# The effect of education on health

## Estimating the heterogeneous treatment effects of education on depression with Causal IV Forests

Magnus Lindgaard Nielsen, WKG579

Autumn 2021

## Abstract

I provide new evidence on the causal effect of education on health using compulsory schooling reforms to obtain exogenous identification of education. I estimate the local average treatment effect to be a reduction in depressive symptoms by 0.11 per extra year of education induced by the schooling reforms, but the effect is not statistically significant. I also show that there exist a lot of heterogeneity in the effects sizes, with the most robust finding being that women benefit most from additional education, but also observe indications that people with lower socioeconomic status also benefit more, but all of the findings are associated with wide confidence intervals. The large amounts of heterogeneity which varies across countries and other covariates indicate that one should be wary of extrapolating findings in the literature to new settings without further research.

**Characters including spaces:** 35.927
**Supervisor:** Torben Heien Nielsen
**Hand-in date:** 01/12/2021
**Course:** Seminar: Empirical Health Economics

## Contents

# Introduction

A correlation between education and health has been shown in prior research, but whether education has a causal effect on health is less clear, with different findings across countries, settings and health outcomes (Cutler & Lleras-Muney, 2006).

In this paper, I add to the evidence of a causal effect of education on health using compulsory schooling length reforms. This has been done before, most famously for the U.S. by Lleras-Muney (2002), but also in Europe (see e.g. Crespo, López-Noval, and Mira (2014)). However, it has been shown by Mazumder (2010) that the effect found in Lleras-Muney (2002) disappears when controlling for state-specific trends, either due to a lack of variation left after controlling for the cohort trends or due to no causal relationship between. The health measure of interest is depressive symptoms, as it is estimated that 5% of all adults globally suffer from depression (World Health Organization, 2021) and WHO estimates that depression will become the largest disease burden by 2030 (World Health Organization, 2011).

The research question governing this paper is twofold. The first part relates to whether an effect exists: 1) *Does more education cause a reduction in depressive symptoms?*, and the second part relates to how this effect varies with childhood factors: 2) *How does the effect of education on depressive symptoms vary with childhood factors?*

I utilize the Grossman model and the efficient producer hypothesis to motivate that a causal link from education to health and depression exists. My data consists of the easySHARE dataset (Börsch-Supan & Gruber, 2020), which is a longitudinal survey, expanded with childhood factors from SHARE wave 3 and 7 (Börsch-Supan, 2019, 2020). Identification of causal effects is obtained in 7 different European countries using instrumental variables with compulsory school reforms compiled by Crespo et al. (2014). Local average treatment effects are obtained using two-stage least-squares (2SLS) (J. D. Angrist & Pischke, 2008) and conditional local average treatment effects are estimated using causal IV forests (Athey, Tibshirani, & Wager, 2019). This allows me to examine how the treatment effects change with socioeconomic status in the childhood home, rather than relying on a more binary subgroup analysis using 2SLS.

In my preferred specification, I find that an extra year of education results in a reduction of depressive symptoms by 0.11 for people who were affected by the schooling reforms, with the unconditional local average treatment effects being insensitive to different robustness checks, but all are associated with wide confidence intervals which almost always cross 0 at the 95% confidence level. The conditional local average treatment effects estimate reductions in depressive symptoms by 0.3 to 0.1 for compliers, with women and people from low socioeconomic status homes generally experiencing larger reductions in depressive symptoms. These findings, however, are also almost always statistically insignificant and are generally not robust and thus hard to extrapolate from.

My findings indicate that education does reduce depressive symptoms, but that there exists a lot of heterogeneity which we are not able to model with sufficient precision and robustness to make strong claims when controlling for trends. The large amount of heterogeneity also means that one should be wary of generalizing causal findings of the influence of education on health from one setting to another, as the effects do not seem to be constant across many covariates, including both socioeconomic status, gender and country in this specific setting.

Section 2 presents the Grossman model and the efficient producer hypothesis. Section 3 presents the econometric models used. Section 4 presents the data and descriptive statistics. Section 5 presents and interprets the results. Section 6 concludes.

## Conceptual Framework

The conceptual framework I use to motivate the causal relationship between education and health is the Grossman model, see Bhattacharya, Hyde, and Tu (2013) for a more thorough walkthrough, and the *efficient producer hypothesis.*

The Grossman model posits the following health production function, $H(\cdot)$ (eq 3.7 in Bhattacharya et al. (2013)):

$$H_t = H\left([1-\gamma]H_{t-1}, T_t^H, M_T\right) \tag{1}$$

Where $H_t$ is health capital at time $t$, $\gamma$ is the depreciation rate of health capital, $T_t^H$ is time spent on health improvement at time $t$ and $M_t$ is health-related market inputs bought at time $t$. The cost of health is the real rental rate plus the depreciation rate of health, $r+\gamma$, which I assume is equal for everyone.

The efficient producer hypothesis then states that well educated people are more efficient producers of health, i.e. that $\frac{\partial H_t}{\partial T_t^H}$ and $\frac{\partial H_t}{\partial M_t}$ is increasing in education, $E_t$. As all people face the same cost of health, $r+\gamma$, this results in a higher level of health for a higher educated person, all else equal. As such, I expect that an increase in education will result in an increase in health, which for the given health measure, depression symptoms, is equal to a reduction in depressive symptoms. To extend this model to accommodate heterogeneous treatment effects, I furthermore assume that how $\frac{\partial H_t}{\partial T_t^H}$ and $\frac{\partial H_t}{\partial M_t}$ changes with education, $E_t$, varies with other covariates.

The posited mechanism behind the relationship between $\frac{\partial H_t}{\partial T_t^H}$, $\frac{\partial H_t}{\partial M_t}$ and $E_t$ is that more well educated people seek help from doctors more often, better follow doctors advice or purchase medicine in a more sophisticated manner. It has been found that the treatment gap for mental health problems is decreasing in education (Evans-Lacko et al., 2018), which supports that more well educated people seek help more. It is also possible to treat both mild, moderate and severe depression (World Health Organization, 2021), and thus it is plausible that an inability or reluctance to follow doctors advice would have an influence on the prevalence of depressive symptoms. This explanation could easily be extended to I will however not be trying to explain or disentangle the mechanism behind the observed causal relationship in this paper.

## Research Design

The causal relationship of interest is how health varies with education, as seen in equation 2:

$$H_{itk} = \beta_0 + \beta_1 X_{itk} + \beta_2 E_{itk} + u_{itk} \tag{2}$$

Where $H_{itk}$ is health measured in depressive symptoms, $X_{itk}$ is a vector of controls, $E_{itk}$ is the amount of education in years, $\mu_{itk}$ is an error term and subscripts $i, t, k$ denote person, time of interview and country, respectively. The parameter of interest is $\beta_2$, the change in depressive symptoms caused by an increase in education by one year.

The ideal experiment to identify $\beta_2$ would randomly allocate an amount of compulsory schooling to each person and prohibit further schooling. To approximate this randomization and deal with potential endogeneity problems, I use instrumental variables estimation. I estimate the local average treatment effect with two-stage least-squares (2SLS) (J. D. Angrist & Pischke, 2008) and conditional local average treatment effects using causal IV forests (Athey et al., 2019). The identifying assumptions (moment conditions) are the same, and I utilize the 2SLS setup to explain them.

I estimate the following first-stage:

$$E_{itk} = \phi_0 + \phi_1 X_{itk} + \phi_2 Z_{itk} + \epsilon_{itk} \tag{3}$$

Where $Z_{itk}$ is the amount of compulsory schooling in years and $\epsilon_{itk}$ is an error term. School reforms that change the amount of compulsory schooling induce exogenous variation in the amount of schooling completed, and is used as an instrument.

I make two main assumptions:

1. The amount of compulsory schooling only influences the amount of depressive symptoms through years of education acquired, i.e. $\text{Cov}(Z_{itk}, u_{itk}) = 0$, the so-called exclusion restriction,

2. That the first-stage exists, such that the amount of compulsory schooling influences the amount of education, i.e. $\text{Cov}(Z_{itk}, E_{itk}) \neq 0$

I can only test the second condition, which can be tested when estimating the first-stage regression. According to Stock, Wright, and Yogo (2002), F statistics larger than 10 is a good rule-of-thumb for a sufficiently strong first-stage, but the sign and size of the parameters should also be in accordance with the theoretical background and reasonable. This corresponds to a positive parameter in this setup. For the first condition to hold, Mazumder (2010) shows, in relation to mortality in the U.S., that it is important to control for country-specific trends. Otherwise the schooling reforms might correlate with historical improvements in nutrition, the exclusion restriction.

If there exists selection or heterogeneity, I am able identify a local average treatment effect (LATE) with 2SLS,[1] which is the effect of a one year increase in education on the people who were induced to increase their education by the treatment (compliers).

To examine the heterogeneity in the LATE, it is common to perform subgroup analysis, where the LATE is estimated for different subsamples of the population (Brunello, Fort, & Weber, 2009; Crespo et al., 2014; Mazzonna, 2014). I instead estimate a conditional local average treatment effect (CLATE) using Causal IV Forests, which is part of the Generalized Random Forest (see Athey et al. (2019) with an application of Causal IV Forests presented in section 7). The CLATE is the LATE given a specific set of covariates, and allows for counterfactual computation of the LATE for any given set of input covariates. All Causal IV Forests are computed using the *EconML* package from Microsoft Research with the recommended standard values (Battocchi et al., 2019)

[1] I assume monotonicity (J. Angrist & Imbens, 1995), i.e. that people are not induced to become less educated by having a higher compulsory school length.

## Data

The main data source used is the easySHARE dataset (Börsch-Supan & Gruber, 2020)[2], which is then expanded with childhood factors from SHARE wave 3 and 7 (Börsch-Supan, 2019, 2020).

The outcome of interest is the EURO-D score, which measures the amount of depressive symptoms in a patient, with values from 0 to 12 inclusive, a measure of depression which is validated across several European countries (Prince et al., 1999) (the variable *eurod* in easySHARE). It is this variable that decides what wave to use for each person, where the last available EURO-D score is used. This is such that I use the most recent data. If people with larger amounts of depressive symptoms are more likely not to be interviewed again (e.g. they have a higher mortality rate), this would cause selection issues. This would then leave the most depressed people in each wave, with the less depressed people being interviewed again later. To address this concern, I perform a robustness check with the first available EURO-D score.

The main dependent variable of interest is the years of education (*eduyears_mod* in easySHARE). People with no information, refusals to answer or not knowing as well as implausible and suspected wrong values are coded as missing. Furthermore, the covariates age, birthyear and sex of the respondent, as well as wave, also stems from the easySHARE data (the variables *gender* (recoded to a binary indicator for female), *byear*, *age* (rounded to nearest whole number) and *wave*).

