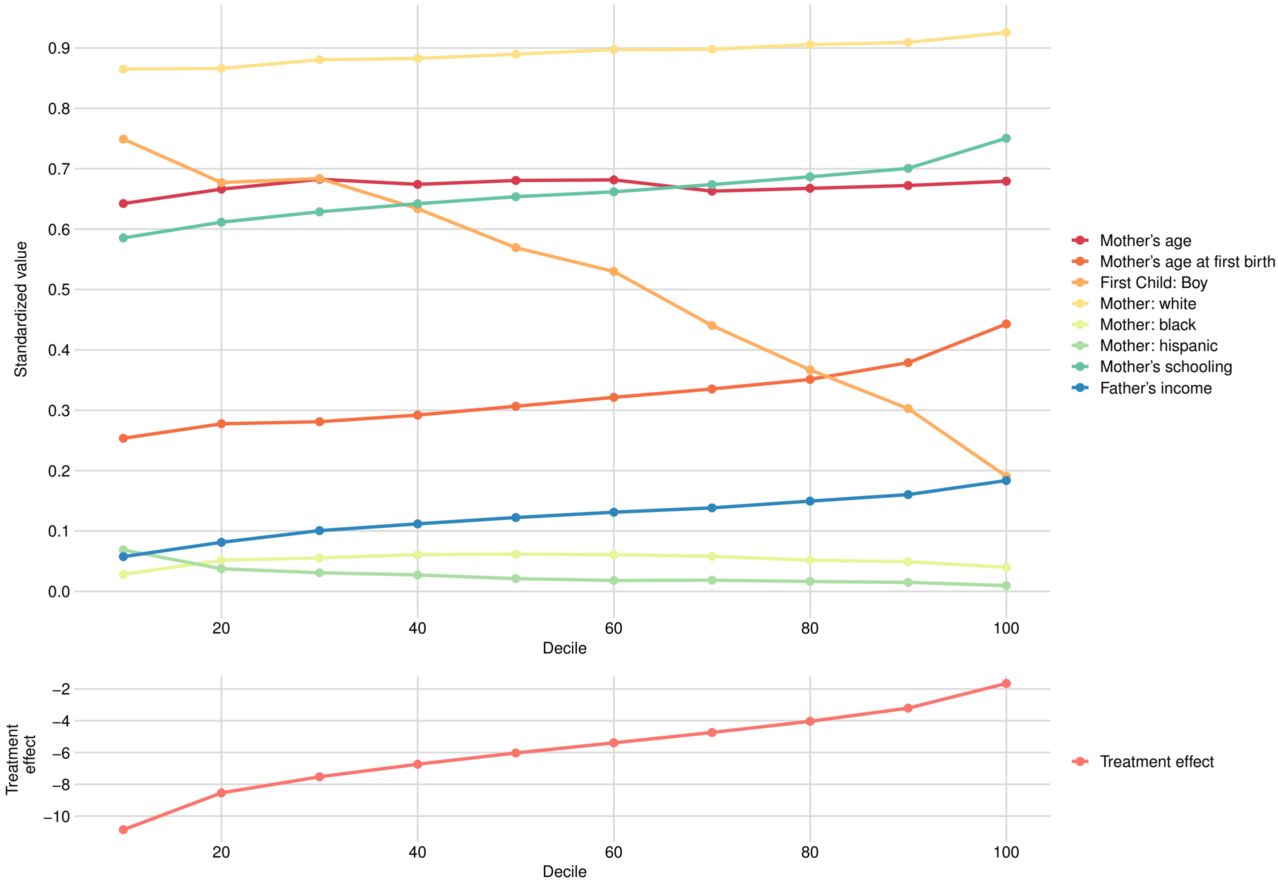


Figure 3.16 – Weeks worked: average values of covariates across different points of the treatment effect distribution

The y-axis shows the mean standardized values of the given covariate for a given decile of the distribution of treatment effects. The standardized values are obtained by dividing the value of each covariate by the difference between the maximal and the minimal value. In the case of dummy variables, this shows the fraction of cases.

The most striking pattern in figure 3.15 is that of the variable indicating whether the first child was a boy. It turns out that extremely negative labor supply reactions are very tightly related to having a boy as the first child, while relatively mild labor market reactions are tightly related to not having a boy as first child. At first glance, this appears to be consistent with the findings in Ichino et al. (2014) who find that women with first-born boys are less likely to work and work fewer hours in variety of countries. However, they argue that a likely channel for this is reduced marital stability after a first-born girl. This channel does not apply to our application (we only consider married women), suggesting an independent effect of having a first-born boy on later labor supply.[7]

Figure 3.16 shows the corresponding results for the number of weeks worked per year. The general patterns are very similar to those in figure 3.15, but the differences are more gradual across percentiles. This is not suprising given that the outcome modeled is a continuous variable (weeks worked). By contrast, the patterns are more nonlinear for the binary case of working vs. not working (mostly horizontal for the central deciles and strongly changing towards the boundaries, see figure 3.15). Also note that the median effects shown in the lower panels of figures 3.15 and 3.16 are close to the estimated two-stage least squares coefficients reported by Angrist and Evans (1998) for the whole sample (-.113 for worked for pay, and -5.15 for weeks worked, Angrist and Evans, 1998, Table 7).

3.4 Conclusion

This paper develops the case of two-stage least squares random forests (2SLS random forests) based on the general framework of Athey et al. (2019). Our application to Angrist and Evans (1998) demonstrates the usefulness of the method for evaluating effect heterogeneity along multiple dimensions, leading to a richer description of effect differences across observable characteristics and their interactions compared to conventional methods for average effects.

[7] See Ichino et al. (2014) for more discussion. A possibility is sample selection bias, i.e. women still married may differ in unobserved ways from those not married.

Chapter 4

Latent variable modeling: Parenting styles, socioeconomic status and (non-)cognitive skills

4.1 Introduction

The ambition to provide equal chances for economic and social participation to every child is broadly voiced among developed societies. Yet, a large body of literature documents a gap in non-cognitive and cognitive skills across parental income and education, even in early childhood (see Heckman and Mosso, 2014; Francesconi and Heckman, 2016; Attanasio, 2015, for extensive reviews of the recent literature). At the same time, early childhood factors are important determinants of economic and social adult outcomes. For example, Keane and Wolpin (1997); Cunha et al. (2005) and Huggett et al. (2011) show that at least half of the lifetime income variability across individuals arises from differences in childhood characteristics, which are primarily influenced by children's environment. In this sense, Cunha et al. (2006) underline the importance of parenting, i.e. all actions taken by parents to support the development of their child. Parenting has been the subject of economic research for several decades, dating back at least to the work on families by Becker (1981) and Becker and Tomes (1986).

Various studies that analyze the behavior of parents, focus on parental investment. In particular, they show how time and monetary investments affect children's skill acquisition and how such investments differ between parents with different socioeconomic status (e.g Cunha et al., 2013; Boneva and Rauh, 2016; Attanasio et al., 2020; Falk et al., 2021). Parenting styles, the broad strategy of how parents interact with their children, can be seen as another dimension of investment. However, the choice and effects of parenting styles is a rather novel topic in the economic literature (see Doepke et al., 2019, for an extensive review of the economic literature on parenting).

We supplement and extent the research on parenting styles. Our data contains a large set of measures about the parent-child interaction from six different domains, such as how parents monitor their child or how much autonomy parents leave to their child. Using latent dirichlet analysis for survey data (LDA-S), a hierarchical model recently proposed by Munro and Ng (2022), we operationalize parenting styles as latent classes. LDA-S differs from conventional models used to recover latent factors. First, LDA-S acknowledges that the survey responses on parent-child interaction are categorical. Second, it provides an economic interpretation to the unobserved heterogeneity. Finally, LDA-S connects unobserved heterogeneity with observed characteristics and survey responses. Therefore, we can directly incorporate and analyze differences in parenting styles along parental socioeconomic status and household composition. Further, we analyze how parenting styles are associated with children's non-cognitive and cognitive skills.

Our paper connects at least three strands of literature. First, we add to the discussion on how to operationalize parenting styles. Economists analyzing parenting styles commonly refer to the theoretical foundations laid out by Baumrind (1971, 1991). The framework classifies parents' behavior into two dimensions, demandingness and responsiveness. This results in four different parenting styles. *Authoritative* parents are both demanding and responsive. This style is defined by parents monitoring and communicating clear rules and standards for their children's behavior. Parents are assertive, but not invasive or restrictive. They support their children rather than punish them with disciplinary methods and raise them to be socially responsible, self-regulated and cooperative. *Authoritarian* parents are demanding as well, but not responsive. This parenting style is characterized by an orderly environment without explanations and a clear set of regulations. These parents are obedience- and status-oriented. In contrast, *permissive* parents are more responsive and less demanding. They allow self-regulation and avoid confrontation. This parenting style is non-traditional and tolerant. Last, *neglecting* parents are neither demanding nor responsive. This style is defined by non-supportive parents, who do not monitor their children, but actively reject them.

Doepke and Zilibotti (2017) develop a theory that rationalizes the choice between Baumrind's parenting styles. The equilibrium of the model results in different parenting styles depending on parental preferences and the socioeconomic environment. Other theoretical models capture key features of Baumrind's parenting styles. For example, Burton et al. (2002) model parenting styles as parent's patience when the child misbehaves. Lundberg et al. (2009) model parenting styles as the control of the parents on the child's decision-making. The model proposed by Cobb-Clark et al. (2019) captures the closeness of the parent-child relationship and the degree of monitoring parents employ. In empirical studies, the parenting styles are commonly modeled as continuous latent factors coming from factor analysis. Cobb-Clark et al. (2019) get two indices to measure parenting styles. One capturing whether parents respect the child's views and opinions, the other how much the parents monitor the child. Falk et al. (2021) rely on three domains of parent-child interaction (i) parental warmth, (ii) parental interest and monitoring, and (iii) parental psychological and behavioral control. They recover one latent factor, for which higher values reflect warm and child-oriented parenting but also a high degree of monitoring, while a lower value is associated with a higher degree of punishment. Fiorini and Keane (2014) identify two latent factors. One is an index of warmth and affection, the other can be interpreted as the effectiveness of imposing discipline.

The second major literature we connect to is that on the relation between parental investment and socioeconomic status. Recent literature established a strong link between both and discusses potential causal channels. Parental time and monetary investment may hinge on parents' objective, resource constraints and incorrect beliefs about the child's production function of human capital (Attanasio, 2015; Dizon-Ross, 2016; Doepke and Zilibotti, 2017). Evidence on the link between parenting styles and socioeconomic status is scarce. In their theoretical framework, Cobb-Clark et al. (2019) model parenting styles as parental investments. The investment depends not only on time and income but on mental effort required to pay attention to engage with, monitor and supervise the child. Their model links socioeconomic status to parental investment by allowing the endowment of a household's attention to depend on socioeconomic status. They empirically support the key features of their model and find that the extent to which parents monitor their children decreases with socioeconomic status. Falk et al. (2021) also link socioeconomic status and parenting styles. They find that parents with low socioeconomic status more often resort to parenting with a higher degree of punishment and less often to warm and child-oriented parenting. In this sense, Weinberg (2001) argues that, because of the scarcity of means, low-income parents have limited access to create incentives for the child. Therefore, they more often resort to authoritarian methods, such as corporal punishment. Doepke and Zilibotti (2017) find similar evidence. They show that parental education is associated with a lower probability to be a neglecting or authoritarian parent. In contrast, the probability to be an authoritative parent increases.[1]

Despite the notion of Becker (1960) that a larger number of children tends to lower investment in each individual child, the link between the household composition and the choice of parenting style is less considered. For parents with more than one child, it may not be possible to follow a warm and child-oriented style due to constraints.

The third and last strand of literature we contribute to is about the link between parenting styles and children's cognitive and non-cognitive skills. In psychology, many studies have attempted to establish this relationship. Often, such studies focus on achievement in school, the child's personality or non-cognitive skills (see for example Aunola et al., 2000; Aunola and Nurmi, 2005; Alegre, 2011; Masud et al., 2015). Most commonly, they find that an authoritative style, or the features that an authoritative style exhibits, are associated with the most favorable outcomes. The economic literature provides similar evidence. Doepke and Zilibotti (2017) show that authoritative parenting is associated with better performance in school and higher educational attainment. Cobb-Clark et al. (2019) find a positive association between respectful parenting and the child's educational outcomes as well as on non-cognitive skills in youth (internal locus of control and less risky behavior). Higher parental monitoring is associated with less risky behavior. Fiorini and Keane (2014) and Falk et al. (2021) study the association between parenting styles and early childhood skills. Falk et al. (2021) analyze the link between the parenting style and the child's patience, risk aversion, behavior, altruism and IQ. They find positive effects of a warmer and more child-oriented parenting style on all these outcomes. Fiorini and Keane (2014) show that non-cognitive skills like behavioral problems, social skills, and emotional problems are especially sensitive to the parenting style. They find that a parenting style combining effective discipline and parental warmth, i.e. an authoritative style in the sense of Baumrind, leads to the most favorable non-cognitive outcomes. In contrast to Falk et al. (2021), they find that cognitive skills are less sensitive to the parenting style.