The childhood factors included in SHARE wave 3 and 7 are very diverse. I follow Crespo et al. (2014) in what childhood factors to include, conditional on the covariates being available for both waves, which leaves the covariates regarding self-assessed math and language skills compared to peers, physical and mental health status, books, rooms and accommodations in home as well as the job occupation of the main breadwinner. The following list goes through the covariates interpretations, whereas the section *Coding of variables* in the appendix includes what variables in the dataset they are based on and transformations more in depth:

- Self-assessed math and language skills are rated on a scale from 1 to 5, with 1 being much worse relative to peers and 5 being much better.
- Physical health status as a child is rated on a scale from 1 to 5, with 1 being poor and 5 being excellent.

[2] See appendix section *Data source* for more information

- Mental health status as a child is a binary indicator indicating the presence of emotional, nervous, or psychiatric problems during childhood.
- Jobtype of the main breadwinner is a binary indicator indicating if the breadwinner has no urban or qualified job.
- Amount of accommodations in the home is a count of whether the home had 5 specific accommodations in the home.
- Amount of rooms in the home is the amount of rooms in the home.
- Amount of books in the home is a scale from 1 to 5 of the amount of books in the home, binned, with higher values corresponding to more books.
- Area of residence as a scale from 1 to 5, with higher values corresponding to more rural areas.

The summary statistics are seen in table 1. The sample consists of 6523 observations, of which 57% is female. The average person is born in 1947 and is 67 years of age at the time of interview. I observe that very few people in the sample experienced childhood mental health problems, and most people rated their childhood physical health as fair. People in the sample were of average socio-economic status, with 49% of the main breadwinners having no urban or qualified job and people on average being in the middle of the book and accommodations scales, with a tendency to live in more rural areas.

Table 1: Summary statistics

| | Mean | SD | Min | Median | Max | N |
|---|---|---|---|---|---|---|
| *General covariates* | | | | | | |
| Education | 11.0 | 3.8 | 0.0 | 11.0 | 18.0 | 6523 |
| EURO-D score | 2.2 | 2.1 | 0.0 | 2.0 | 12.0 | 6523 |
| Female | 0.57 | 0.50 | 0.00 | 1.00 | 1.00 | 6523 |
| Age | 67 | 5 | 49 | 68 | 87 | 6523 |
| Birthyear | 1947 | 5 | 1928 | 1947 | 1961 | 6523 |
| *Socio-economic covariates* | | | | | | |
| Books | 2.3 | 1.2 | 1.0 | 2.0 | 5.0 | 6523 |
| Rooms | 3.7 | 1.6 | 0.0 | 3.0 | 10.0 | 6523 |
| Area of residence | 3.7 | 1.5 | 1.0 | 4.0 | 5.0 | 6523 |
| Accommodations | 2.72 | 1.80 | 0.00 | 3.00 | 5.00 | 6523 |
| Breadwinner jobtype | 0.49 | 0.50 | 0.00 | 0.00 | 1.00 | 6523 |
| *Ability covariates* | | | | | | |
| Math skills | 3.3 | 0.9 | 1.0 | 3.0 | 5.0 | 6523 |
| Language skills | 3.4 | 0.9 | 1.0 | 3.0 | 5.0 | 6523 |
| *Health covariates* | | | | | | |
| Childhood physical health | 3.94 | 1.05 | 1.00 | 4.00 | 5.00 | 6523 |
| Childhood mental health | 0.02 | 0.13 | 0.00 | 0.00 | 1.00 | 6523 |

To identify which people were affected by the compulsory school reforms, I follow Crespo et al. (2014), who in turn follow Brunello et al. (2009), and utilize a window based method. Based on the first cohort affected by a compulsory school reform in a country, I include all people born up to seven years earlier than the first cohort, and people born in the first cohort affected or up to six years later.[3] To calculate the amount of compulsory schooling required I also follow Crespo

[3]I also refers to landers as countries.

et al. (2014). For each country and birth year they calculate the amount of compulsory schooling by subtracting the enrollment age from the minimum dropout age, which are tabulated in table 2.

To determine where people lived during their childhood I utilize variables that include information about residences that people have lived in throughout their life. People list up to 29 different residences in wave 3 and 30 different residences in wave 7. Original variable names are noted in italics in parenthesis, with wave 3 names mentioned before wave 7 names. The information includes both when people moved to a location (*sl_ac006_1* to *sl_ac006_29*, *ra006_1* to *ra006_30*), when they moved out (*sl_ac021_1* to *sl_ac021_28*, *ra021_1* to *ra021_30*) as well as the regions of residence (*sl_ac015c_1* to *sl_ac015c_29*, *ra021_1* to *ra021_30*). This allows me to relax the critical assumption that people who live in region $j$ at the time of interview also lived in region $j$ during their childhood, an assumptiom made by e.g. Crespo et al. (2014). This enhances the credibility of the identification strategy. I instead include people that have lived in the country 7 years before or after the compulsory school reform in the sample, but make no requirements for the length of stay, mirroring the reform window strategy.

On the basis of the people identified, I create a regression discontinuity plot with second order polynomials fitted on either side of the cutoff, which can be seen in figure 1, using *rdplot* (Calonico, Cattaneo, Farrell, & Titiunik, 2017). This is merely a graphical inspection, as I control for no other covariates and only plot the years of education of people against relative time to implementation of reform. No clear gap is seen, but a clear increasing trend in the years of education is seen, and it is important that I control for this later on.

To obtain continuous latent factors in regards to the socioeconomic status of a persons childhood home, I perform a principal component analysis (Jollife & Cadima, 2016) on the three variables regarding rooms, books and accommodations in childhood homes. This reconfiguration of the data is not necessary for inference, but it is easier to graphically inspect heterogeneity using continuous variables. Principal components are ordered after how much of the variation they explain, and as such the principal component of main interest is the first, but I include all principal components as I want to control for rooms, books and accommodations in full. As such, I merely utilize principal component analysis to reconfigure the three covariates in three new covariates with a linear function obtained through an eigenvector decomposition of the correlation matrix. This has the important consequence that, when utilizing OLS and 2SLS and not including interaction terms with either rooms, books and accommodations, the parameters on the remaining covariates do not change dependent upon whether the untransformed or transformed

Table 2: Compulsory schooling reforms, recreated verbatim from Crespo et al. (2014)

| Country | Source | Year of reform | FCA[a] | Left window | Right window | YCE[b] before | YCE[b] after |
|---|---|---|---|---|---|---|---|
| Austria | Brunello et al. (2009) | 1962 | 1947 | 1940 | 1953 | 8 | 9 |
| Germany1[c] | Brunello et al. (2009) | 1956 | 1941 | 1934 | 1947 | 8 | 9 |
| Germany2[c] | Brunello et al. (2009) | 1949 | 1934 | 1927 | 1940 | 8 | 9 |
| Germany3[c] | Brunello et al. (2009) | 1962 | 1947 | 1940 | 1953 | 8 | 9 |
| Germany4[c] | Brunello et al. (2009) | 1958 | 1943 | 1936 | 1949 | 8 | 9 |
| Germany5[c] | Brunello et al. (2009) | 1967 | 1953 | 1946 | 1959 | 8 | 9 |
| Germany6[c] | Brunello et al. (2009) | 1967 | 1953 | 1946 | 1959 | 8 | 9 |
| Germany7[c] | Brunello et al. (2009) | 1967 | 1953 | 1946 | 1959 | 8 | 9 |
| Germany8[c] | Brunello et al. (2009) | 1967 | 1953 | 1946 | 1959 | 8 | 9 |
| Germany9[c] | Brunello et al. (2009) | 1969 | 1955 | 1948 | 1961 | 8 | 9 |
| Germany10[c] | Brunello et al. (2009) | 1964 | 1949 | 1942 | 1955 | 8 | 9 |
| Sweden | Brunello et al. (2009) | 1962 | 1950 | 1943 | 1956 | 8 | 9 |
| Netherlands | Murtin and Viarengo (2008) | 1950 | 1937 | 1930 | 1943 | 6 | 8 |
| Italy | Murtin and Viarengo (2008) | 1963 | 1949 | 1942 | 1955 | 5 | 8 |
| France | Brunello et al. (2009) | 1967 | 1953 | 1946 | 1959 | 8 | 10 |
| Denmark[d] | Arendt (2005) | 1958 | 1945 | 1938 | 1951 | 5 | 7 |

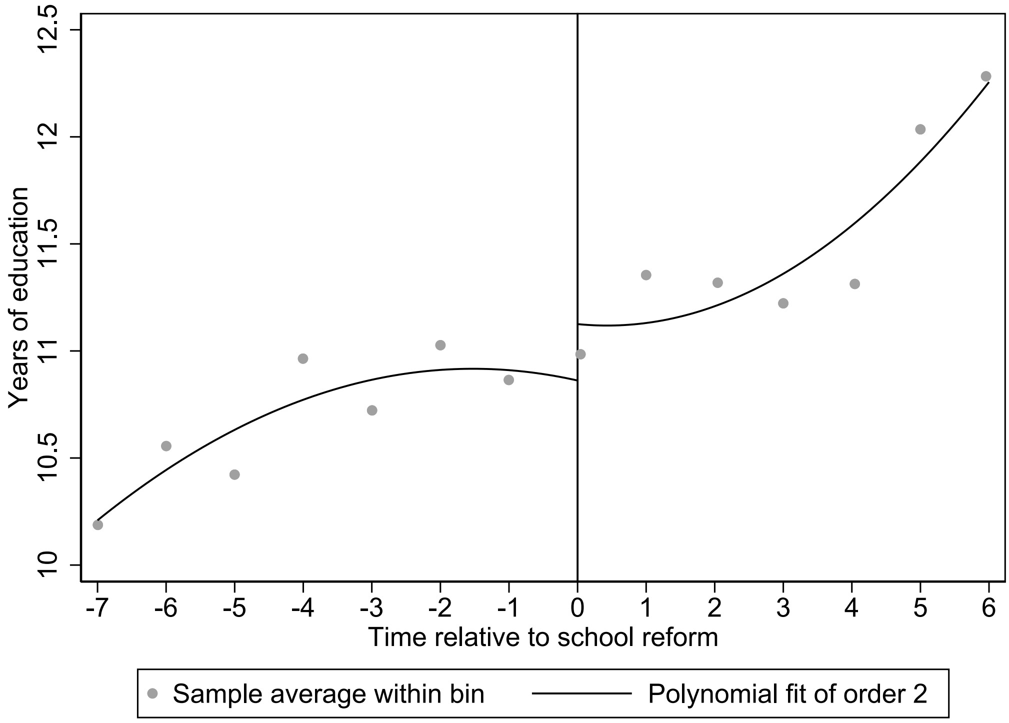
[a] First cohort affected.

[b] Years of compulsory education

[c] Germany1:Schleswig-Holstein; Germany2:Hamburg; Germany3:Niedersachsen; Germany4:Bremen; Germany5:Nordrhein-Westphalia; Germany6:Hessen; Germany7:Rheinland-Pfalz; Germany8:Baden-Wurtemberg; Germany9:Bayern; Germany10:Saarland.

[d] Danish 1958 reform is more complicated than others, but to start with I interpreted it as an increase by two years of compulsory schooling. See Arendt (2005).

Figure 1: Regression discontinuity plot




variables are included. This, however, is not the case for causal IV forests.

Table 3: Principal component variable loadings and explained variance

| | Principal component 1 | Principal component 2 | Principal component 3 |
|---|---|---|---|
| Books | 0.62 | -0.31 | -0.72 |
| Rooms | 0.49 | 0.87 | 0.04 |
| Accommodations | 0.61 | -0.38 | 0.69 |
| Explained variance | 0.57 | 0.26 | 0.17 |

I report the loadings (how much a principal component changes with a one standard deviation increase in a given input variable) and the explained variance of each principal component in table 3. I see that the first principal component loads positively on both books, rooms and accommodations, with loadings of 0.62, 0.49 and 0.61 respectively, and that it explains 57% of the variation in the input data. Given that we believe that increases in books, rooms or accommodations corresponds to higher socioeconomic status, this allows me to interpret increases in the first principal component as having a higher socioeconomic background.

# Results

This section is split into three parts: The first part covers the first stage regression, the second part covers the results and the third part examines the robustness of the results.

## First stage regression

Figure 1 established that there exists no clear discontinuity when not controlling for other covariates. Following Mazumder (2010), I utilize four different ways of controlling for unobserved confounding at the cohort and country level. Model 1 controls for cohort-country specific fixed effects, model 2 controls for age cubic and cohort-country specific fixed effects, model 3 controls for wave specific age cubic and cohort-country specific fixed effects and model 4 controls for wave specific age cubic and country specific cohort trends. As noted by Glymour, Kawachi, Jencks, and Berkman (2008), structural confounders seem to most plausibly violate the exclusion restriction, whereas innate characteristics, such as the childhood covariates, would require selective immigration based on changes in the compulsory school reforms, which is why I focus on how I control for the structural confounders.

I use additional instruments in addition to compulsory schooling length, as the first stages are relatively weak when only utilizing the compulsory school length, shown in the appendix, figure 7. The specific instruments selected were based on a first stage equation with all instruments included (interaction of compulsory schooling length with all untransformed childhood covariates conditional on the covariate also being included in the first and second stage)[4] (Crespo et al., 2014), shown in the appendix, figure 8, where the interaction with math skills and area of residence consistently stand out. I also estimate a model with all instruments as a robustness check.

Table 4 displays the first stage for the four different types of controls and the selected instruments (compulsory schooling length and interactions with area of residence and math skills relative to peers). I see that coefficient on compulsory schooling does indeed change dependent upon how one controls for country, age and cohort. The smallest coefficient on compulsory school appears in model 4 (disregarding interaction effects, which are relatively constant across specifications), with an extra year of compulsory schooling inducing people to take 0.17 years of extra schooling.

[4]Technically, the transformed covariates can also be utilized as instruments by interacting them with the compulsory schooling length. I refrain from doing this due to two reasons: The interpretations of the interactions would be more difficult than with untransformed covariates, and 2) it would change the parameter estimates whether transformed and untransformed covariates are used if they enter in an interaction.

The largest F statistic also appears in model 4, with an F statistic of 24.32, which is well above the rule of thumb of 10 or higher (J. D. Angrist & Pischke, 2008; Stock et al., 2002). Given the visual trend in figure 1 and the risk of catching other societal trends with the compulsory schooling reforms (Mazumder, 2010) and having the largest F statistic, I prefer model 4 which controls for country-specific trends and cubic age effects that depend on the year of interview.

To evaluate how reasonable the first stage is, I plot the average marginal effect of an extra year of compulsory schooling at different covariate levels, seen in figure 2. I note that the average marginal effect never is significantly below zero, and thus an extra year of compulsory schooling never results in a significant decrease in years of education, which seems reasonable. The effect of compulsory schooling is also increasing in rurality and how good one is at math relative to ones peers, which I also deem reasonable. Compulsory schooling reforms being more binding in rural areas (a simple univariate regression of area of residence on years of education results in a coefficient of -0.47) could explain the first relationship, whereas higher returns to schooling for more abled people, inducing a higher incentive to increase schooling, could explain the second relationship.

Although not of primary interest in the first stage equation, I also note that the first principal component is associated with an increased amount of schooling in all specifications. This is what I would expect a factor that captures how socioeconomically well off a family is, which reinforces the interpretation of the first principal component.

Figure 2: Average marginal effect of compulsory schooling with 95% CIs

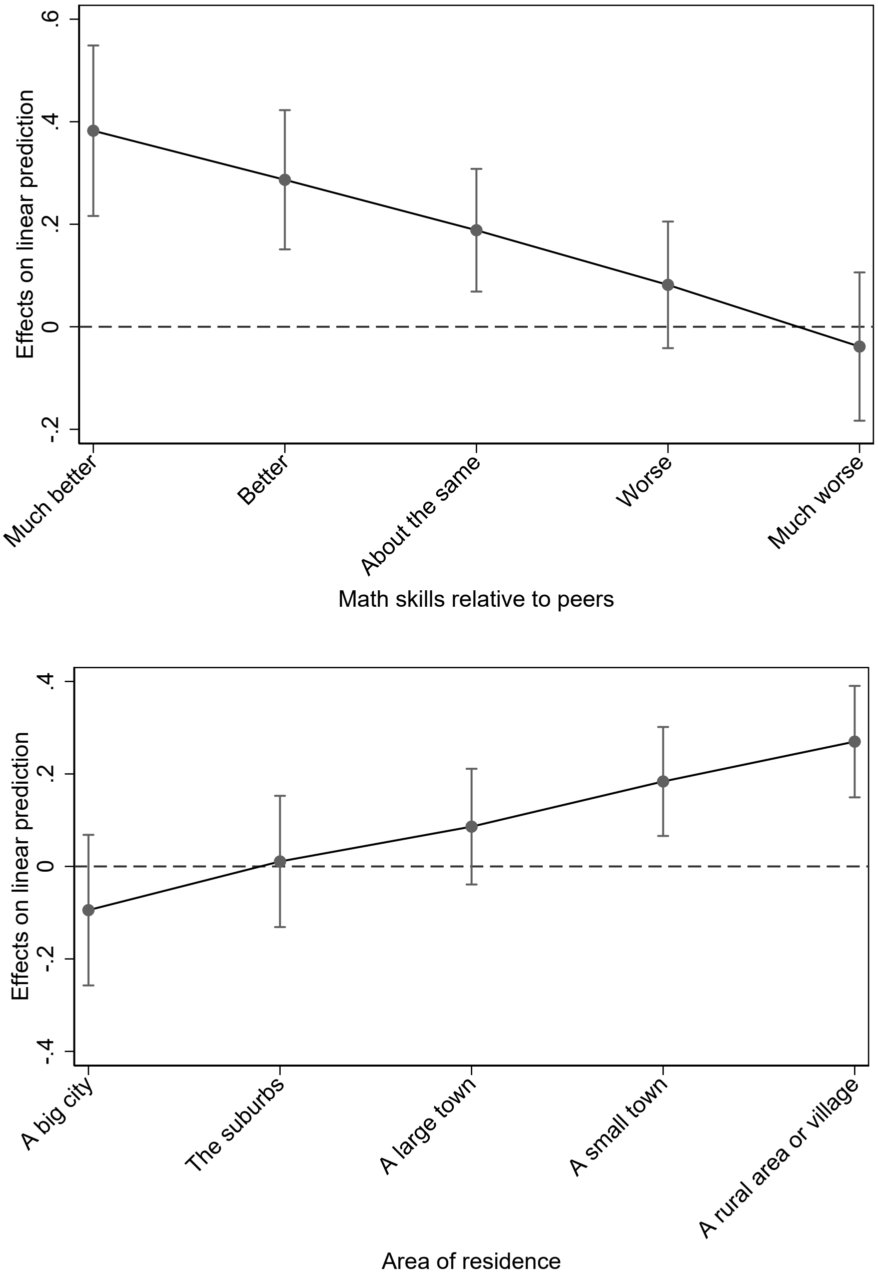


Table 4: First stage regressions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Compulsory school | 0.203 | 0.345 | 0.791*** | 0.166 |
| | (0.206) | (0.217) | (0.273) | (0.139) |
| Female | -0.449*** | -0.456*** | -0.459*** | -0.483*** |
| | (0.0930) | (0.0932) | (0.0936) | (0.0910) |
| First PC | 0.879*** | 0.878*** | 0.878*** | 0.875*** |
| | (0.0409) | (0.0411) | (0.0418) | (0.0398) |
| Second PC | -0.423*** | -0.422*** | -0.424*** | -0.427*** |
| | (0.0543) | (0.0543) | (0.0548) | (0.0523) |
| Third PC | -0.244*** | -0.238*** | -0.238*** | -0.234*** |
| | (0.0519) | (0.0520) | (0.0518) | (0.0509) |
| Math skills | 1.195*** | 1.193*** | 1.206*** | 1.242*** |
| | (0.198) | (0.198) | (0.193) | (0.186) |
| Language skills | 0.618*** | 0.620*** | 0.624*** | 0.616*** |
| | (0.0634) | (0.0631) | (0.0633) | (0.0629) |
| Breadwinner jobtype | -0.320*** | -0.319*** | -0.315*** | -0.292*** |
| | (0.0794) | (0.0794) | (0.0795) | (0.0769) |
| Childhood physical health | 0.0768* | 0.0765* | 0.0714* | 0.0664 |
| | (0.0429) | (0.0429) | (0.0430) | (0.0427) |
| Childhood mental health | 0.384 | 0.377 | 0.388 | 0.376 |
| | (0.292) | (0.292) | (0.292) | (0.280) |
| Area of residence | -0.733*** | -0.738*** | -0.745*** | -0.795*** |
| | (0.144) | (0.144) | (0.141) | (0.138) |
| Compulsory school × Math skills | -0.0987*** | -0.0989*** | -0.101*** | -0.105*** |
| | (0.0265) | (0.0266) | (0.0261) | (0.0249) |
| Compulsory school × Area of residence | 0.0816*** | 0.0820*** | 0.0834*** | 0.0895*** |
| | (0.0184) | (0.0183) | (0.0181) | (0.0178) |
| Observations | 6523 | 6523 | 6523 | 6523 |
| F statistic | 16.22 | 7.239 | 20.49 | 24.32 |
| Country × cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic × wave | | | YES | YES |
| Country × cohort trend | | | | YES |

Standard errors in $(\cdot)$, clustered on birth cohort and country

$^{*}$ = p < 0.05, $^{**}$ = p < 0.01, $^{***}$ = p < 0.001

### Estimated treatment effects

I first present the results from OLS and 2SLS, which can be seen in table 5. We see that the OLS estimates are closer to 0 than either of the 2SLS estimates in all specifications. As 2SLS is biased towards OLS if the first stage is weak (J. D. Angrist & Pischke, 2008), the estimated LATEs are conservative rather than liberal, although the first stage is not especially weak.