We contribute to the literature in many ways. First, we apply a novel method which can handle a large set of measures on parent-child interactions. Therefore, we are able to describe

[1]Doepke and Zilibotti (2017) classify parents into Baumrind's (1991) four parenting styles using two questions asking children whether their parents are (i) supportive and (ii) strict/demanding.

parenting styles in more detail than previous studies. This allows us to separate styles that differ only in terms of a few, but important, dimensions. Second, the theoretical framework of LDA-S provides an economically interpretable link between parent-child interactions and parents' socioeconomic environment. Third, in contrast to continuous latent factors, latent classes more easily refer to theoretical models such as Baumrind (1971, 1991). In this way, the data driven approach can be embedded into theoretical frameworks. Fourth, we fill the gap on the link between parenting styles and household composition. Fifth, rich data on children's (non-)cognitive skills allow us to explore the association between parenting styles and children's skills. Finally, we are able to analyze the role of parenting styles in the emergence of skill gaps between children from different socio-economic environments in early childhood.

Applying LDA-S results in four parenting styles. Two styles closely resemble Baumrind's (1991) authoritative and authoritarian style. The two other styles can be interpreted as variations. One style is very similar to an *authoritative* style, which we call *democratic-loving*. The democratic-loving style differs from the authoritative style as such parents do not enforce their will, leave more autonomy to their child and communicate with the child more positively. The last style is like an *authoritarian* style, but the parents are much more inconsistent in their parenting. We call this style *authoritarian-inconsistent*. Our results show that parenting styles are strongly associated with household income, education and whether the child is an only child. Although our model does not directly allow for the identification of potential channels, the results suggest that constraints in both time and (non-)cognitive skills of the parents play an important role in choosing a parenting style. We find that children's (non-)cognitive skills are strongly associated with the parenting style. In line with Fiorini and Keane (2014), this link is more pronounced for non-cognitive than for cognitive skills. An *authoritative* and a *democratic-loving* style are associated with the highest skills, whereas children who are raised with an *authoritarian-inconsistent* style have the lowest skills. Our results show how differences in parenting styles contribute to the skill gap between children from different socioeconomic environments. We find that in particular styles associated with low household income are linked with lower skills. Parents with high household income are more likely to choose a style which is associated with higher skills. Further, parents with more than one child are more likely to choose a style that is related to lower skills. In contrast, having an only child is associated with a style that is associated with higher skills. Interestingly, parents' education is not systematically connected to parenting styles which are related to more favorable outcomes.

The remaining paper is structured as follows. Section 4.2 briefly describes the data. In section 4.3, we describe the method applied in the empirical analysis. Section 4.4 describes the parenting styles which are identified by our model. In section 4.5, we show how these parenting styles are linked with parental socioeconomic environment. Section 4.6 presents and discusses how the parenting styles are linked to (non-)cognitive skills. Section 4.7 concludes.

4.2 Data

This paper uses the first Starting Cohort (NEPS-SC1) of the German National Educational Panel Study (NEPS, 2021b). The panel study follows children born between February and July 2012 since they were six months old. One parent of every child is interviewed as part of the study. The data is perfectly suited to answer our research questions. It contains extensive information on each child, the child's development, the household in which the child lives in as well as on parents and how they rear their child. Our analysis mainly relies on questions about the parent-child interaction and measures on (non-)cognitive skills.

4.2.1 Parent-child interaction

To identify parenting styles, we rely on $J = 23$ questions about the parent-child interaction when the child was 5 and 6 years old. Broadly, the parent-child interaction can be classified into six categories: (1) How parents monitor their child, (2) how parents enforce their will, (3) how inconsistent parents are in their parenting, (4) how emotionally warm parents are with their child, (5) how parents communicate with their child and (6) how much autonomy parents leave to their child. Table 4.1 summarizes the questions and shows the response behavior of parents from 1530 children. For most of the questions, there is a rather large amount of response heterogeneity. The distribution is mostly concentrated around one option, i.e. multiple mass points at extremes of the response distribution do not exist. Our goal is to link parenting styles with household income, parental education and household composition. To this end, we compute the monthly household equivalence income (Hagenaars et al., 1994) and split it into three categories using the 33%- and 66%-quantile of its distribution. We use an indicator showing whether at least one parent has a university degree to measure parental education and an indicator that shows whether the child is an only child to measure the household composition. In the right panel of table 4.1 we depict p-values of a Chi-Square

Test that tests the null of independence between parent-child interaction x_{ij} and the parental characteristics. Results show that there are substantial differences in parenting between parents with different characteristics. In section 4.5, we analyze whether there are any patterns behind these differences.

	never	seldom	some-times	often	very often	χ^{inc}	χ^{educ}	χ^{sib}
Monitoring of Parents (M)								
If your child has new friends, you talk to him/her about them.	0.00	0.03	0.17	0.56	0.24	0.01	0.01	0.26
If your child went out, you ask him/her what he/she did and experienced.	0.00	0.00	0.01	0.28	0.71	0.05	0.79	0.75
If your child's out, you know exactly where he/she is.	0.00	0.00	0.01	0.11	0.88	0.18	0.46	0.92
If your child has new friends, you will meet them soon.	0.00	0.01	0.09	0.37	0.53	0.04	0.17	0.25
Enforcement of Will (E)								
When your child starts to negotiate with you, you exercise your authority.	0.01	0.11	0.52	0.33	0.03	0.59	0.00	0.11
You set clear limits for your child so that it does not exploit your goodwill.	0.00	0.04	0.24	0.59	0.12	0.01	0.00	0.68
If you want your child to do something, you give a clear command and don't tolerate any great detours.	0.01	0.11	0.35	0.47	0.06	0.39	0.00	0.48
If your child wants you to make an exception, you insist on your rules so that it is clear who is in charge in the family.	0.05	0.23	0.50	0.20	0.02	0.06	0.01	0.67
Inconsistency of parenting (I)								
You soften a punishment or terminate it prematurely.	0.04	0.33	0.49	0.12	0.02	0.05	0.34	0.11
On some days you are stricter with your child than on the others.	0.01	0.18	0.69	0.11	0.01	0.20	0.04	0.51
You threaten to punish your child, but you don't punish him/her.	0.18	0.48	0.28	0.06	0.00	0.00	0.02	0.02
It's hard for you to be resolute in your parenting.	0.09	0.47	0.37	0.06	0.01	0.03	0.00	0.05
Emotional Warmth (W)								
You show your child with words and gestures that you love him/her.	0.00	0.00	0.02	0.27	0.71	0.05	0.01	0.30
You comfort your child, when it is sad.	0.00	0.00	0.02	0.30	0.68	0.77	0.46	0.74
You praise your child.	0.00	0.00	0.03	0.56	0.41	0.39	0.00	0.14
Communication of parents (C)								
You criticize you child.	0.04	0.30	0.56	0.09	0.00	0.00	0.00	0.28
You shout at your child, if he/she has done something wrong.	0.15	0.55	0.27	0.03	0.00	0.17	0.19	0.04
You insult your child when you are angry at him/her.	0.46	0.34	0.18	0.02	0.00	0.10	0.06	0.04

Autonomy of child (A)	does not apply at all	does rather not apply	does rather apply	does completely apply	χ^{inc}	χ^{educ}	χ^{sib}
I think it's good if my child says what it thinks.	0.00	0.01	0.36	0.64	0.49	0.03	0.93
If my child wants something and doesn't get it, I'll explain why.	0.00	0.01	0.28	0.72	0.24	0.17	0.34
I often ask my child for opinion.	0.00	0.10	0.61	0.28	0.69	0.00	0.01
I let my child make its own plans for the things it wants to do.	0.01	0.17	0.66	0.15	0.88	0.00	0.34
If I want my child to do something, I'll explain why.	0.00	0.06	0.55	0.39	0.67	0.10	0.24

Table 4.1 – Measures on parent-child interaction

The table summarizes the measures on parent-child interaction. The panel in the middle depicts the survey responses. The right panel shows the p-values of a Chi-Square Test testing the null of independence between each parent-child interaction and monthly equivalence household income in three categories (χ^{inc}), whether at least parent has a university degree (χ^{educ}) or whether the child is an only child (χ^{sib}).

4.2.2 Non-cognitive and cognitive skills

The final part of this paper analyzes how parenting styles affect (non-)cognitive skills of the child. NEPS-SC1 collects an extensive set of different skill measures. In our main analysis, we focus on outcomes that were surveyed when the child was 4 and 7 years old. Cognitive skills are assessed via standardized tests (Berendes et al., 2013; NEPS, 2020, 2021a). The measurement of basic cognitive skills is based on tests that are as education-independent and domain-unspecific as possible. To measure mental performance, we rely on the child's ability to reason. Linguistic skills are undisputedly very import determinants for explaining social disparities in school careers. These are captured via listening comprehension. To test the mathematical literacy, the child is required to recognize and flexibly apply mathematics in realistic, mainly extra-mathematical situations.

To analyze the effect of parenting styles on non-cognitive skills, we use the Goodman's (1997) *Strengths and Difficulties Questionnaire* (SDQ) to measure social behavior (see Wohlkinger et al., 2019). We also observe the patience of the child. It is measured by a classical experiment on the delay of gratification, where the child could choose between one gift now or two gifts tomorrow (NEPS, 2021a).

In addition, we conduct supplementary analyses for outcomes that are only surveyed once. We analyze the effect of parenting styles on the child's personality traits measured by Big Five and how children cope with their every-day school life. This includes the child's autonomy, enjoyment of learning, willingness to make an effort, and social integration into the class.

4.3 Latent dirichlet analysis for survey data

Motivated by the differences shown in table 4.1, our goal is to explain the heterogeneity in parent-child interaction given the parental education, household income and whether the child is an only child. We apply an adapted version of latent dirichlet analysis (see Blei et al., 2003) for survey data (LDA-S) proposed by Munro and Ng (2022). LDA-S connects unobserved heterogeneity with observed characteristics and survey responses, explicitly acknowledges that survey responses are categorical and provides an economic interpretation of the unobserved heterogeneity. Throughout the paper, italic symbols denote scalars and bold symbols denote vectors that collect the respective scalars along their indices.

Assume we observe N parents indexed by i. Each parent belongs to one of $d_i \in \mathbb{G} = \{1, ..., G\}$ observable groups. In our case, individuals are grouped by all possible combinations of three categories of household equivalence income, an indicator that shows whether one parent has a university degree and whether the child has siblings living in the same household, i.e. $G = 12$. We observe J dimensions of the interaction between the target child of the survey and the parents x_{ij}. Each dimension j has L^j possible responses, where parents choose a single response v from $x_{ij} \in \mathbb{L}^j = \{1, ..., L^j\}$. We model the heterogeneous parent-child interaction as coming from K possible strategies to raise a child $z_i \in \mathbb{K} = \{1, ..., K\}$ (i.e. parenting styles). Parents choose z_i such that their utility is maximized. The model incorporates a group-affinity, which allows parents with similar income, education and number of children to choose the same parenting style. An individual effect allows parents to deviate from their group affinity, though.

$$z_i = \underset{k \in 1,\ldots,K}{\arg\max}\, U(k) = \underset{k \in 1,\ldots,K}{\arg\max} \sum_{j}^{K} \mathbb{1}\left(k = j\right)\left(u_{d_i,j} + e_{ij}\right), \tag{4.1}$$

where $u_{d_i,j}$ denotes the group affinity of $d_i = g$ for style $j = k$ and e_{ij} is an individual effect that captures everything else. The observed heterogeneity of an individual's group membership d_i and unobserved heterogeneity of an individual's chosen parenting style is linked by a random variable that gives the probability to choose parenting style $z_i = k$ given group membership $d_i = g$

$$\pi_{gk} = \mathbb{P}\left(z_i = k | d_i = g\right) = \mathbb{P}\left(u_{gk} + e_{ik} = \max_{j \in \mathbb{K}}\left(u_{gj} + e_{ij}\right)\right). \tag{4.2}$$

The chosen parenting style influences the observed parent-child interaction x_{ij}. Parents optimally interact with their child by maximizing their individual score function for each survey question.

$$x_{ij} = \underset{v \in \{1,\ldots,L^j\}}{\arg\max} \sum_{u=1}^{L_j} \mathbb{1}\left(v = u\right)\left(q^j_{z_i,u} + s^j_{iu}\right), \tag{4.3}$$

where $q^j_{z_i,u}$ denotes a parenting style-specific effect and s^j_{iu} an individual-specific effect, which allows parents to deviate from the usual parent-child interaction with parenting style $z_i = k$. The expected parent-child interaction is described by a random variable $\beta^j_{k,v}$ that captures the probability that an individual with parenting style $z_i = k$ chooses v as the response to survey question j.

$$\beta^j_{kv} = \mathbb{P}\left(x_{ij} = v | z_i = k\right) = \mathbb{P}\left(q^j_{z_i,v} + s^j_{iv} = \max_{u \in \mathbb{L}^j}\left(q^j_{z_i,u} + s^j_{iu}\right)\right) \tag{4.4}$$

Since we neither observe $\boldsymbol{u}_{g,:}, \boldsymbol{e}_{i,:}, \boldsymbol{q}^j_{k,:}$ and $\boldsymbol{s}^j_{i,:}$ nor their distribution, $\boldsymbol{\pi}_{g,:}$ and $\boldsymbol{\beta}^j_{k,:}$ are treated as random. Each $\boldsymbol{\pi}_{g,:}$ and each $\boldsymbol{\beta}^j_{k,:}$ is multinomial distributed. Therefore, Munro and Ng (2022) specify priors that follow a Dirichlet distribution with hyperparameters $\boldsymbol{\alpha}_{g,:} \in \mathbb{R}^K$ and

$\boldsymbol{\eta}^j_{k,:} \in \mathbb{R}^{Lj}$ respectively. In summary, LDA-S is identified by a hierarchical model

$$x_{ij}|\boldsymbol{\beta}, z_i \sim \text{Multinomial}\left(\boldsymbol{\beta}^j_{z_i,:}\right) \tag{4.5a}$$

$$z_i|\boldsymbol{\pi}_{d_i,:} \sim \text{Multinomial}\left(\boldsymbol{\pi}_{d_i,:}\right) \tag{4.5b}$$

$$\boldsymbol{\pi}_{d_i,:} \sim \text{Dirichlet}\left(\boldsymbol{\alpha}_{d_i,:}\right) \tag{4.5c}$$

$$\boldsymbol{\beta}^j_{z_i,:} \sim \text{Dirichlet}\left(\boldsymbol{\eta}^j_{z_i,:}\right). \tag{4.5d}$$

Using this, we can write down the joint distribution and estimate the model using MCMC methods (see Munro and Ng, 2022). The Gibbs Sampler iteratively samples each variable from its conditional distribution conditional on all other variables. Each iteration comprises three steps. First, z_i conditional on $\boldsymbol{x}_{i,:}$, $\boldsymbol{\beta}$, and $\boldsymbol{\pi}_{d_i,:}$ is sampled form a multinomial distribution. Second, $\boldsymbol{\beta}$ conditional on $\boldsymbol{\eta}$, $\boldsymbol{x}$, and $\boldsymbol{z}$ is sampled form a Dirichlet distribution. Third, $\boldsymbol{\pi}_{g,:}$ conditional on $\boldsymbol{\alpha}$, $\boldsymbol{x}$, and $\boldsymbol{z}$ is sampled form a Dirichlet distribution. Throughout the process, new values of the variables are used as soon as they are obtained. Draws of z_i depend on the values of $\boldsymbol{\beta}$ and $\boldsymbol{\pi}_{d_i,:}$ from the previous iteration, whereas draws of $\boldsymbol{\beta}$ and $\boldsymbol{\pi}_{g,:}$ depend on z_i from the current one. In our analysis, we conduct 20000 iterations. The results shown in sections 4.4, 4.5 and 4.6 are based on the the sample averages over the whole process, as it is usually done. To account for the bias caused by starting the system with randomly chosen initial values (initial transient), we burn the first 10000 iterations.

To estimate the model, hyperparameters of the Dirichlet distributions, $\boldsymbol{\alpha}_{g,:}$ and $\boldsymbol{\eta}^j_{k,:}$, and the number of parenting styles, K, have to be specified. Hyperparameters of the Dirichlet prior specify the researcher's beliefs about the importance of the group-specific terms ($\boldsymbol{u}_{g,:}$ and $\boldsymbol{q}^j_{k,:}$) relative to the individual-specific ones ($\boldsymbol{e}_{i,:}$ and $\boldsymbol{s}^j_{i,:}$). For example, we would specify $\alpha_{g,k} < 1$ if we believe that members of the same observable group g are likely to choose the same parenting style k. Similarly, $\eta^j_{kv} < 1$ reflects the belief that all individuals who choose the same parenting style, are likely to respond the same way. The opposite is true for $\alpha_{g,k} > 1$ and $\eta^j_{kv} > 1$. We impose uninformative priors to the relationship between observed group affinity and parenting style or parenting style and response behavior, i.e. $\alpha_{g,k} = 1\, \forall g, k$ and $\eta^j_{kv} = 1\, \forall j, k, v$. Regarding the number of parenting styles, we follow Munro and Ng (2022) and choose the optimal K according to the minimum of an approximated Bayesian information criterion (BIC). In our case $K = 4$.

In summary, LDA-S imposes structure on observable group indicators and parents' responses in the questionnaires by assuming that parents optimally choose parenting styles given their group membership and optimally select responses given their chosen parenting styles. The optimal choices are affected by individual terms and group commonalities in the first or parenting style commonalities in the second case. The individual effects allow parents to deviate from the choices usually made by other parents with the same group or parenting style.

4.4 Identification of parenting styles

In this section, we present the parenting styles defined by LDA-S. Our results show that Parenting Style 1 is chosen slightly more often (31%) than Parenting Style 2 (28%). Parenting Style 3 and 4 are chosen less often (20%). We want to give each parenting style a meaningful interpretation. Figure 4.1 and 4.2 depict the probability for an individual with parenting style $z_i = k$ to choose v as response to survey question j, i.e. $\boldsymbol{\beta}^j_{k,:}$.

The far left area of figure 4.1 shows how the typical **monitoring** behavior of parents, given their parenting style, looks like. Parents who choose Style 1 state (1) with a probability of 0.40 that they talk about the child's new friends *very often*, (2) with a probability of 0.93 that they ask about the child's experiences *very often*, (3) with a probability of 0.95 that they know where the child is *very often* and (4) with a probability of 0.69 that they meet their child's new friends *very often*. Typical parents with Style 4 behave similarly. In contrast, the respective probabilities of parents who choose Style 2 or Style 3 are much smaller. The left area of figure 4.1 shows that the styles also differ in the way how parents **enforce their will**. Parents with Style 1, Style 2, or Style 3 *often* enforce their will with a rather high probability. In contrast, the respective probabilities are much smaller for parents with Style 4. The parent-child interaction along **emotional warmth** shows that parents who choose Style 1 or Style 4 are likely to be very warm in their parenting. In contrast, parents with Style 2 or Style 3 are emotionally warm with a much smaller probability. With regard to **inconsistent** parenting, parents who choose Style 3 stand out. They state (1) with a probability of 0.34 that they *often* soften a punishment, (2) with a probability of 0.27 that they *often* are stricter on some days, (3) with a probability of 0.22 that they *often* inconsistently threaten their child and (4) with a probability of 0.21 that it is *often* hard for them to be resolute in their parenting. The respective probabilities for Style 1, Style 2, or Style 4 are very close to zero. The parent-child interaction along the dimension of **communication** shows that typical parents who choose Style 4 communicate with the child in a negative way with a very low probability. They state (1) with a probability of 0.45 that they *seldom* criticize the child, (2) with a probability of

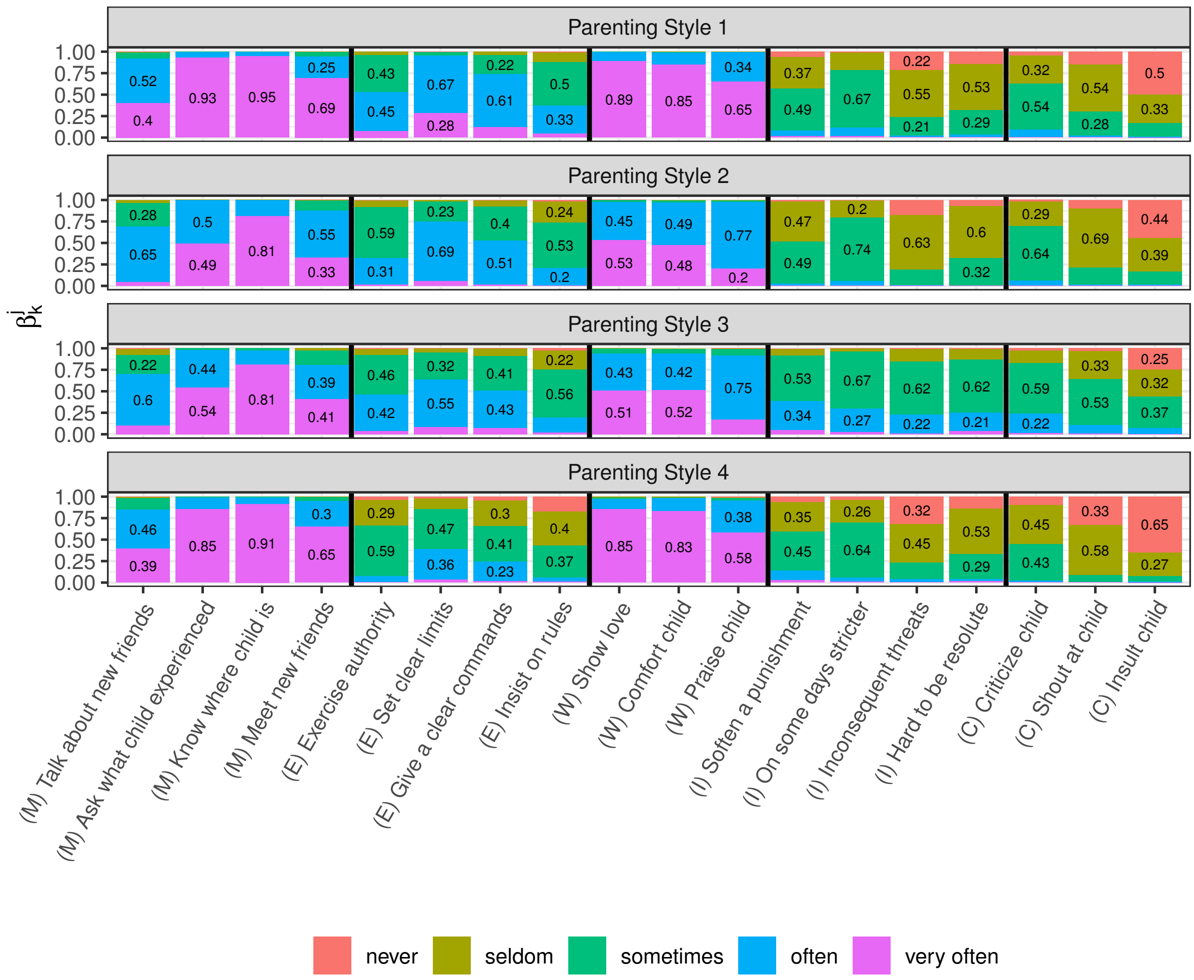


Figure 4.1 – Probability to respond given choice of parenting style I

The figure depicts $\beta^j_{k,:}$, i.e. the probability for an individual with parenting style $z_i = k$ to choose v as response to survey question j.

0.33 that they *never* shout at the child and (3) with a probability of 0.65 that they *never* insult the child. In contrast, parents with Style 3 state (1) with a probability of 0.22 that they *often* criticize the child, (2) with a probability of 0.53 that they *sometimes* shout at the child and (3) with a probability of 0.37 that they *sometimes* insult the child. The respective probabilities of Style 1 and Style 2 are somewhere between those of Style 3 and Style 4.

In figure 4.2, we show how much **autonomy** parents leave to their child. The behavior of parents with Style 4 stands out. They typically leave their child much autonomy. Children who are raised with Style 2 and 3 are likely to be less autonomous. The probabilities for parents with Style 1 lie in between. Their child is likely to be more autonomous than those of Style 2 or Style 3 but less than those of Style 4.

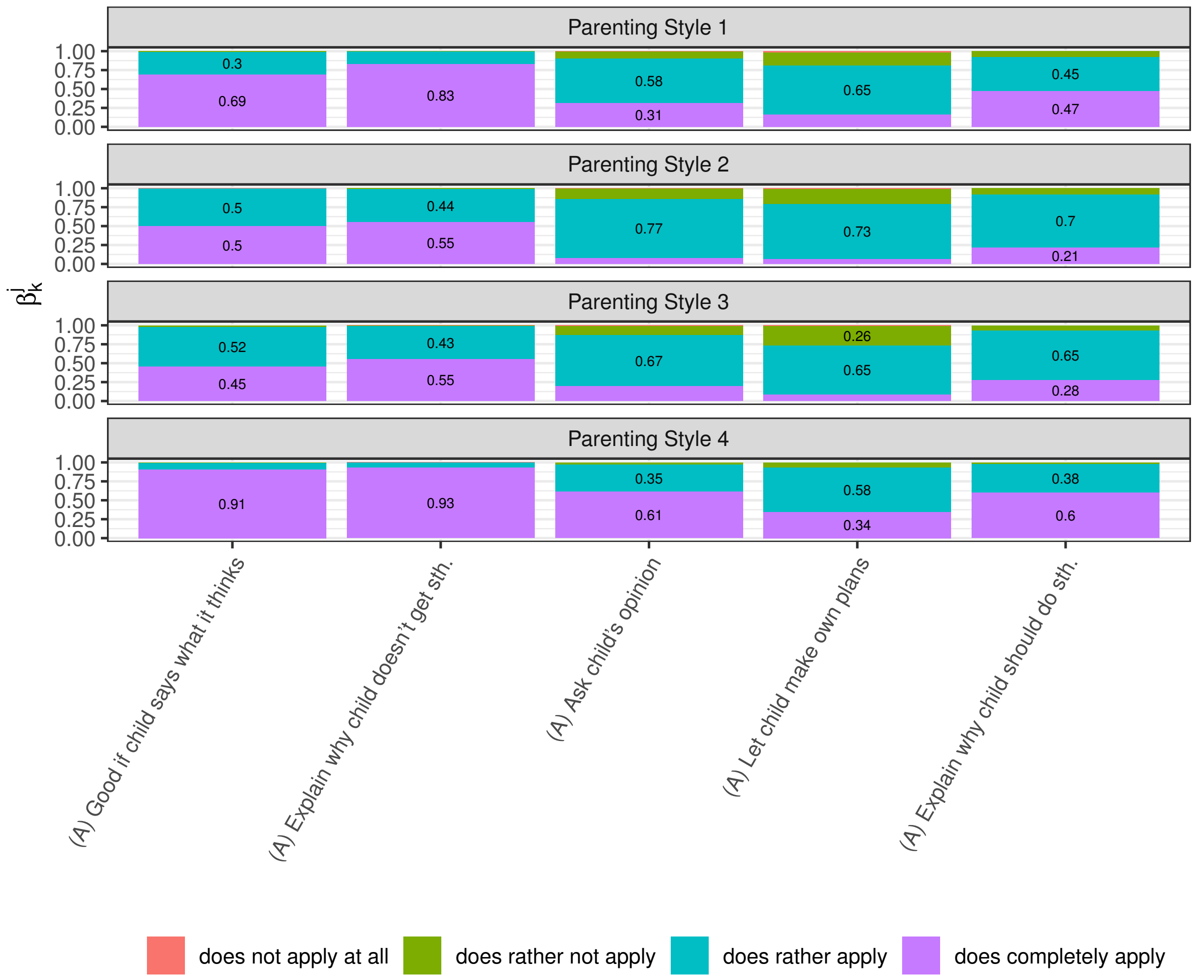


Figure 4.2 – Probability to respond given choice of parenting style II

The figure depicts $\beta^j_{k,:}$, i.e. the probability for an individual with parenting style $z_i = k$ to choose v as response to survey question j.