The LATE varies from negligible to substantive dependent upon how I control for structural confounders. In the preferred specification, the estimates are -0.11 and -0.14 for three and all instruments, respectively, but neither is statistically significant at a 5% significance level. This corresponds to a decrease of 0.11 depressive symptoms per extra year of education for people who were induced to take an extra year of education by the schooling reforms (compliers). Compared to the mean amount of depressive symptoms, 2.2, this corresponds to a decrease in depressive symptoms by approximately 5%.

Table 5: Estimated models

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| OLS | -0.0193** | -0.0189** | -0.0181** | -0.0193** |
| | (0.00874) | (0.00874) | (0.00870) | (0.00856) |
| IV, three inst. | -0.0220 | -0.0235 | -0.0251 | -0.110 |
| | (0.126) | (0.125) | (0.121) | (0.114) |
| IV, all inst. | -0.0701 | -0.0699 | -0.0743 | -0.137 |
| | (0.115) | (0.114) | (0.112) | (0.105) |
| Observations | 6523 | 6523 | 6523 | 6523 |
| Country × cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic × wave | | | YES | YES |
| Country × cohort trend | | | | YES |

Standard errors in $(\cdot)$, clustered on birth cohort and country

$^{*} = p < 0.05$, $^{**} = p < 0.01$, $^{***} = p < 0.001$

To increase clarity, I report a smoothed version of all the CLATE graphs estimated using causal IV forests, by creating a moving average of the CLATE and confidence intervals with 100 counterfactual first principal components (the principal component increases by 0.1 in this window). The unsmoothed version of figure 4 is included in the appendix, figure 11, to exemplify how the graphs look before smoothing. As the CLATE is dependent on all input covariates, I must decide both on what covariates to vary and at what values the other covariates should be held constant at. As a rule, I evaluate at the mean, median or most common value of covariates where there is no middle category, but if a middle category exists, I utilize that.[5]

Figure 4 shows the estimated CLATE as a function of socioeconomic status and gender, which is are my main covariates of interest. In concurrence with the findings from Crespo et al. (2014), the absolute size of estimated CLATE in the lower part of the socioeconomic part is higher than in the upper part, and most of this increased effect is driven by women, with an estimated CLATE of about -0.25 at the lowest end of the socioeconomic scale, linearly decreasing until the middle of the scale, and being constant at about -0.1, whereas men only have slightly increased absolute effects in the lower end of the socioeconomic scale. This means that women in the lower end of the socio-economic scale that were induced to take another year of education by the reform experience a reduction in depressive symptoms by 0.25. I note, however, that the estimated CLATEs are almost always not statistically different from zero, with very wide confidence intervals that overlap to a large degree.

I report the same figures with the country of origin changed in figures 12 to 24 in the appendix, and the pattern is robust across countries, except for France, where females generally have lower returns to education than men, and the estimated effects are smallest in the lower and upper ends of the socioeconomic scale, and largest in the middle, see figure 13.

[5] Based on the summary statistics in table 1 and the sample by country table, table 9 in the appendix, this leads me to estimate the CLATE for a 68 year old male, interviewed in wave 6, thus born in year 1947, from Sweden, living in a large town, with no mental health problems and fair physical health during childhood, with the main breadwinner having an urban or qualified job, math and language skills about the same as peers and with principal component 2 and 3 fixed at zero, unless otherwise noted.

Figure 4: CLATE heterogeneity by gender

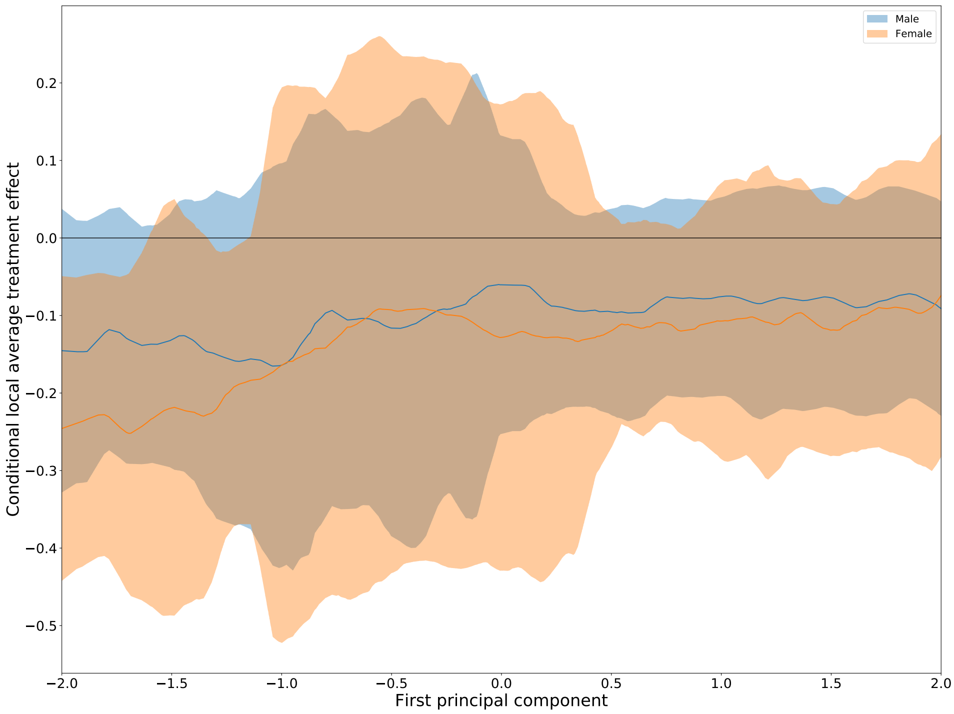
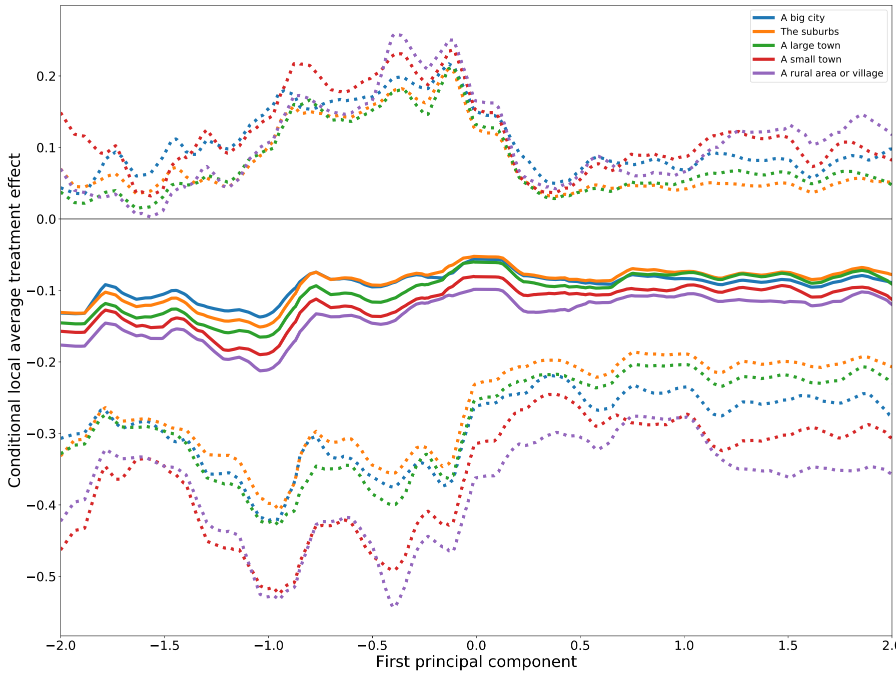
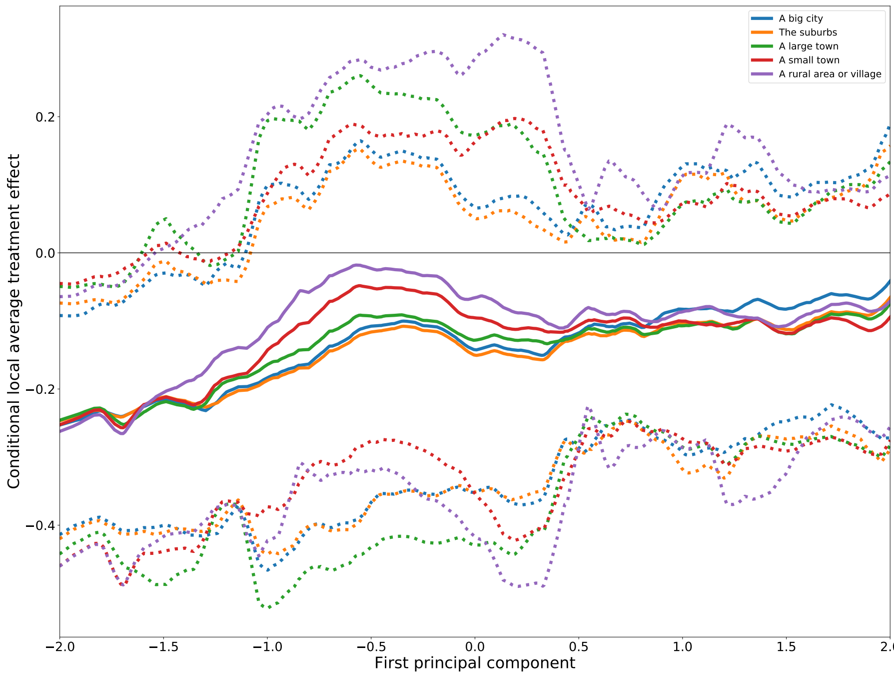



Figure 5 shows the estimated CLATE as a function of the first principal component and area of residence, for both men and women. For men, the difference in CLATE between areas of residence is roughly equal for all socioeconomic status', with the absolute CLATE increasing in rurality and the difference in general being a reduction of depressive symptoms by 0.05 from the most rural to most urban. The opposite is true for women, where the absolute CLATE is decreasing in rurality and not constant, no differences found in the lowest and highest parts of the socioeconomic spectrum, but a large difference of up to 0.1 based on rurality in the socioeconomic range of -1 to 0.