Figure 4.1 and 4.2 show major differences between the four parenting styles. To summarize these differences, we compute the Rao distance between $\beta^j_{k,:}$ and $\beta^j_{m,:}$ for all $k \neq m$ (Munro and Ng, 2022). Table 4.2 depicts the five dimensions of parent-child interaction where the parenting styles differ most from each other for each parenting style.

The results show that talking about new friends and asking what the child experienced are the two biggest differences between Style 1 and Style 2. In addition, four out of the five biggest differences between Style 1 and Style 4 can be followed back to the way parents enforce their will. We conclude that typical parents who choose Style 1 monitor their child, are consistent, powerfully enforce their will yet leave the child autonomy and are emotionally warm. This style closely mirrors Baumrind's (1991) *authoritative* style.

	Parenting Style 2	Parenting Style 3	Parenting Style 4
Parenting Style 1	(M) Talk about new friends (M) Ask what child experienced (W) Praise child (E) Set clear limits (W) Gestures	(I) Threaten child (I) Hard to be resolute (W) Praise child (I) Soften a punishment (E) Set clear limits	(E) Set clear limits (E) Exercise authority (E) Give a clear commands (E) Insist on rules (C) Shout at child
Parenting Style 2		(I) Threaten child (I) Hard to be resolute (I) Soften a punishment (C) Shout at child (I) On some days stricter	(A) Ask child's opinion (A) Good if child says what it thinks (M) Talk about new friends (A) Explain why child doesn't get sth. (W) Praise child
Parenting Style 3			(I) Threaten child (C) Shout at child (I) Hard to be resolute (A) Good if child says what it thinks (E) Exercise authority

Table 4.2 – Largest differences between parenting styles

The table summarizes the five biggest differences between each parenting style. Differences are computed using the Rao distance.

Besides the major difference between Style 1 and Style 4 in that parents with Style 4 do not powerfully enforce their will, one of the key difference between Style 4 and Style 1 in table 4.2 is shouting at the child. This is also one of the main differences between Style 4 and Style 3. Thinking that it's good if the child says what she thinks, explaining why the child doesn't get something and asking the child for her opinion are among the five biggest differences between Style 4 and Style 2. The latter also belongs to the biggest differences between Style 4 and Style 3. In summary, Style 4 is similar to Style 1 in many aspects. They differ, as parents with Style 4 typically do not enforce their will but leave their child more autonomy, and do not communicate negatively. This style is not only closely related to Baumrind's (1991) *authoritative* style, but also to Baumrind's (1991) *permissive* style. However, *permissive* parents do not extensively monitor their child. As positive and participative communication distinguish this style, we define Style 4 as *democratic-loving*.

Style 2 strongly differs from Style 4 in talking about new friends. Further, table 4.2 underlines that parents with Style 2 are emotionally much colder than parents with Style 1 or Style 4. Showing love with words or gestures belongs to the biggest differences between Style 1 and Style 2. Praising the child is one of the biggest differences between Style 4 and Style 2. We conclude that parents who choose Style 2 powerfully enforce their will, are not as emotionally warm as *authoritative* or *democratic-loving* parents and typically do not take the child's will into account as much as *authoritative* or *democratic-loving* parents. In line with Baumrind

(1991), we call such parents *authoritarian*.

Table 4.2 also shows that differences between Style 3 and the other styles are mainly due to inconsistent behavior. Other than that, Style 3 closely mirrors an *authoritarian* style. Therefore, we refer to this style as *authoritarian-inconsistent*.

4.5 Parenting styles and socio-economic environment

By modeling group affinities for parenting styles, our model directly incorporates differences in parenting styles along parents' education, household equivalence income and whether the child is an only child. In this section, we interpret $\boldsymbol{\pi}_{g,:}$, i.e. the probability to choose style k given membership of observable group g. Table 4.3 shows the average probabilities of $\boldsymbol{\pi}_{g,:}$ for each parental characteristic separately.

	Authoritative	Authoritarian	Authoritarian-inconsistent	Democratic-loving	Number of Observations
Education					
No University	0.34	0.26	0.22	0.18	643
University	0.27	0.31	0.19	0.23	887
Household equivalence income					
Low	0.26	0.28	0.25	0.21	431
Middle	0.29	0.30	0.21	0.20	504
High	0.34	0.28	0.17	0.21	595
Siblings					
Only child	0.32	0.23	0.21	0.25	323
Siblings	0.30	0.30	0.20	0.20	1207

Table 4.3 – Average probabilities

The table depicts the average probabilities to choose each style for each parental characteristic separately. Probabilities are computed by averaging $\boldsymbol{\pi}_{g,:}$ along observable groups weighted by the number of observations in each observable group.

The results show that parental education is an important determinant in choosing a parenting style. On average, parents with a university degree are more likely to raise their child with an *authoritarian* or *democratic-loving* style than parents without a university degree. In contrast, they are less likely to choose an *authoritative* or *authoritarian-inconsistent* style. The average probability to choose an *authoritative* style becomes higher with rising household income. In contrast, it is less likely to raise the child with an *authoritarian-inconsistent* style for parents with higher household income. The average probability to choose an *authoritarian* or *democratic-loving* style is not associated with household income. The household composition is strongly associated with the average probability to choose an *authoritarian* and *democratic-loving* style. Whereas the probability to raise the child with an *authoritarian* style is smaller for parents with an only child than for parents with more than one child. The probability to choose a *democratic-loving* style is higher for parents with an only child.

Figure 4.3 shows the link between parenting style and parents' socioeconomic status as well es as household composition in more detail. It depicts the probability to choose a parenting style given the membership of observable group g. An *authoritative* style is most likely chosen by

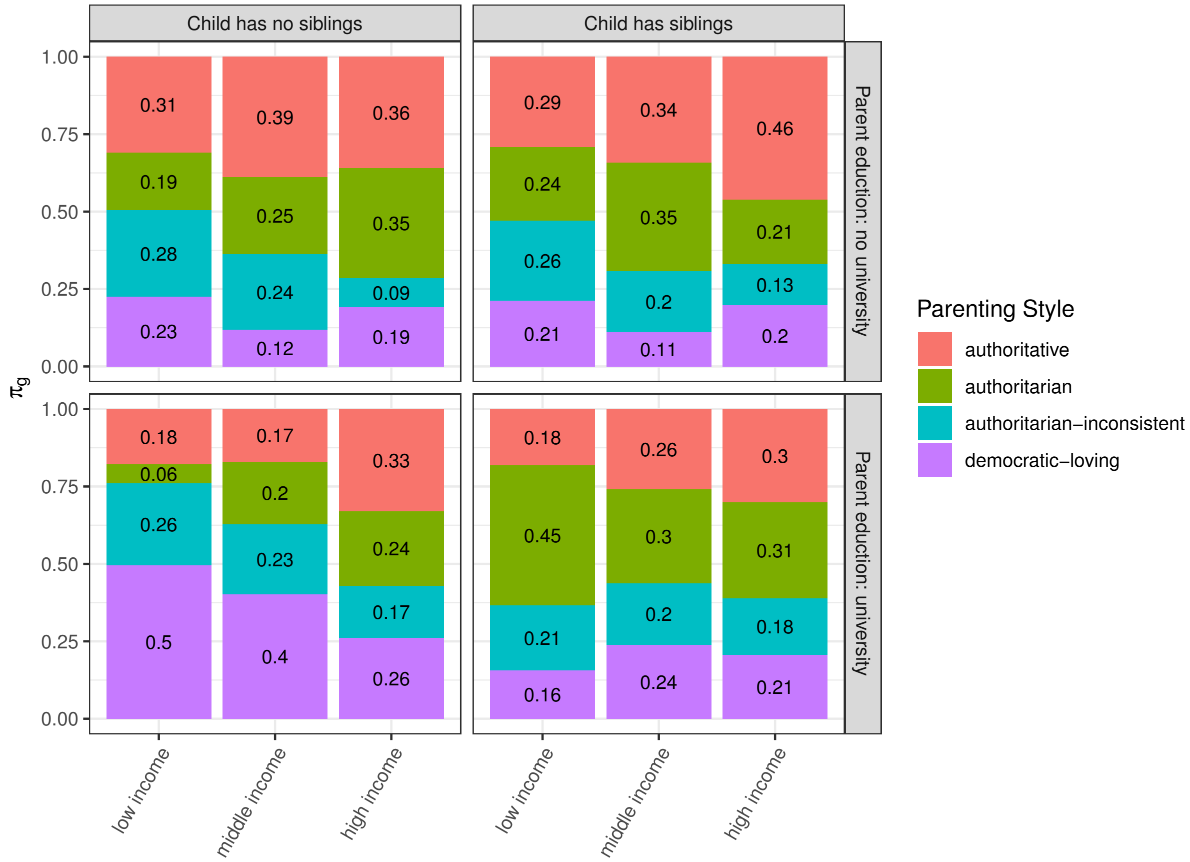


Figure 4.3 – Probability to choose parenting style k given membership of observable group

The figure shows $\boldsymbol{\pi}_{g,:}$, i.e. the probability to choose style k given membership of observable group g.

parents with more than one child, without a university degree and with high household income (0.46). The probability becomes smaller for lower income, however stays on a high level (0.29 and 0.34). For parents with lower education and an only child the probabilities to choose an *authoritative* style are also high (upper left panel). Those with low income have the lowest probability, although differences between income levels are less pronounced. The lower panels reveal that parents with low income and university degree are less likely to raise their child *authoritatively*. The probability becomes larger for higher income.

Parents with a university degree and more than one child are very likely to choose an *authoritarian* style (lower right panel). The probability is the highest if the income of the household is low (0.45). For parents with a middle and a high income the probability is much smaller, but stays on a high level (0.3). In comparison, parents with a university degree and an only child are less likely to choose an *authoritarian* style, especially if the income of the household is small (0.06) or in the middle (0.2). Opposed to highly educated parents with more than one child, the probability becomes higher with increasing income. For parents with lower education and an only child, the probability also systematically increases with higher income. The level of the probability is consistently higher compared to parents with a university degree and an only child, but smaller compared to parents without university education, more than one child and a low or middle household income. Thus, figure 4.3 shows that the number of children is strongly associated with the decision to choose an *authoritarian* style. Parents with an only child are less likely to choose such a style. The role of education and income in choosing an *authoritarian* style clearly differs with the number of children.

The probability to choose an *authoritarian-inconsistent* style does not vary by observable groups as much as the probability to choose one of the other styles. Mostly, the probability lies between 0.2 and 0.28. However, we find a clear pattern, where parents with lower income are systematically more inclined to choose this style.

Among all parents, those with one child and a university degree have the highest probability to choose a *democratic-loving* style (lower left panel). The probability is especially high for a low household income (0.5) and decreases for parents in households with middle or high income (0.4 and 0.26 respectively). Highly educated parents with more than one child or parents with low education are less inclined to raise their child with a *democratic-loving* style.

In summary, average probabilities depicted in table 4.3 show a strong association between parenting styles and parental characteristics. However, figure 4.3 indicates, that the link between parenting styles, parental socioeconomic status and household composition is complex. The probabilities strongly differ along observable groups, i.e. the combination of all three variables. The results show that there is no style which is clearly preferred by one group. This indicates that other characteristics not captured by the observable groups are very important determinants.