Figure 5: CLATE heterogeneity by area of residence



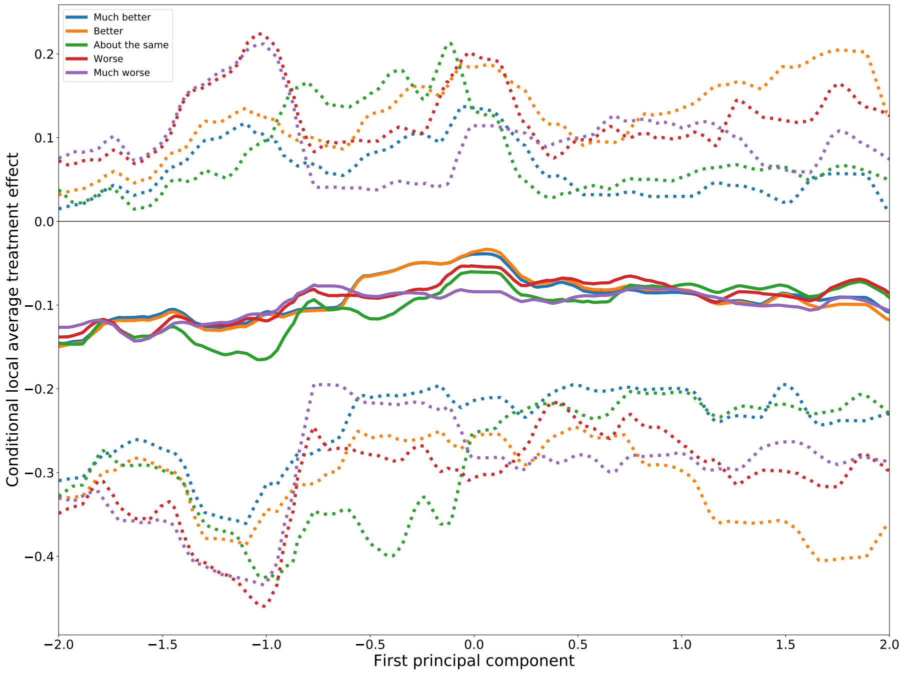

(a) Male



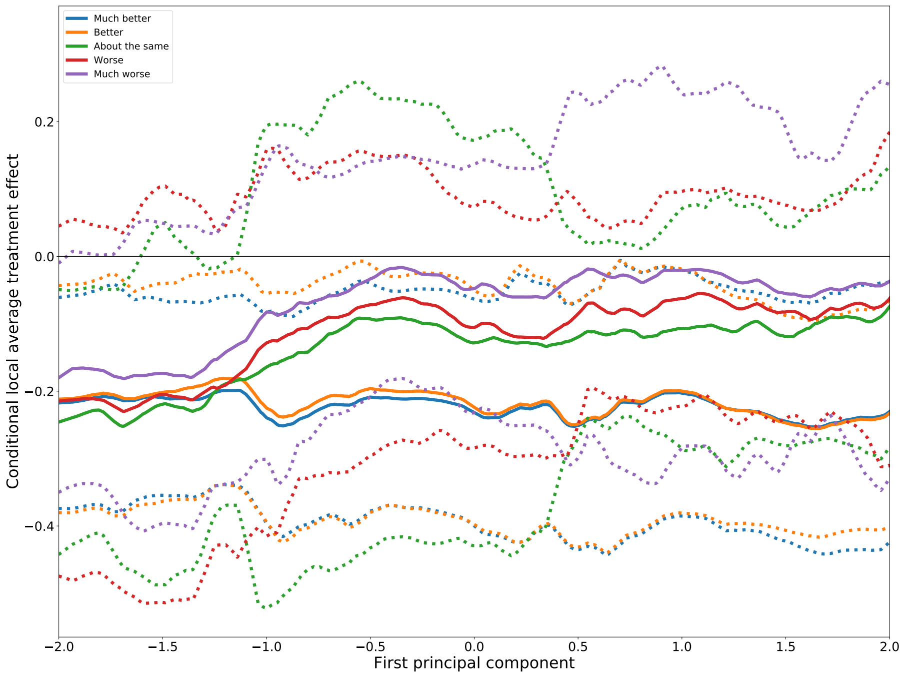

(b) Female

Figure 7 shows the estimated CLATE as a function of the first principal component and math and language skills relative to peers. The estimated differences between males with different levels of ability do not differ much, and do not vary a lot with socioeconomic status. This is not the case for women, where women that are much better or better than their peers have estimated CLATEs that are significantly below zero for all socioeconomic statuses and do not vary much with it, with all estimates being around -0.2, such that the effect of one additional year of schooling on depression scores is about -0.25 for women from low SES families that were affected by a schooling reform.

Figure 7: CLATE heterogeneity by ability relative to peers




(a) Male

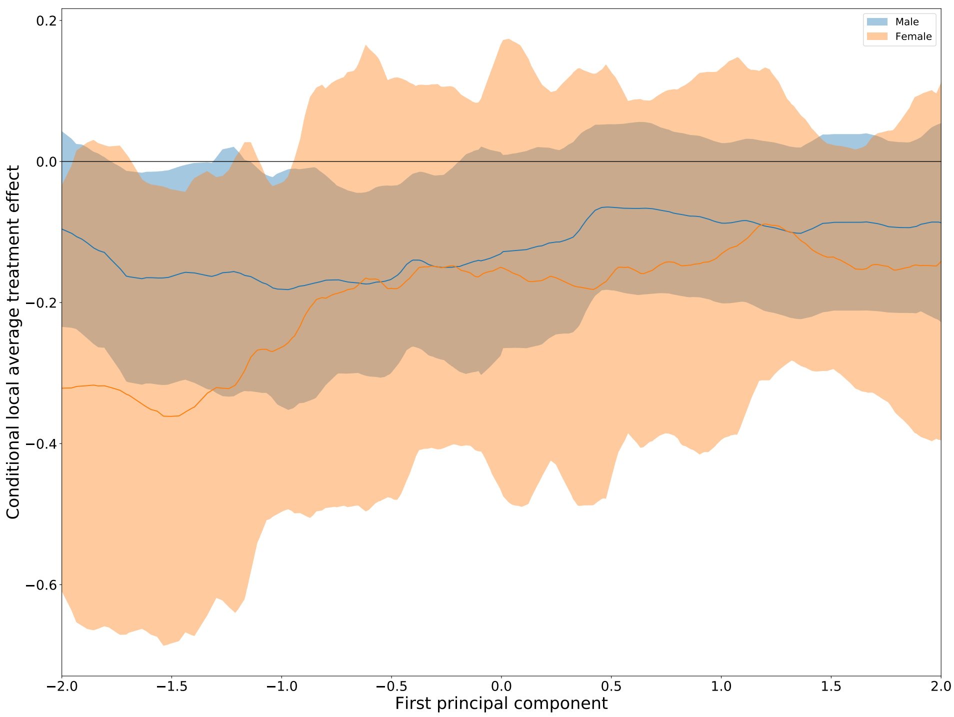



(b) Female

### Robustness checks

To assess how sensitive the estimated CLATEs were to my selection of instruments, I also include a graph of the estimated CLATEs had I utilized all instruments, which can be seen in figure 9 (in addition to the estimated LATE in table 5). The relationship is unchanged, but the estimated absolute CLATEs for women in the lower half of the socioeconomic spectrum are larger, with reductions in depressive symptoms of up to 0.3 for the lowest part of the socioeconomic spectrum. However, as seen in table 8 in the appendix, the first stage is much weaker than with the three selected instruments (F statistic of 12.16 compared to 24.32) and thus I prefer the model with three instruments.

Figure 9: Heterogeneity by gender using all instruments




To assess robustness of the arbitrary 7-year cutoff utilized when I calculate what people to include in the sample, both in regards to who lived in the country, but also who was and was not affected by the school reform, I estimate models with a 5 and 9 year cutoff. The estimated LATEs and CLATEs can be seen in the appendix, in tables 11 & 12 and figures 27 & 28. Focusing on the preferred models the estimated LATEs are decreasing for larger cutoffs. The estimated CLATEs are relatively unchanged, with the 5 year cutoff resulting in a larger difference between men and women, primarily driven by a reduction in the estimated absolute CLATEs for men.

To assess the robustness of the model to the decision to use the most recent observation, I estimate models that use the first observation instead. The estimated LATE can be seen in table 6 and the estimated CLATEs can be seen in figure 10. The estimated LATEs are similar to previously estimated LATEs, but the heterogeneity observed along the socioeconomic status is changed, with women from a higher socioeconomic status having higher estimated absolute CLATEs, contrary to previous findings, but the difference between men and women persists.

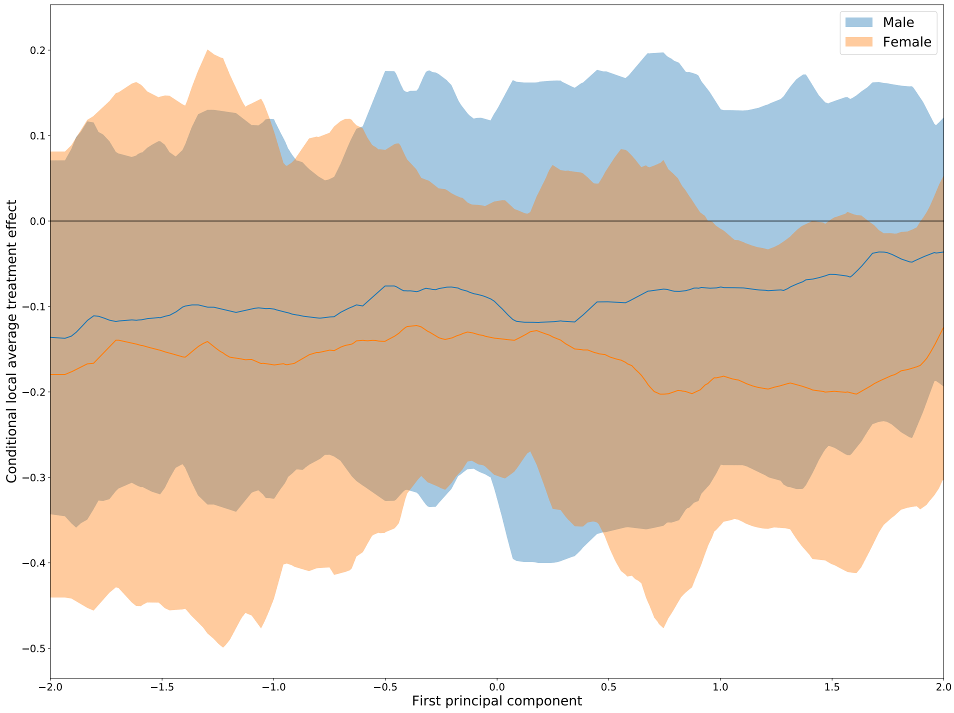
Table 6: Estimated models with earliest observation

| | OLS | IV, three inst. | IV, all inst. |
|---|---|---|---|
| Years of education | -0.0263*** | -0.108 | -0.102 |
| | (0.00954) | (0.0941) | (0.0885) |
| Observations | 6523 | 6523 | 6523 |
| F statistic | | 24.21 | 4.690 |
| Country × cohort | | | |
| Age cubic | | | |
| Age cubic × wave | YES | YES | YES |
| Country × cohort trend | YES | YES | YES |