The results point to the importance of parents' time resources and (non-)cognitive skills when choosing a parenting style. Parents who choose a *democratic-loving* style strongly focus on the needs of the child and do not impose their will by directly restricting the child's actions. Such parents rather enforce their will by persuading the child. In contrast, an *authoritative* or *authoritarian* style directly restrict the child's actions. At the same time, holding a university degree is a good predictor for choosing a *democratic-loving* style. An *authoritative* style is generally more likely chosen by parents without a university degree. For parents with an only child, an *authoritarian* style is more likely chosen by parents with lower education. Therefore, we conclude that a *democratic-loving* style demands high (non-)cognitive skills of the parents. In contrast, an *authoritative* or an *authoritarian* style are not as (non-)cognitively demanding as a *democratic-loving* style. Therefore, (non-)cognitive skills may play an important role in the choice of a parenting style. Further, a *democratic-loving* style requires to discuss issues in case of disagreement between child and parent or to let the child make own plans. In contrast, an *authoritarian* style leaves less autonomy to the child. The number of children indicates whether parents are able to give their full attention to only one child or whether they have to allocate their time to multiple children. The results of the model show that parents with more than one child are less likely to choose a *democratic-loving* style, but more likely choose an *authoritarian* style. Therefore, we conclude that a *democratic-loving* style requires more time resources compared to an *authoritarian* style. This underlines the importance of time resources in choosing a parenting style.

4.6 Parenting styles and children's skill development

4.6.1 Estimation strategy

To analyze how parenting styles are related to the skill development of children, we estimate a simple static model, where we look at one specific period in childhood. Our data contains M skills S_i^{7m} when the child was 7 years old indexed by m. Our model uses predictions $z_{i1}, ..., z_{i4}$ for the parenting style of parents i derived from the estimation in section 4.4 and 4.5 (z_{ik} equals one for the most likely parenting style for parents i and zero otherwise). We assume that S_i^{7m} is linearly affected by the parenting style z_{ik}, the initial endowment of that skill S_i^{4m}

at age 4, and a vector of child's and parent's individual characteristics captured by $\boldsymbol{X}_i^m$.

$$S_i^{7m} = \beta_{0g}^m + \sum_{k \in 2,3,4} \gamma_{1k}^m z_{ik} + \delta_1^m S_i^{4m} + \boldsymbol{\delta}_2^{m'} \boldsymbol{X}_i^m + \eta_i^m, \tag{4.6}$$

where the base category is z_{i1}, i.e. whether i chooses an *authoritative* style or not, and β_{0g}^m is an intercept that varies with the observed group membership. The coefficient γ_{1k}^m can be interpreted as the effect of the parenting style on the change in skills between age 4 and 7. This effect is biased if, given $\boldsymbol{X}_i^m$, unobserved factors affect both the choice of the parenting style and the change in skills of the child. For example, the speed at which children learn may be related to factors that are also associated with the parents' choice of parenting style (genetics, neighborhood, etc.). Unfortunately, we cannot use an instrumental variable approach to solve this issue, since we could not find any exogenous source of variation that affects the choice of the parenting style. Therefore, we use an extensive set of control variables to mitigate the bias induced by potential confounding factors. First, other skills at age 4 may affect both, the change in skills of the child and the choice of parenting style. Hence, $\boldsymbol{X}_i^m$ contains the initial endowment of all other considered skills and additional measures on (non-)cognitive skills at age 4 (e.g. measures on the child's temperament). Second, the choice of the parenting style may depend on parental skills and preferences, which may also have a strong effect on the change in the child's skills. To account for this, we control for the respondents personality traits, patience and risk aversion. Third, peers may be important confounding factors. Therefore, we control for the share of parents' friends who hold a university degree, the share of parents' friends with migration background and the share of the child's friends with migration background. Fourth, the quality time parents spend with their child may be correlated with both, the change in skills of the child and the choice of the parenting style. To address this source of bias, we control how much quality time parents spent with their child when the child was 6 years old, i.e. in our considered period of childhood between age 4 and 7. The indicator is constructed by taking the average over how often parents (1) read a story to their child, (2) show single letters or the alphabet to the child, (3) practice numbers with the child, (4) teach short poems, rhymes or songs to the child, (5) paint, draw or craft with the child, (6) go to the library with the child, and (7) tell a story to the child. Finally, we control for demographic characteristics of the child and the parents. All controls are summarized in table A2 in the appendix.

Using equation (4.6), we estimate how parenting styles are related to cognitive skills (mathematical literacy, listening comprehension, reasoning) and non-cognitive skills (problem behavior, prosocial behavior and patience). We also report supplementary results on how parenting styles are associated with the child's personality traits measured by the Big Five. However, the child's initial endowments of personality traits at age 4 are not available in the data due to the children's young age. Since parents pass on their skills and preferences to their child through genetic, social or other channels, we use the personality traits of the interviewed parent to measure the child's initial endowment (comparable to Falk et al., 2021). Further, we analyze how the children cope with their school day. No initial endowments can be observed since children were recently enrolled.

4.6.2 Results

We estimate equation (4.6) using ordinary least squares. All regression models contain the same control variables. The outcomes are normalized to have mean 0 and standard deviation 1. Table 4.4 shows the main results of our analysis.

	Authoritative vs. ...			Authoritarian vs. ...		Authoritarian-inconsistent vs. ...	
	Authoritarian	Authoritarian-inconsistent	Democratic-loving	Authoritarian-inconsistent	Democratic-loving	Democratic-loving	N
Cognitive skills							
Reasoning	-0.01	0.07	0.00	0.08	0.02	-0.07	1373
Mathematical literacy	0.05	0.09	-0.08	0.03	-0.13**	-0.17**	1364
Listening comprehension	-0.02	0.12	-0.14*	0.14*	-0.12	-0.26***	1228
Non-cognitive skills							
Prosocial	0.10	0.30***	-0.08	0.19**	-0.18**	-0.38***	1182
Problem behaviour	-0.13*	-0.17**	-0.01	-0.04	0.12	0.16*	1184
Patience	0.13*	0.07	0.23***	-0.07	0.10	0.16*	1227

Table 4.4 – Parenting styles and skills - main results

The table depicts the effect of each parenting style on child's (non-)cognitive skills. All controls are summarized in table A2. Controls include the initial endowment of the skill observed at age 4. Outcomes are normalized to have mean 0 and standard deviation 1. Significance of the coefficients at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, *, respectively. The last column N shows the number of observations.

The upper panel shows the effect of the parenting style on cognitive skills. The ability to reason is not significantly associated with the parenting style. A *democratic-loving* style is associated with a significant higher mathematical literacy compared to an *authoritarian* or *authoritarian-inconsistent* style. Further, the listening comprehension of children raised with a *democratic-loving* style is significantly higher than the listening comprehension of children with *authoritarian-inconsistent* or *authoritative* parents. Children who are raised with an *authoritarian* style have a higher listening comprehension compared to children raised with an *authoritarian-inconsistent* style.

In the lower panel, we report how parenting styles are related to non-cognitive skills. Children raised by *authoritarian-inconsistent* parents are less prosocial than children with *authoritative*, *democratic-loving* or *authoritarian* parents. Further, they exhibit problem behavior more frequently than children with parents who choose an *authoritative* or a *democratic-loving* style. An *authoritarian* style is associated with less prosocial behavior than a *democratic-loving* style and with a more frequent problem behavior than an *authoritative* style. Regarding social behavior (i.e. prosocial and problem behavior) *authoritative* and *democratic-loving* styles do not differ. However, we find that children who are raised by *authoritative* parents are more patient than those raised by *democratic-loving* parents. They are also significantly more patient than children of *authoritarian* parents. Children raised with *authoritarian-inconsistent* style are significantly more patient than children with *democratic-loving* parents.

Table 4.5 summarizes supplementary results on how parenting styles are related to the child's non-cognitive skills.

	Authoritative vs. ...			Authoritarian vs. ...		Authoritarian-inconsistent vs. ...	
	Authoritarian	Authoritarian-inconsistent	Democratic-loving	Authoritarian-inconsistent	Democratic-loving	Democratic-loving	N
Personality							
Extraversion	0.21***	0.18**	0.20**	-0.02	-0.01	0.01	1203
Conscientiousness	0.05	0.28***	-0.02	0.23***	-0.07	-0.30***	1202
Agreeableness	0.05	0.28***	-0.07	0.23***	-0.12	-0.35***	1191
Openness	0.19***	0.18**	0.02	-0.01	-0.17**	-0.16*	1202
Neuroticism	-0.09	-0.18**	0.02	-0.09	0.12	0.20**	1202
Coping with school day: Child's autonomy							
Doing homework independently	0.03	0.25***	0.00	0.22**	-0.03	-0.24**	1084
Needs support with homework	0.03	-0.28***	-0.01	-0.32***	-0.04	0.28***	1123
Can cope with many tasks easily	0.13*	0.20**	0.04	0.07	-0.09	-0.16*	1261
Coping with school day: Enjoyment of learning							
Likes to go to school	0.02	0.05	-0.08	0.03	-0.10	-0.13	1275
Having fun at school	0.06	0.20**	-0.04	0.14*	-0.10	-0.24**	1274
Having fun studiying	0.12*	0.19**	0.04	0.06	-0.09	-0.15*	1272
Coping with school day: Willingness to make an effort							
Treats working materials careful	0.11	0.29***	-0.04	0.19**	-0.15*	-0.33***	1273
Completes tasks with care	0.04	0.24***	-0.04	0.20**	-0.08	-0.29***	1270
Gives up fast	-0.01	-0.13	-0.00	-0.12	0.01	0.13	1271
Tries hard if task are difficult	0.13*	0.33***	0.04	0.19**	-0.09	-0.28***	1263
Coping with school day: Social integration							
Integrated well in class	0.10	0.18**	-0.07	0.08	-0.17**	-0.25***	1275
Has many friends in class	0.16**	0.12	-0.06	-0.05	-0.23***	-0.18**	1271
Has many new friends in class	0.08	0.05	0.00	-0.03	-0.08	-0.05	1275

Table 4.5 – Parenting styles and skills - further results

The table depicts the effect of each parenting style on child's skills. All controls are summarized in table A2. Initial endowment of personality traits are measured using the Big Five of the interviewed parent. Outcomes are normalized to have mean 0 and standard deviation 1. Significance of the coefficients at conventional significance levels $1\%, 5\%, 10\%$ are indicated by stars ***, **, *, respectively. The last column N shows the number of observations.

The results in the upper panel show how the child's personality traits are associated with the parenting style. Children who are raised with an *authoritative* or a *democratic-loving* style are more conscientious, more agreeable, more open and less neurotic than children who are raised with an *authoritarian-inconsistent* style. Children with *authoritarian* parents are less open than children of *democratic-loving* or *authoritative* parents, but more agreeable and more conscientious than children who are raised with an *authoritarian-inconsistent* style. An *authoritative* style is associated with a higher extraversion compared to all other styles.

In the remaining panels, we analyze how parenting styles are related to how children cope with their everyday school life. Children with *authoritarian-inconsistent* parents are less autonomous, have less pleasure in learning, show less willingness to make an effort and are worse integrated in the class than those with *authoritative* or *democratic-loving* parents. Further, they are also less autonomous, have less fun in school and show less willingness to make an effort than children who are raised with an *authoritarian* style. Compared to an *authoritarian* style, children with *authoritative* parents can cope with many tasks more easily, have fun studying more often, try hard if a task is difficult more often and have many friends in class more often. Children with *democratic-loving* parents treat their working material more careful, are better integrated and have more friends in class. Whether parents raise their child with an *authoritative* or *democratic-loving* style is not related to how the child copes with everyday school life.

In summary, we find that both non-cognitive and cognitive skills are sensitive to the parenting style. The results shown in table 4.4 indicate that differences in non-cognitive skills are more pronounced than in cognitive skills. In general, an *authoritative* and a *democratic-loving* style are associated with similar cognitive and non-cognitive skills. Children with *authoritative* parents have a lower listening comprehension but are more patient than children with *democratic-loving* parents. In comparison, both an *authoritarian* and an *authoritarian-inconsistent* style are systematically associated with lower skills. Children with *authoritarian-inconsistent* parents have lower skills compared to all three alternative parenting styles. A *democratic-loving* style is associated with less patience than any other parenting style, even an *authoritarian-inconsistent* style.

4.7 Discussion

Recent literature established a strong link between children's skill development and parental monetary and time investments. In this paper, we focus on the role of parenting styles, a type of parental investment that has only recently become the focus of economic research. We use a novel latent class model (LDA-S, Munro and Ng, 2022) to investigate which parenting styles can actually be observed in the data. The model directly incorporates a link between the latent classes, i.e. parenting styles, and parental education, household income and household composition. We identify four parenting styles. Two styles closely resemble Baumrind's (1991) authoritative and authoritarian style. The other two are variations of these styles. We find that parenting styles are strongly associated with household income, education and whether the child is an only child. The results suggest that constraints in both time and (non-)cognitive skills of the parents play an important role in choosing a parenting style. Analyzing how the observed styles are associated with the child's (non-)cognitive skill development, we find that children raised with an authoritative or a democratic-loving style have the most favorable outcomes. Our results show how differences in parenting styles contribute to the skill gap between children from different socioeconomic environment. Parenting styles that are associated with low household income and having more than one child are associated with lower skills of the child. As much of the literature, we rely on observational data to estimate the effect of parenting styles on skills. Therefore, one has to keep in mind that our results can only be interpreted as causal under the strong assumption that we control all factors that affect both, the choice of the parenting style and the change in skills of the child between age 4 and 7.