Standard errors in (·), clustered on birth cohort and country

$^{*}$ = p < 0.05, $^{**}$ = p < 0.01, $^{***}$ = p < 0.001

Figure 10: Heterogeneity by gender with earliest observation




## Conclusion

I find results that are of a meaningful and substantive size, with estimated local average treatment effects in the range of a reduction of depressive symptoms by 0 to 0.3 for each extra year of school induced by the reforms, dependent on model and covariates. This indicates that more education causes a reduction in depressive symptoms, but the estimated effects are never statistically significant at a 5% significance level. The estimated local average treatment effect of -0.11 corresponds to a decrease in depressive symptoms by 5% compared to the mean prevalence. The model finds a lot of heterogeneity in the estimated effects based on childhood conditions and gender. My most robust finding is that women generally have higher returns to education as measured by reduction in depressive symptoms. The results along the socioeconomic scale are not very robust, with some models estimating higher returns to education in the lower end of the socioeconomic spectrum, but some reversing the trend with higher returns in the upper end of the socioeconomic spectrum dependent upon how I construct the data.

The estimated effects generally indicate that there exists some interplay between gender, socioeconomic status and the returns to education, but due to wide confidence intervals, it is hard to say exactly what the interplay is, and further research is needed before conclusions can be made. The large amounts of heterogeneity and the wide confidence intervals also means that one should be wary of generalizing findings from other studies with different samples, settings and outcomes, as the heterogeneity might cause one to draw wrong conclusions. Further research might benefit from examining more specific mechanisms and settings (for i.e. a specific subgroup of people or focusing on a single country with larger sample sizes), compared to the rather broad and cross-country setting of this study.

## References


Angrist, J., & Imbens, G. (1995). *Identification and estimation of local average treatment effects.* National Bureau of Economic Research Cambridge, Mass., USA.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics.* Princeton university press.

Arendt, J. N. (2005, 4). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, *24*(2), 149–160. doi: 10.1016/j.econedurev.2004.04.008

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1179–1203. doi: 10.1214/18-AOS1709

Battocchi, K., Dillon, E., Syrgkanis, V., Hei, M., Oprescu, M., Lewis, G., & Oka, P. (2019). *EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation.* https://github.com/microsoft/EconML.

Bhattacharya, J., Hyde, T., & Tu, P. (2013). *Health economics.* Macmillan International Higher Education.

Börsch-Supan, A. (2019). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 – SHARELIFE. Release version: 7.1.0. SHARE-ERIC. Data set.* doi: 10.6103/SHARE.w3.710

Börsch-Supan, A. (2020). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.1.1. SHARE-ERIC. Data set.* doi: 10.6103/SHARE.w7.711

Börsch-Supan, A., & Gruber, S. (2020). *easySHARE. Release version: 7.1.0. SHARE-ERIC. Dataset.* doi: 10.6103/SHARE.easy.710

Brunello, G., Fort, M., & Weber, G. (2009). Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe. *The Economic Journal*, *119*(536), 516–539. Retrieved from `https://doi.org/10.1111/j.1468-0297.2008.02244.x` doi: 10.1111/j.1468-0297.2008.02244.x

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404.

Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, *43*, 36–46. doi: 10.1016/j.econedurev.2014.09.003

Cutler, D. M., & Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. *National bureau of economic research*. doi: 10.1590/s0100-40422000000600014

Evans-Lacko, S., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bruffaerts, R., ... Wojtyniak, B. (2018, 7). Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: Results from the WHO World Mental Health (WMH) surveys. *Psychological Medicine*, *48*(9), 1560–1571. doi: 10.1017/S0033291717003336

Glymour, M. M., Kawachi, I., Jencks, C. S., & Berkman, L. F. (2008, 6). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*, *62*(6), 532–537. doi: 10.1136/jech.2006.059469

Gruber, S., Hunkler, C., Stuck, S., Bohacek, R., Christelis, D., Cavapozzi, D., ... Wahrendorf, M. (2014). *Generating easySHARE Guidelines, Structure, Content and Programming* (Tech. Rep.). Retrieved from `www.share-project.org`

Jann, B. (2019, 7). *ISCOGEN: Stata module to translate ISCO codes.* Statistical Software Components, Boston College Department of Economics. Retrieved from `https://ideas.repec.org/c/boc/bocode/s458665.html`

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). doi: 10.1098/rsta.2015.0202

Lleras-Muney, A. (2002, 6). *The Relationship Between Education and Adult Mortality in the United States* (Tech. Rep.). Cambridge, MA: National Bureau of Economic Research. Retrieved from `http://www.nber.org/papers/w8986.pdf` doi: 10.3386/w8986

Mazumder, B. (2010). *Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws* (Tech. Rep.). Federal Reserve Bank of Chicago.

Mazzonna, F. (2014, 1). The long lasting effects of education on old age health: Evidence of gender differences. *Social Science and Medicine*, *101*, 129–138. doi: 10.1016/j.socscimed.2013.10.042

Murtin, F., & Viarengo, M. (2008). *The convergence of compulsory schooling in Western Europe: 1950-2000* (Tech. Rep.). London: Centre for the Economics of Education.

Prince, M. J., Reischies, F., Beekman, A. T., Fuhrer, R., Jonker, C., Kivela, S. L., ... Copeland, J. R. (1999). Development of the EURO-D scale - A European Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*, *174*(APR.), 330–338. doi: 10.1192/bjp.174.4.330

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529.

World Health Organization. (2011). *Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level* (Tech. Rep.).

World Health Organization. (2021, 9). *Depression fact sheet.* Retrieved from `https://www.who.int/news-room/fact-sheets/detail/depression`

## Appendix

### Data source

This paper uses data from the generated easySHARE data set (DOI: 10.6103/SHARE.easy.710), see Gruber et al. (2014) for methodological details. The easySHARE release 7.1.0 is based on SHARE Waves 1, 2, 3 (SHARELIFE), 4, 5, 6 and 7 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, 10.6103/SHARE.w7.710).

### Coding of variables

Original variable names are noted in italics in parenthesis, with wave 3 names mentioned before wave 7 names. If a person has answered in both waves (very few have), the most recent values from wave 7 are used. I stress that the scales I list are the final scales, and not necessarily how the variables appear in the SHARE data.

The self-assessed math (*sl_cs010_*, *cc010_*) and language skills (*sl_cs010a_*, *cc010a_*) are rated on a scale from 1 to 5, with 1 being much worse relative to peers and 5 being much better. I code refusals to answer and not knowing as missing.

The physical health status as a child (*sl_hs003_*, *hs003_*) is rated on a scale from 1 to 5, with 1 being poor and 5 being excellent. I code refusals to answer, not knowing as well as the spontaneous answer 'health varied a great deal' as missing.

The mental health status as a child (*sl_hs009d3*, *hs009d3*) is a binary indicator indicating the presence of emotional, nervous, or psychiatric problems during childhood. I code refusals to answer and not knowing as missing.

The jobtype of the main breadwinner (*sl_cs009_*, *cc009isco*) is an indicator variable based on the

ISCO-08 major (1-digit) classifications. *cc009isco* is still in raw ISCO-08 codes and I transform it to the major ISCO groups using the package *iscogen* (Jann, 2019). I follow Crespo et al. (2014) in my coding of breadwinner jobtype. As such, 1 corresponds to a breadwinner in one of the following occupations: Craft or related trades worker; a plant/machine operator or assembler; elementary occupation; armed forces. This corresponds to a breadwinner with no urban or qualified job. I code refusals to answer, not knowing and the spontaneous response 'there was no main breadwinner' as missing.

The amount of accommodations in the home (*sl_cs007d1* to *sl_cs007d5*, *cc007d1* to *cc007d5*) is a count of whether the home had 5 specific accommodations. The accommodations listed are access to a bath, cold water, hot water, inside toilet and central heating. I code refusals and not knowing as not having.

The amount of rooms in the home (*sl_cs002_*, *cc002_*) is the amount of rooms in the home when ten. As some people report having up to 50 rooms, I elect to limit the sample to people with 15 or less rooms in the home. I code refusals to answer and not knowing as missing.

The amount of books in the home (*sl_cs008_*, *cc008_*) is a scale from 1 to 5 of the amount of books in the home when ten in bins, with higher values corresponding to more books. The bins used are none or very few and enough to fill one shelf, one bookcase, two bookcases and two or more bookcases. I code refusals to answer and not knowing as missing.

The area of residence (*sl_ac017_1*, *ra017_1*) is a scale from 1 to 5, with higher values corresponding to more rural areas. I code refusals to answer and not knowing as missing.

## Tables and figures

Table 7: First stage equation with single instrument

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Compulsory school | 0.411** | 0.550** | 0.969*** | 0.182*** |
| | (0.200) | (0.214) | (0.276) | (0.0640) |
| Observations | 6523 | 6523 | 6523 | 6523 |
| F statistic | 4.205 | 6.603 | 12.32 | 8.051 |
| Country $\times$ cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic $\times$ wave | | | YES | YES |
| Country $\times$ cohort trend | | | | YES |

Standard errors in $(\cdot)$, clustered on birth cohort and country

$^{*}$ = $p < 0.05$, $^{**}$= $p < 0.01$, $^{***}$ = $p < 0.001$

Table 8: First stage equation with all possible instrument interactions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Compulsory school | 0.165 | 0.312 | 0.762** | 0.174 |
| | (0.301) | (0.299) | (0.337) | (0.209) |
| Compulsory school × Breadwinner jobtype | 0.0788 | 0.0802 | 0.0745 | 0.0637 |
| | (0.0489) | (0.0488) | (0.0495) | (0.0466) |
| Compulsory school × Language skills | -0.0410 | -0.0422 | -0.0435 | -0.0587 |
| | (0.0364) | (0.0362) | (0.0361) | (0.0368) |
| Compulsory school × Math skills | -0.0816*** | -0.0812*** | -0.0833*** | -0.0824*** |
| | (0.0269) | (0.0269) | (0.0266) | (0.0257) |
| Compulsory school × Childhood mental health | -0.425* | -0.415* | -0.422* | -0.398* |
| | (0.233) | (0.234) | (0.233) | (0.236) |
| Compulsory school × Childhood physical health | 0.0110 | 0.0106 | 0.0130 | 0.0197 |
| | (0.0303) | (0.0303) | (0.0302) | (0.0301) |
| Compulsory school × Area of residence | 0.0821*** | 0.0824*** | 0.0837*** | 0.0889*** |
| | (0.0183) | (0.0182) | (0.0180) | (0.0175) |
| Observations | 6523 | 6523 | 6523 | 6523 |
| F statistic | 8.923 | 10.21 | 10.87 | 12.16 |
| Country × cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic × wave | | | YES | YES |
| Country × cohort trend | | | | YES |

Standard errors in (·), clustered on birth cohort and country

* = $p < 0.05$, **= $p < 0.01$, *** = $p < 0.001$

Table 9: Summary statistics

| | Percent |
|---|---|
| *Not Germany* | |
| Austria | 0.132 |
| Denmark | 0.146 |
| France | 0.170 |
| Italy | 0.195 |
| Netherlands | 0.056 |
| Sweden | 0.197 |
| *Germany* | |
| Schleswig-Holstein | 0.005 |
| Hamburg | 0.002 |
| Niedersachsen | 0.015 |
| Bremen | 0.001 |
| Nordrhein-Westphalia | 0.028 |
| Hessen | 0.008 |
| Rheinland-Pfalz | 0.004 |
| Baden-Würtemberg | 0.017 |
| Bayern | 0.023 |
| Saarland | 0.001 |

Figure 11: Heterogeneity by gender, unsmoothed

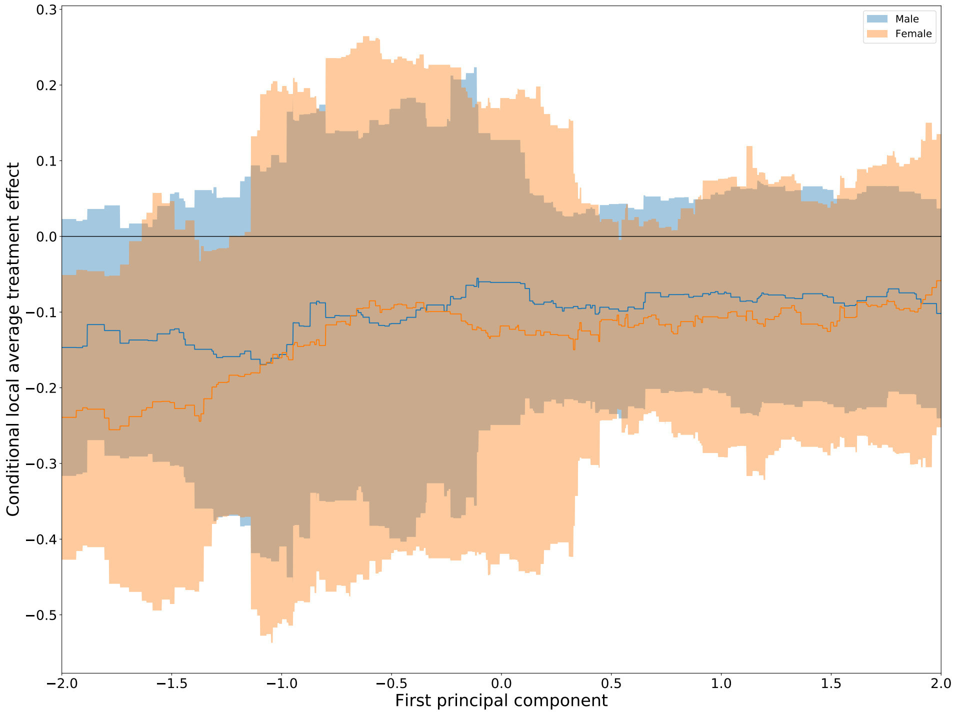



Figure 25: Heterogeneity by gender with a 5 year cutoff

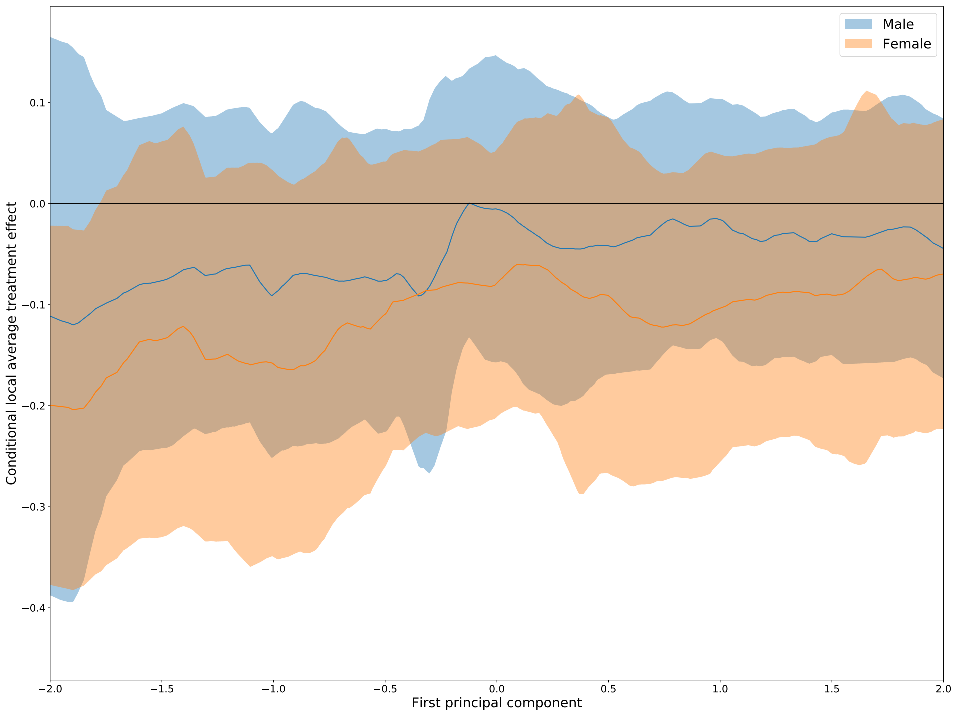


Figure 12: Heterogeneity by gender, Denmark

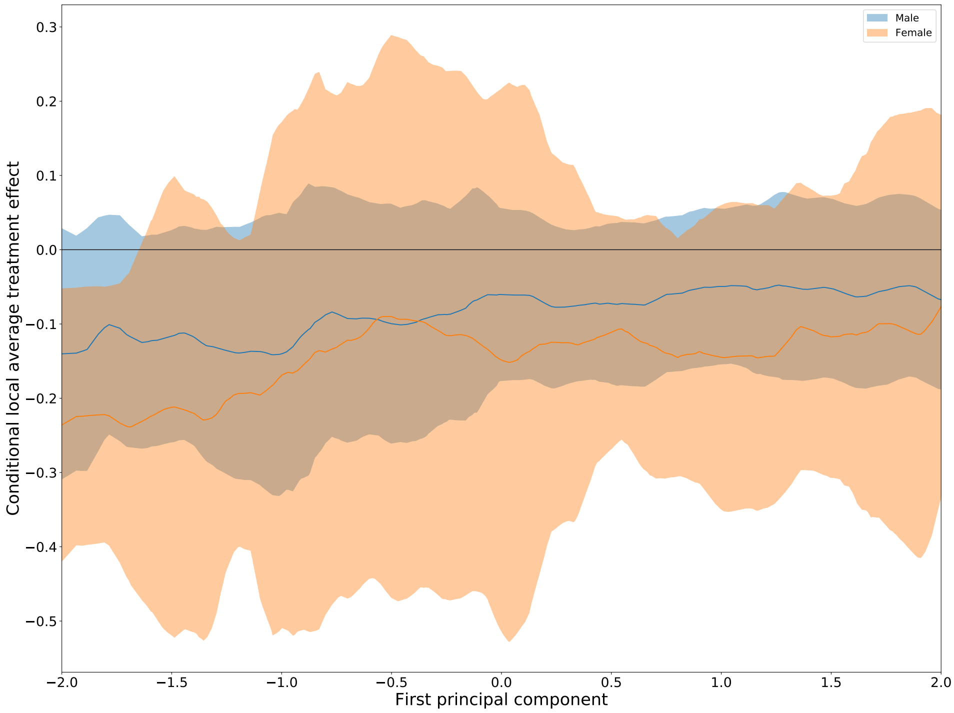



Table 10: Estimated models with a 9 year cutoff

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| OLS | -0.0224*** | -0.0222*** | -0.0216** | -0.0230*** |
| | (0.00837) | (0.00836) | (0.00835) | (0.00823) |
| IV, three inst. | 0.0129 | 0.0112 | 0.0117 | -0.122 |
| | (0.131) | (0.130) | (0.127) | (0.107) |
| IV, all inst. | 0.0314 | 0.0296 | 0.0275 | -0.107 |
| | (0.123) | (0.122) | (0.120) | (0.102) |
| Observations | 7866 | 7866 | 7866 | 7866 |
| Country × cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic × wave | | | YES | YES |
| Country × cohort trend | | | | YES |