Our paper gives important implications for future research. As parenting styles are not directly observable, we emphasize the challenge to operationalize them in future research, a point also made by Doepke and Zilibotti (2021). Our results suggest three important considerations. First, most commonly, the researcher does not know in which dimensions parenting styles differ. Our model identifies an authoritative and authoritarian parenting style in the sense of Baumrind (1971, 1991). The two other styles are variations of them which would have been overlooked if we relied solely on the classic theoretical framework. Therefore, along with theoretical models, data-driven approaches are a crucial tool for identifying parenting styles. Second, it is important to rely on a large set of different dimensions. This helps to properly describe the latent variables or classes and to fully understand the differences between them. More importantly, an extensive set of dimensions is crucial to separate parenting styles that are similar to each other. Our results show that the parenting styles may only differ regarding a few dimensions (e.g. authoritative vs. democratic-loving or authoritarian vs. authoritarian-inconsistent). Missing dimensions which are important could lead to misleading results. In our case, a lack of distinction between authoritarian from authoritarian-inconsistent parents would make authoritarian parenting appear worse than it actually is. Authoritarian-inconsistent parenting is associated with much less favorable outcomes. Third, data-driven methods which can handle many different dimensions in an interpretable way, such as LDA-S, are a crucial tool to handle a large set of measures for parent-child interaction.

Our paper also gives important directions for policy-makers. The recent literature finds that non-cognitive skills foster cognitive skills but not vice versa and that non-cognitive skills mainly develop in childhood and hardly change in adulthood (Cunha and Heckman, 2007, 2008; Cunha et al., 2010). Since parenting styles are strongly associated with non-cognitive skills, our results point to parenting styles as an important driver of the skill gap between children with different background. To reduce this gap, a policy measure, which may be easy to implement, could be to promote styles that are associated with the most favorable outcomes (authoritative or democratic-loving). However, the effectiveness of parents in implementing certain parenting styles may depend on their personal characteristics. Our results suggest that both (non-)cognitive skills and time resources of parents might limit the choice of parenting style. For example, some parents will find it harder to convince their child of their own opinion. Such parents may have difficulties to exercise an authoritative or democratic-loving style properly. Others might just not be able to give their full attention to only one child as they have more than one. Hence, they would not be able to apply time consuming styles, e.g. democratic-loving. Policy-makers could foster parents' (non-)cognitive skills which are important to raise a child or help parents to allocate their available time between children more efficiently.

Chapter 5

Dissertation conclusion

Over the last decades, machine learning became increasingly popular as a toolbox of methods for making precise predictions on a wide spectrum of different tasks. Despite their success, economists only slowly started to incorporate them in their research – maybe because of the conceptual difference between predictive and causal queries. As of now, the literature combining conventional econometric approaches with machine learning methods is growing fast and new methods to answer economic questions are developed and applied by practitioners.

In this doctoral thesis, I explore three novel methods by applying each of them to one relevant research question. Chapter 2 applies post-double-selection to investigate whether and how wage beliefs and information influence the decision to become a nurse. Post-double-selection is a comprehensible data-driven method which is easy to implement in a variety of settings involving the estimation of average effects. It improves upon conventional approaches through data-driven model selection to handle many potential controls relative to sample size, and to build models with a more flexible functional form. Moreover, post-double-selection increases the credibility of research by making model selection more traceable. This is in particular important since publication bias towards *statistically significant* results is still an issue (see e.g., Card and Krueger, 1995; Ashenfelter et al., 1999; Doucouliagos et al., 2012; Havránek, 2015). It has been shown that researchers respond to it by changing their behavior, and thus amplify the bias (e.g., Franco et al., 2014; Gelman and Loken, 2014; Brodeur et al., 2016).

In chapter 3, we use the generalized random forest framework to develop a 2SLS random forest. Generalized random forests flexibly model interactions between covariates in high dimensions.

Using our 2SLS random forest, we are able to plot detailed maps of heterogeneous effects across multiple dimensions which is not possible using standard regression techniques. Moreover, the framework addresses the model selection which plays a crucial role in the analysis of effect heterogeneity. There is much concern that the researcher systematically searches for subgroups with high effects and only report the results that highlight heterogeneity (Assmann et al., 2000). Therefore, researchers are often required to submit a pre-analysis plan specifying the subgroups that will be analyzed. This procedure restricts the researcher in finding unexpected but strong heterogeneity. Generalized random forests deal with the limitations of pre-analysis plans by relying on a data-driven, systematic search for effect heterogeneity while still providing valid asymptotic confidence intervals.

The last chapter applies latent dirichlet analysis for survey data to infer parenting styles. The model corresponds to a structural model of utility maximization which guides the interpretation of the model parameters. Therefore, we are able to identify well-interpretable latent structures from a large set of covariates.

In summary, machine learning methods applied in this doctoral thesis contribute to economic research in many ways. First, they allow to flexibly model the relationship between variables and to account for high-level interactions. Second, the methods are designed to handle a large number of variables. Third, most of the machine learning methods limit the freedom of the researcher in making rather arbitrary decisions. This makes empirical research more traceable and increases the trust in empirical work. Finally, new tools to analyze data entail new perspectives and new questions which can be answered. The ability to estimate personalized effects is the key to efficiently assign policies on an individual level (see chapter 3). Moreover, the machine learning literature provides methods for dimensionality reduction which lead to well-interpretable results despite their complexity (see chapter 4). As discussed in chapter 1 there are many other methods to estimate average effects, to estimate heterogeneous effects or to model latent variables (e.g., Chernozhukov et al., 2017; Athey et al., 2018b; Nie and Wager, 2021). All these methods have the same advantages as those discussed here.

Research at the intersection of economics and machine learning covers other topics which have not been discussed in this doctoral thesis. Because the data from which algorithms learn is not free from prejudice, concerns that algorithm guided decisions discriminate against certain groups become larger. Therefore, one branch of the literature tries to figure out how fairness of such algorithms can be ensured (Kleinberg et al., 2018; Rambachan et al., 2020). Another branch of the literature is about using machine learning methods to test assumptions. For example, Farbmacher et al. (2022) use a causal forest to detect local violations of instrument validity. A third branch of the literature uses machine learning methods for supplementary analysis. Athey and Imbens (2015), for example, develop a measure of robustness inspired by regression trees.

Machine learning methods for image and text analysis appear promising to generate and analyze new data. For example, applying machine learning methods for image recognition, satellite data and street map data can be used to predict poverty, safety, and property values at deep regional levels (e.g., Naik et al., 2017; Glaeser et al., 2018). Textual data can be used to access detailed product information for estimating supply and demand, or to analyze diary entries of individuals or interview data with open questions to evaluate social policies. Due to the unstructured nature of textual data, the analysis of texts has long been a topic in the machine learning community (see Gentzkow et al. 2019 for a review). Both, methods for image recognition and text analysis, are not (yet) among the tools commonly used by economists. Finally, economists pay less attention to the machine learning literature on causality. The reason for this may be because of the conceptual differences. In contrast to most of the econometric literature which relies on the potential outcome framework (Rubin, 1974), much of the machine learning literature relies on directed acyclic graphs (Pearl et al., 2000) to conduct causal inference. Imbens (2020) explains the differences between these approaches and discusses the reasons why economists seldom rely on the approach commonly adopted by the machine learning community.

Bibliography

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Alegre, A. (2011). Parenting styles and children's emotional intelligence: What do we know? *The Family Journal*, 19(1):56–62.

Altonji, J. G., Arcidiacono, P., and Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, volume 5, pages 305–396. Elsevier.

Andreassen, L., Di Tommaso, M. L., and Strøm, S. (2017). Nurses and physicians: A longitudinal analysis of mobility between jobs and labor supply. *Empirical Economics*, 52(4):1235–1269.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–477.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Ashenfelter, O., Harmon, C., and Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, 6(4):453–470.

Asher, S., Nekipelov, D., Novosad, P., and Ryan, S. P. (2016). Classification trees for heterogeneous moment-based models. Working Paper 22976, National Bureau of Economic Research.

Assmann, S. F., Pocock, S. J., Enos, L. E., and Kasten, L. E. (2000). Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet*, 355(9209):1064–1069.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, pages 507–547. University of Chicago Press.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021a). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Athey, S., Blei, D., Donnelly, R., Ruiz, F., and Schmidt, T. (2018a). Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers and Proceedings*, 108:64–67.

Athey, S., Friedberg, R., Mühlbach, N., Steimer, H., and Wager, S. (2020). Estimating heterogeneous treatment effects of an early retirement reform. Unpublished manuscript.

Athey, S. and Imbens, G. (2015). A measure of robustness to misspecification. *American Economic Review*, 105(5):476–480.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725.

Athey, S., Imbens, G. W., Metzger, J., and Munro, E. (2021b). Using wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*. Advance online publication. https://doi.org/10.1016/j.jeconom.2020.09.013.

Athey, S., Imbens, G. W., and Wager, S. (2018b). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

Attanasio, O. P. (2015). The determinants of human capital formation during the early years of life: Theory, measurement, and policies. *Journal of the European Economic Association*, 13(6):949–997.

Aunola, K. and Nurmi, J.-E. (2005). The role of parenting styles in children's problem behavior. *Child Development*, 76(6):1144–1159.

Aunola, K., Stattin, H., and Nurmi, J.-E. (2000). Parenting styles and adolescents' achievement strategies. *Journal of Adolescence*, 23(2):205–222.

Bajari, P., Nekipelov, D., Ryan, S. P., and Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5):481–485.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology*, 4(1):1–103.

Baumrind, D. (1991). The influence of parenting style on adolescent competence and substance use. *The Journal of Early Adolescence*, 11(1):56–95.

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and Economic Change in Developed Countries*, pages 209–240. Columbia University Press.

Becker, G. S. (1981). *A treatise on the family*. Harvard University Press.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3):S1–S39.

Beffy, M., Fougère, D., and Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics*, 94(1):334–347.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *Review of Economics and Statistics*, 81(2):608–650.

Berendes, K., Weinert, S., Zimmermann, S., and Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, 5(2):15–49.

Björkegren, D. and Grissen, D. (2017). Behavior revealed in mobile phone usage predicts loan repayment. *arXiv preprint arXiv:1712.05840*.

Blei, D. M. and Lafferty, J. D. (2009). Topic models. In *Text mining: Classification, clustering, and applications*, pages 71–89. CRC press.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(7):993–1022.

Blossfeld, H.-P. and von Maurice, J. (2011). Education as a lifelong process. *Zeitschrift für Erziehungswissenschaft*, 14(2):19–34.

Bomball, J., Schwanke, A., Stöver, M., Schmitt, S., and Görres, S. (2010). Imagekampagne für Pflegeberufe auf der Grundlage empirisch gesicherter Daten: Einstellungen von Schüler/innen zur möglichen Ergreifung eines Pflegeberufes. Ausgabe 05/2010, Institut für Public Health und Pflegeforschung.

Boneva, T. and Rauh, C. (2016). Human capital production and parental beliefs. Unpublished Manuscript.

Boudarbat, B. (2008). Field of study choice by community college students in Canada. *Economics of Education Review*, 27(1):79–93.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and regression trees*. Wadsworth.

Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.

Brunello, G., Lucifora, C., and Winter-Ebmer, R. (2004). The wage expectations of European business and economics students. *Journal of Human Resources*, 39(4):1116–1142.

Burton, P., Phipps, S., and Curtis, L. (2002). All in the family: A simultaneous model of parenting style and child conduct. *American Economic Review*, 92(2):368–372.

Card, D. and Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review*, 85(2):238–243.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. Working Paper 9546, National Bureau of Economic Research.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–265.

Chernozhukov, V., Fernández-Val, I., and Luo, Y. (2018). The sorted effects method: Discovering heterogeneous effects beyond their averages. *Econometrica*, 86(6):1911–1938.

Chernozhukov, V. and Hansen, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

Cho, S.-H., Jung, S. Y., and Jang, S. (2010). Who enters nursing schools and why do they choose nursing? A comparison with female non-nursing students using longitudinal data. *Nurse Education Today*, 30(2):180–186.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):39–67.

Cobb-Clark, D. A., Salamanca, N., and Zhu, A. (2019). Parenting style as an investment in human development. *Journal of Population Economics*, 32(4):1315–1352.

Cunha, F., Elo, I., and Culhane, J. (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working Paper 19144, National Bureau of Economic Research.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In *Handbook of the Economics of Education*, volume 1, pages 697–812. Elsevier.

Cunha, F., Heckman, J. J., and Navarro, S. (2005). Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191–261.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Dante, A., Rizzi, L., Ianderca, B., and Palese, A. (2013). Why do university students not choose a nursing degree at matriculation? An Italian cross-sectional study. *International Nursing Review*, 60(1):129–135.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Dizon-Ross, R. (2016). Parents' beliefs and children's education: Experimental evidence from Malawi. Unpublished Manuscript.

Doepke, M., Sorrenti, G., and Zilibotti, F. (2019). The economics of parenting. *Annual Review of Economics*, 11:55–84.

Doepke, M. and Zilibotti, F. (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5):1331–1371.

Doepke, M. and Zilibotti, F. (2021). Do rising returns to education justify "helicopter" parenting? *IZA World of Labor*.

Doiron, D., Hall, J., Kenny, P., and Street, D. J. (2014). Job preferences of students and new graduates in nursing. *Applied Economics*, 46(9):924–939.

Doucouliagos, C., Stanley, T. D., and Giles, M. (2012). Are estimates of the value of a statistical life exaggerated? *Journal of Health Economics*, 31(1):197–206.

Dudík, M., Erhan, D., Langford, J., and Li, L. (2014). Doubly robust policy evaluation and optimization. *Statistical Science*, 29(4):485–511.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*.

Falk, A., Kosse, F., Pinger, P., Schildberg-Hörisch, H., and Deckers, T. (2021). Socioeconomic status and inequalities in children's IQ and economic preferences. *Journal of Political Economy*, 129(9):2504–2545.

Farbmacher, H., Guber, R., and Klaassen, S. (2022). Instrument validity tests with causal forests. *Journal of Business & Economic Statistics*, 40(2):605–614.

Fernandes, A., Huber, M., and Vaccaro, G. (2020). Gender differences in wage expectations. *arXiv preprint arXiv:2003.11496*.

Fiorini, M. and Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, 32(4):787–836.

Francesconi, M. and Heckman, J. J. (2016). Child development and parental investment: Introduction. *The Economic Journal*, 126(596):F1–F27.

Franco, A., Malhotra, N., and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203):1502–1505.

Gelman, A. and Loken, E. (2014). The statistical crisis in science data-dependent analysis – a "garden of forking paths" – explains why many statistically significant comparisons don't hold up. *American Scientist*, 102(6):460.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.

German Employment Agency (2018). Entgeltatlas. https://entgeltatlas.arbeitsagentur.de/entgeltatlas.

German Employment Agency (2020). Arbeitsmarkt im Pflegebereich. Berichte: Blickpunkt Arbeitsmarkt 05/2020.

Glaeser, E. L., Kominers, S. D., Luca, M., and Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, 56(1):114–137.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 1–9. Curran Associates, Inc.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5):581–586.

Gopalan, P., Hofman, J. M., and Blei, D. M. (2015). Scalable recommendation with hierarchical poisson factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 326–335. AUAI Press.

Grammig, J., Hanenberg, C., Schlag, C., and Sönksen, J. (2020). Diverging roads: Theory-based vs. machine learning-implied stock risk premia. Working Papers in Business and Economics, University of Tübingen.

Green, D. P. and Kern, H. L. (2012). Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public Opinion Quarterly*, 76(3):491–511.

Hagenaars, A. J., De Vos, K., Asghar Zaidi, M., et al. (1994). Poverty statistics in the late 1980s: Research based on micro-data. Publications office, European Commission.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46.

Hanel, B., Kalb, G., and Scott, A. (2014). Nurses' labour supply elasticities: The importance of accounting for extensive margins. *Journal of Health Economics*, 33:94–112.

Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep IV: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1414–1423. PMLR.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, 13(6):1180–1204.

Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2):261–294.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, 6:689–733.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24.

Ichino, A., Lindström, E.-A., and Viviano, E. (2014). Hidden consequences of a first-born boy for mothers. *Economics Letters*, 123(3):274–278.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263.

Imbens, G. and Athey, S. (2021). Breiman's two cultures: A perspective from econometrics. *Observational Studies*, 7(1):127–133.

Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, 58(4):1129–1179.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Kaji, T., Manresa, E., and Pouliot, G. (2020). An adversarial approach to structural estimation. *arXiv preprint arXiv:2007.06169*.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5):491–495.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, 108:22–27.

Knaus, M. C. (2020). Double machine learning based program evaluation under unconfoundedness. *arXiv preprint arXiv:2003.03191*.

Knoll, B., Riedel, N., and Schlenker, E. (2017). He's a chip off the old block - the persistence of occupational choices across generations. *Labour*, 31(2):174–203.

Li, L., Chen, S., Kleban, J., and Gupta, A. (2015). Counterfactual estimation and optimization of click metrics in search engines: A case study. In *Proceedings of the 24th International Conference on World Wide Web*, pages 929–934. Association for Computing Machinery.

Lin, Y. and Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, 101(474):578–590.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.

Lundberg, S., Romich, J. L., and Tsang, K. P. (2009). Decision-making by children. *Review of Economics of the Household*, 7(1):1–30.

Masud, H., Thurasamy, R., and Ahmad, M. S. (2015). Parenting styles and academic achievement of young adolescents: A systematic literature review. *Quality & Quantity*, 49(6):2411–2433.

Matthes, S. (2019). *Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions-und Aversionsfaktoren in der Berufsfindung*. Verlag Barbara Budrich.

McCabe, R., Nowak, M., and Mullen, S. (2005). Nursing careers: What motivated nurses to choose their profession? *Australian Bulletin of Labour*, 31(4):380–402.

McCaffrey, D. F., Ridgeway, G., and Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9(4):403–425.

Meinshausen, N. and Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(6):983–999.

Montmarquette, C., Cannings, K., and Mahseredjian, S. (2002). How do young people choose college majors? *Economics of Education Review*, 21(6):543–556.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Munro, E. and Ng, S. (2022). Latent dirichlet analysis of categorical survey responses. *Journal of Business & Economic Statistics*, 40(1):256–271.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., and Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576.

NEPS (2020). Information on competence testing – wave 7. NEPS Starting Cohort 1 – Newborns education from the very beginning, LIfBi.

NEPS (2021a). Information on competence testing – wave 8. NEPS Starting Cohort 1 – Newborns education from the very beginning, LIfBi.

NEPS (2021b). National Educational Panel Study, Scientific Use File of Starting Cohort Newborns. *LIfBi Leibniz Institute for Educational Trajectories*.

Nie, X. and Wager, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Pearl, J. et al. (2000). *Models, reasoning and inference*. Cambridge University Press.

Petrucci, C., La Cerra, C., Aloisio, F., Montanari, P., and Lancia, L. (2016). Empathy in health professional students: A comparative cross-sectional study. *Nurse Education Today*, 41:1–5.

Rambachan, A., Kleinberg, J., Ludwig, J., and Mullainathan, S. (2020). An economic perspective on algorithmic fairness. *AEA Papers and Proceedings*, 110:91–95.

Robins, J., Sued, M., Lei-Gomez, Q., and Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, 22(4):544–559.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). Dicekriging, diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Ruiz, F. J., Athey, S., and Blei, D. M. (2020). Shopper: A probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics*, 14(1):1–27.

Schweri, J. and Hartog, J. (2017). Do wage expectations predict college enrollment? Evidence from healthcare. *Journal of Economic Behavior & Organization*, 141:135–150.

Sexton, J. and Laake, P. (2009). Standard errors for bagged and random forest estimators. *Computational Statistics & Data Analysis*, 53(3):801–811.

Shields, M. A. (2004). Addressing nurse shortages: What can policy makers learn from the econometric evidence on nurse labour supply? *The Economic Journal*, 114(499):F464–F498.

Speer, J. D. (2020). Where the girls are: Examining and explaining the gender gap in the nursing major. *Scottish Journal of Political Economy*, 67(3):322–343.

Stock, J. H. and Watson, M. (2011). Oxford University Press.

Strehl, A., Langford, J., Li, L., and Kakade, S. M. (2010). Learning from logged implicit exploration data. In *Advances in neural information processing systems*, pages 1–9. Curran Associates, Inc.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., and Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10(2):141–158.

Swaminathan, A. and Joachims, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *The Journal of Machine Learning Research*, 16(1):1731–1755.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567.

Urminsky, O., Hansen, C., and Chernozhukov, V. (2016). Using double-lasso regression for principled variable selection. Available at SSRN: https://ssrn.com/abstract=2733374.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Weinberg, B. A. (2001). An incentive model of the effect of parental income on children. *Journal of Political Economy*, 109(2):266–280.

Wilkes, L., Cowin, L., and Johnson, M. (2015). The reasons students choose to undertake a nursing degree. *Collegian*, 22(3):259–265.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Economics of Education Review*, 82(2):791–824.

Wohlkinger, F., Blumenfelder, A. R., Bayer, M., Maurice, J. v., Ditton, H., and Blossfeld, H.-P. (2019). Measuring motivational concepts and personality aspects in the National Educational Panel Study. In *Education as a Lifelong Process*, pages 155–169. Springer.

Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M., and Blossfeld, H.-P. (2011). Motivational concepts and personality aspects across the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2):155.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Wyss, R., Ellis, A. R., Brookhart, M. A., Girman, C. J., Jonsson Funk, M., LoCasale, R., and Stürmer, T. (2014). The role of prediction modeling in propensity score estimation: An evaluation of logistic regression, bCART, and the covariate-balancing propensity score. *American Journal of Epidemiology*, 180(6):645–655.

Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2):492–514.

Zhou, Z., Athey, S., and Wager, S. (2022). Offline multi-action policy learning: Generalization and optimization. *Operations Research*. Advance online publication. https://doi.org/10.1287/opre.2022.2271.

Zubizarreta, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922.

Appendix A

Tables

Variable	Mean	Standard deviation	Min	Max	Number of non-missing values	Share of missing values
Demographic						
Gender: female	0.4978	0.5000	0.0	1.0	7089	0.00
Migration background	0.1035	0.3047	0.0	1.0	7089	0.00
Opportunities						
Higher secondary track in 9th grade	0.3957	0.4890	0.0	1.0	6907	0.03
Competencies						
Science	0.1561	0.9793	-2.6	5.3	6985	0.01
Mathematics	0.2077	1.2220	-4.4	4.6	7002	0.01
Information and communication technology	0.1461	0.9002	-3.3	4.1	6989	0.01
Reading	34.6639	8.3705	0.0	51.0	7003	0.01
Reading speed	0.1577	1.2019	-4.0	3.3	6950	0.02
Metacognition	0.8185	0.1175	0.0	1.0	6965	0.02
Attitude to school & school performance						
Math grade: over average (class)	0.5445	0.4981	0.0	1.0	6981	0.02
German grade: over average (class)	0.5083	0.5000	0.0	1.0	7016	0.01
Grade: math	2.8770	0.9989	1.0	6.0	6981	0.02
Grade: german	2.8077	0.7970	1.0	6.0	7016	0.01
Ever retent a grade	0.1534	0.3604	0.0	1.0	6984	0.01

Table A1 continued from previous page

School concept: german	2.9434	0.6233	1.0	4.0	7022	0.01
School concept: math	2.5750	0.9186	1.0	4.0	6993	0.01
School concept: general	2.9264	0.5630	1.0	4.0	7019	0.01
Interests in math	2.2180	0.7900	1.0	4.0	6862	0.03
Interests in german	2.1874	0.8047	1.0	4.0	6865	0.03
Personality & behavior						
Big Five: artistic	2.7984	1.3196	1.0	5.0	7070	0.00
Big Five: crticize	2.8645	1.0243	1.0	5.0	7068	0.00
Big Five: easy-going/lazy	3.2171	1.1723	1.0	5.0	7074	0.00
Big Five: nervous	2.8271	1.0795	1.0	5.0	7066	0.00
Big Five: imaginative	3.7465	1.0257	1.0	5.0	7062	0.00
Big Five: relaxed	3.2995	1.0612	1.0	5.0	7072	0.00
Big Five: cautious/relaxed	2.6416	1.1134	1.0	5.0	7071	0.00
Big Five: sensitive	3.8503	0.9112	1.0	5.0	7068	0.00
Big Five: sociable	3.5319	0.9446	1.0	5.0	7061	0.00
Big Five: thorough	3.5920	0.9284	1.0	5.0	7071	0.00
Big Five: trusting	3.3983	1.0098	1.0	5.0	7068	0.00
Considerate	2.6108	0.5218	1.0	3.0	7011	0.01
Gets mobbed	1.1665	0.4243	1.0	3.0	6980	0.02
Has friends	2.8863	0.3434	1.0	3.0	7000	0.01
Helpful	2.6887	0.4996	1.0	3.0	6997	0.01
Kind to younger	2.4744	0.5924	1.0	3.0	6992	0.01
Likes to help	2.2183	0.5872	1.0	3.0	6994	0.01
Loner	1.4503	0.6032	1.0	3.0	6986	0.01
Popular	2.3546	0.5770	1.0	3.0	6940	0.02
Global self-esteem	2.4496	1.1497	1.0	5.0	7004	0.00
Likes to share	2.5696	0.5503	1.0	3.0	7005	0.01
Gets along with adults	1.6721	0.6596	1.0	3.0	6990	0.01
Good as others	3.9471	0.7928	1.0	5.0	7068	0.00
Be a failure	1.6985	0.9576	1.0	5.0	7045	0.00
Good qualities	3.9625	0.7622	1.0	5.0	7060	0.00
No pride	2.0150	0.9865	1.0	5.0	7059	0.00
Positive attitude towards myself	3.9402	0.9088	1.0	5.0	7056	0.00
Satisfied with myself	3.9442	0.8371	1.0	5.0	7078	0.00
No good	2.3150	1.0705	1.0	5.0	7053	0.01
Feel useless	1.8723	1.0089	1.0	5.0	7057	0.01
Be at least as valuable as others	4.0033	0.9849	1.0	5.0	7054	0.01

Table A1 continued from previous page

TenFlex: flexible	16.0394	3.2374	5.0	25.0	7007	0.01
TenFlex: persistent	18.4130	2.9034	5.0	25.0	7063	0.00
Religious	2.2474	0.8976	1.0	4.0	6810	0.04
Disadvantage: gender	0.0740	0.2617	0.0	1.0	6341	0.11
Disadvantage: foreign name	0.3457	0.4756	0.0	1.0	6411	0.11
Disadvantage: foreign looks	0.3505	0.4772	0.0	1.0	6400	0.10
Disadvantage: lower secondary	0.7906	0.4069	0.0	1.0	6620	0.07
Disadvantage: head scarf	0.5563	0.4969	0.0	1.0	6052	0.15
Disadvantage: overweight	0.2038	0.4028	0.0	1.0	6370	0.10
Disadvantage: bad german	0.8574	0.3497	0.0	1.0	6583	0.07
Family & career planning						
Important to form family	0.6812	0.4661	0.0	1.0	7085	0.00
Child before age 25	0.2307	0.4213	0.0	1.0	7079	0.00
Moving away for training	0.4160	0.4929	0.0	1.0	6019	0.15
Satisfaction						
Satisfaction with life	7.5462	1.9546	0.0	10.0	7089	0.00
Satisfaction with living conditions	8.0968	1.8901	0.0	10.0	7089	0.00
Satisfaction with family	8.3861	2.1703	0.0	10.0	7089	0.00
Satisfaction with friends	8.6148	1.8390	0.0	10.0	7089	0.00
Satisfaction with school	6.8558	2.2203	0.0	10.0	7089	0.00
Satisfaction with health	8.3168	2.0770	0.0	10.0	7089	0.00
Leisure						
Time gaming	3.0198	1.5115	1.0	6.0	6910	0.03
Time reading	3.1293	1.4694	1.0	5.0	6927	0.02
Visiting museum	2.2284	1.0784	1.0	5.0	7057	0.00
TV-shows: science	1.9890	0.7473	1.0	4.0	7024	0.01
Books: science	1.4038	0.6478	1.0	4.0	7024	0.00
Web: science	1.7473	0.7735	1.0	4.0	7012	0.01
Magazines: science	1.7091	0.7909	1.0	4.0	7013	0.01
Science club	1.1463	0.4802	1.0	4.0	7020	0.01
Course: music	1.7950	0.4037	1.0	2.0	7089	0.00
Number of books	3.9537	1.4359	1.0	6.0	7064	0.00
Meaning of work and interests						
Importance of comfort aspects	4.6524	0.9508	1.0	6.0	7089	0.00
Importance of economic aspects	5.1635	0.7465	1.0	6.0	7089	0.00
Importance of expressive aspects	4.9322	0.6508	1.0	6.0	7052	0.00
IILS-Interests: social	3.0449	0.9829	1.0	5.0	7089	0.00

Table A1 continued from previous page

IILS-Interests: conventional	2.5018	0.8550	1.0	5.0	7057	0.00
IILS-Interests: art	2.5329	1.0149	1.0	5.0	7065	0.00
IILS-Interests: analytical	2.6614	0.9723	1.0	5.0	7076	0.00
IILS-Interersts: practical	2.8324	1.0586	1.0	5.0	7066	0.00
IILS-Interests: business	3.0338	0.8357	1.0	5.0	7060	0.00
Parental background						
Parental education (highest): studied	0.2918	0.4546	0.0	1.0	5452	0.23
Parental education (highest): university entrance quali.	0.1970	0.3978	0.0	1.0	5452	0.23
Household income per capita	859.8624	392.4184	200.0	2666.7	4198	0.41
Parental occupation (at least one parent): MINT	0.5144	0.4998	0.0	1.0	5476	0.23
Parental occupation (at least one parent): business	0.5338	0.4989	0.0	1.0	5476	0.23
Parental occupation (at least one parent): care	0.0942	0.2922	0.0	1.0	5476	0.23
Parental occupation (at least one parent): health	0.1348	0.3415	0.0	1.0	5476	0.23
Parental occupation (at least one parent): education	0.1715	0.3770	0.0	1.0	5476	0.23
Parental occupation (at least one parent): hairdresser	0.0197	0.1391	0.0	1.0	5476	0.23
Parental occupation (at least one parent): banking	0.0499	0.2177	0.0	1.0	5476	0.23
Parental occupation (at least one parent): automotive mechanic	0.0268	0.1616	0.0	1.0	5476	0.23
Parental occupation (at least one parent): teacher	0.0575	0.2329	0.0	1.0	5476	0.23
Parental occupation (at least one parent): physician	0.0247	0.1551	0.0	1.0	5476	0.23
Broken home	0.0900	0.2862	0.0	1.0	6832	0.04
Behavior and values of parents						
Discuss books	1.8053	1.0344	1.0	5.0	6930	0.02
Discuss movies	3.2535	1.1332	1.0	5.0	6928	0.02
Discuss politcs	2.5871	1.2831	1.0	5.0	6943	0.02
Discuss arts	1.5531	0.9378	1.0	5.0	6950	0.02
Importance to maintain mother's status (education)	3.6373	1.3416	1.0	5.0	6515	0.08

Table A1 continued from previous page

Importance to maintain father's status (education)	3.6483	1.3606	1.0	5.0	6381	0.10
Importance of grades	4.3354	0.8863	1.0	6.0	7015	0.01
Importance of parent's opinion	3.9303	0.9589	1.0	5.0	7013	0.01
Gender role: duties in household	3.2981	0.8462	1.0	4.0	6032	0.15
Gender role: technology	2.7174	0.8996	1.0	4.0	5937	0.02
Gender role: politics	3.2233	0.8659	1.0	4.0	6883	0.16
Gender role: earning money	1.8850	0.9479	1.0	4.0	6992	0.01
Gender role: occupations	3.0044	0.9044	1.0	4.0	6991	0.01
Importance career	4.0437	1.1019	0.0	5.0	7004	0.01
Importance to maintain mothers status (occupation)	3.7709	1.2631	1.0	5.0	6966	0.02
Importance to maintain fathers status (occupation)	3.7208	1.2600	1.0	5.0	6923	0.02
Expectations of son: living close	2.0428	0.7851	1.0	4.0	6334	0.02
Expectations of son: housekeeping	2.5314	0.9255	1.0	4.0	6536	0.08
Expectations of son: financially support younger siblings	1.8419	0.8128	1.0	4.0	6235	0.08
Expectations of daughter: living close	2.2826	0.8827	1.0	4.0	6295	0.11
Expectations of daughter: housekeeping	2.8661	0.8996	1.0	4.0	6475	0.11
Expectations of daughter: financially support younger siblings	1.8036	0.7886	1.0	4.0	6104	0.14
Expectations to study	0.4272	0.4947	0.0	1.0	6645	0.06
Costs of lower secondary degree	3.3527	1.1535	1.0	5.0	6941	0.02
Costs of middle secondary degree	3.7966	0.8572	1.0	5.0	6914	0.02
Costs of higher secondary degree	3.9547	1.0164	1.0	5.0	6908	0.03
Social environment						
Organization: sports	0.6558	0.4751	0.0	1.0	6990	0.01
Organization: religion	0.2136	0.4099	0.0	1.0	6946	0.02
Organization: culture	0.1434	0.3505	0.0	1.0	6932	0.02
Friends: share migration background	2.6436	1.3255	1.0	7.0	7085	0.00
Friends: share ambitious	3.1854	0.7628	1.0	5.0	7012	0.01
Friends: share try	2.7581	1.0162	1.0	5.0	7068	0.00
Friends: share don't care	2.5054	0.9814	1.0	5.0	7012	0.01

Table A1 continued from previous page

Friends: important to have a career	3.6128	0.8680	1.0	5.0	7066	0.00
Class: share migration background	2.6724	1.1206	1.1206	7.01	6979	0.02
Class: share ambitious	3.0810	0.7637	1.0	5.0	6999	0.01
Class: share try	2.4859	0.9372	1.0	5.0	7065	0.00
Class: share don't care	2.7202	0.9269	1.0	5.0	6987	0.01
School						
Teacher: further education about voc. orientation	3.1783	0.9487	1.0	5.0	5681	0.20
Contact: organization	3.6825	0.9318	1.0	5.0	5603	0.21
Contact: firms	3.9968	0.9016	1.0	5.0	5642	0.20
Programs for voc. orientation	4.1166	0.9513	1.0	5.0	5704	0.19
Contact: counseling	3.8845	0.9966	1.0	5.0	5669	0.20
Contact: local network	3.6833	1.1447	1.0	5.0	5671	0.20
Parental support in voc. orientation	3.6765	0.9647	1.0	5.0	5685	0.20
Testing of interests	4.2433	1.7749	1.0	6.0	5474	0.23
Individual support plans	2.5473	1.5010	1.0	6.0	5375	0.24
Voc. orientation by teachers	4.9588	1.6463	1.0	6.0	5458	0.23
Practice: writing applications	5.7043	0.9610	1.0	6.0	5523	0.22
Practice: job interview	5.1684	1.4346	1.0	6.0	5468	0.23
Train social competencies	4.3466	1.8184	1.0	6.0	5366	0.24
Assisted internship	5.1491	1.6617	1.0	6.0	5487	0.23
External counseling	4.6583	1.7494	1.0	6.0	5502	0.22
Voc. orientation in institutions	3.4604	1.9691	1.0	6.0	5426	0.23
Individual counseling	3.1498	1.6759	1.0	6.0	5499	0.22
Individual support by career choice assistance	2.2650	1.5128	1.0	6.0	5399	0.24
Support by educators	2.0501	1.1373	1.0	6.0	5472	0.23
Regional and labor market characteristics						
Share age 15 to 25	11.7035	0.7721	9.9	14.4	7072	0.00
Firm density	42.0207	39.1636	2.4	186.7	7072	0.00
Regional unemployment rate	2.4753	1.0799	1.0	4.0	7072	0.00
Residence in east-germany	0.1251	0.3308	0.0	1.0	6835	0.04

Table A1 – Summary statistics of potential controls

	Mean	Standard deviation	Min	Max
Skills at age 4				
Reasoning	0.3037	2.3476	-4.1	6.1
Mathmatical literacy	0.0444	0.9813	-3.6	3.2
Listening Comprehension	49.5915	24.0283	0.0	121.0
Prosocial behavior (SDQ)	7.7081	1.4843	1.0	10.0
Problem behavior (SDQ)	1.2479	1.3648	0.0	8.0
Patience	0.7787	0.3954	0.0	1.0
Short term memory	2.3747	0.9009	0.0	5.0
Frustration (temperament)	4.2296	1.3682	0.0	6.0
Is concentrated (temperament)	4.5126	1.2958	0.0	6.0
Feeling down when failing (temperament)	3.0783	1.5319	0.0	6.0
Gets lost in books (temperament)	4.6621	1.4315	0.0	6.0
Full of energy in the evening (temperament)	4.9430	1.3063	0.0	6.0
Difficult to calm down (temperament)	2.9776	1.5640	0.0	6.0
Parental skills and preferences				
Risk tolerance	4.6654	2.0122	0.0	10.0
Patience	5.6288	2.2663	0.0	10.0
Conscientiousness (Big Five)	4.0765	0.6682	1.0	5.0
Extraversion (Big Five)	3.5510	0.8313	1.0	5.0
Agreeableness (Big Five)	3.6374	0.5472	1.7	5.0
Openness (Big Five)	3.7362	0.8855	1.0	5.0
Neuroticism (Big Five)	2.8150	0.7745	1.0	5.0
Peers				
Parents' peers: University degree	4.4034	1.5460	1.0	7.0
Parents' peers: Migration background	2.9817	1.3385	1.0	7.0
Child's peers: Migration background	3.2237	1.4099	1.0	7.0
Quality time				
Quality time	9.0064	1.5770	2.5	28.3
Demographic characteristics				
Child female	0.4974	0.5002	0.0	1.0
Child migration background	0.1262	0.3321	0.0	1.0
Responded female	0.4662	0.4990	0.0	1.0
State of Residence: SH	0.0438	0.2047	0.0	1.0
State of Residence: HH	0.0647	0.2461	0.0	1.0
State of Residence: NI	0.0752	0.2637	0.0	1.0
State of Residence: HB	0.0222	0.1475	0.0	1.0
State of Residence: NW	0.2301	0.4210	0.0	1.0

Table A2 continued from previous page

State of Residence: HE	0.0569	0.2317	0.0	1.0
State of Residence: RP	0.0281	0.1653	0.0	1.0
State of Residence: BW	0.0915	0.2884	0.0	1.0
State of Residence: BY	0.1641	0.3704	0.0	1.0
State of Residence: SL	0.0137	0.1164	0.0	1.0
State of Residence: BE	0.1078	0.3103	0.0	1.0
State of Residence: BB	0.0124	0.1108	0.0	1.0
State of Residence: MV	0.0163	0.1268	0.0	1.0
State of Residence: SN	0.0425	0.2018	0.0	1.0
State of Residence: ST	0.0235	0.1516	0.0	1.0
State of Residence: TH	0.0072	0.0845	0.0	1.0
Group: No sibling, low income, no univ.	0.0333	0.1796	0.0	1.0
Group: No sibling, low income, univ.	0.0144	0.1191	0.0	1.0
Group: No sibling, middle income., no univ.	0.0477	0.2132	0.0	1.0
Group: No sibling, middle income, univ.	0.0242	0.1537	0.0	1.0
Group: No sibling, high income, no univ.	0.0255	0.1577	0.0	1.0
Group: No sibling, high income, univ.	0.0660	0.2484	0.0	1.0
Group: Sibling, low income, no univ.	0.1614	0.3681	0.0	1.0
Group: Sibling., low income, univ.	0.0725	0.2595	0.0	1.0
Group: Sibling., middle income, no univ.	0.0915	0.2884	0.0	1.0
Group: Sibling., middle income, univ.	0.1660	0.3722	0.0	1.0
Group: Sibling., high income, no univ.	0.0608	0.2390	0.0	1.0
Group: Sibling, high income, univ.	0.2366	0.4251	0.0	1.0
N	1530			

Table A2 – Summary statistics of controls