Standard errors in (·), clustered on birth cohort and country

$^{*}$ = p < 0.05, $^{**}$= p < 0.01, $^{***}$ = p < 0.001

Figure 13: Heterogeneity by gender, France

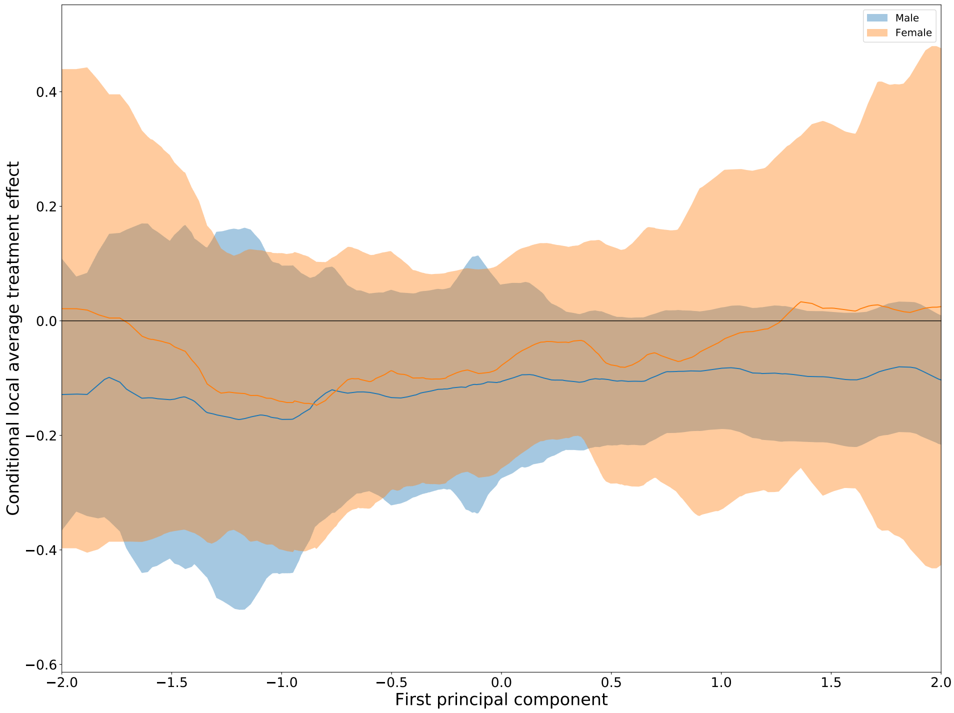



Figure 26: Heterogeneity by gender with a 9 year cutoff

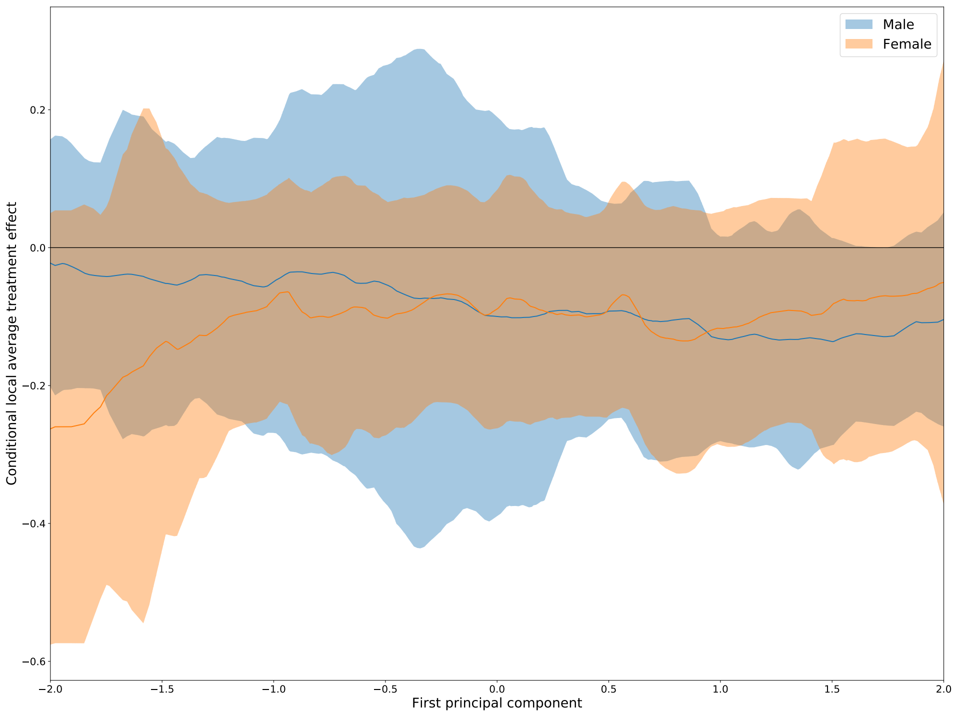


Figure 14: Heterogeneity by gender, Italy

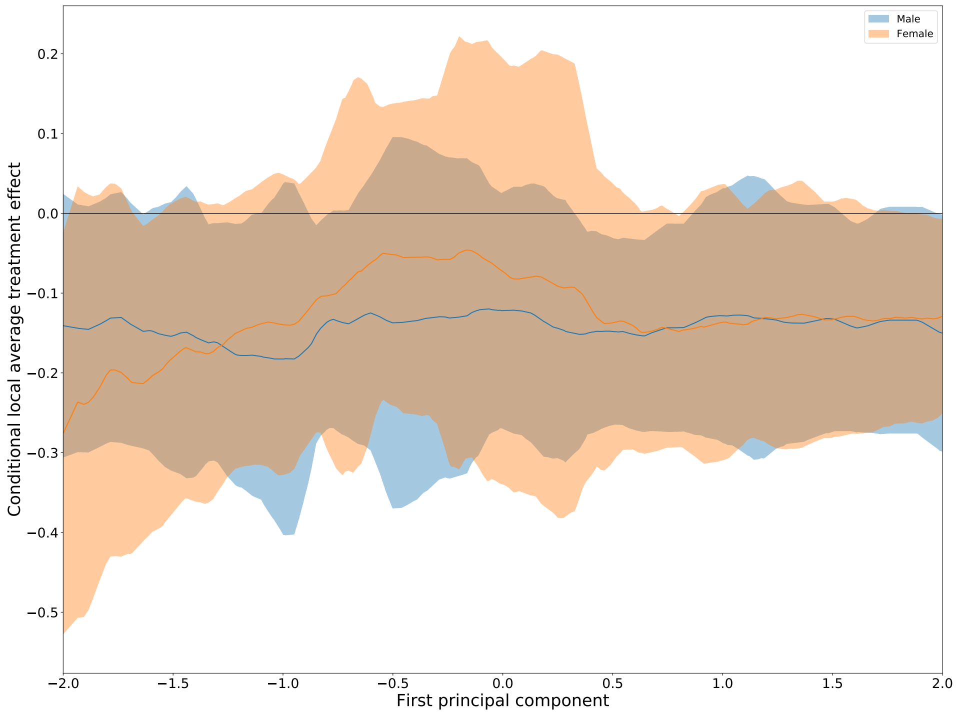



Figure 27: Heterogeneity by gender with a 5 year cutoff

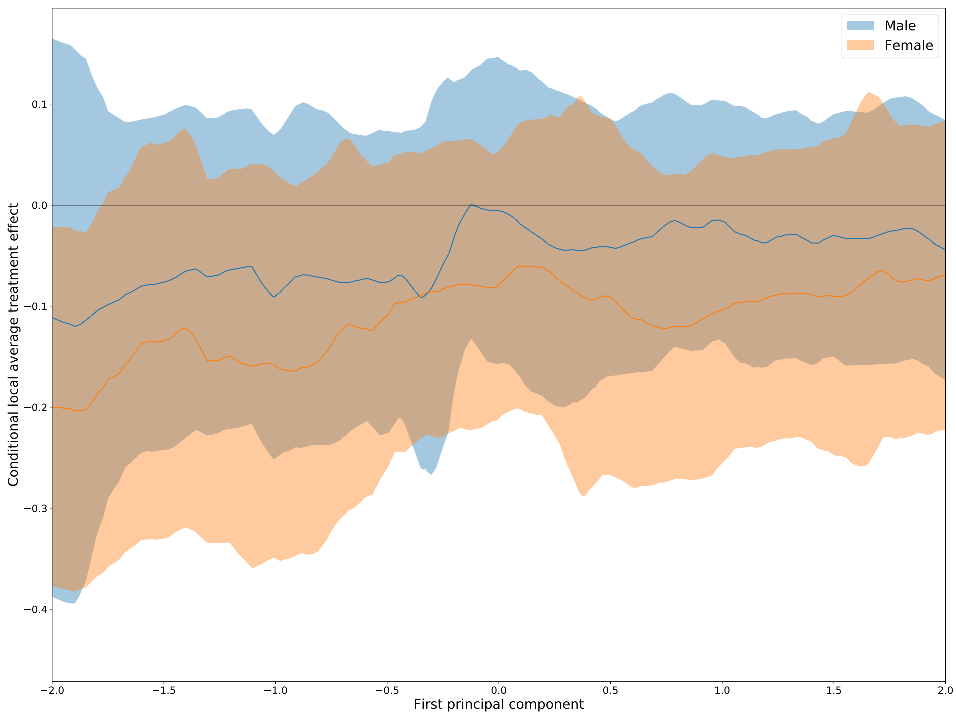


Figure 15: Heterogeneity by gender, The Netherlands

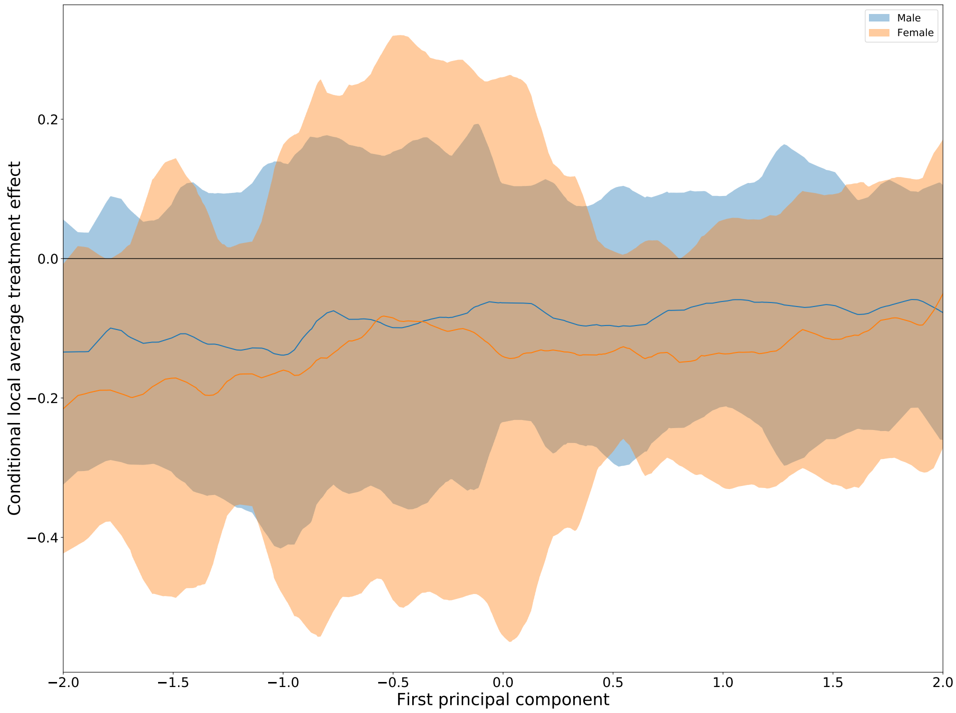



Table 12: Estimated models with a 9 year cutoff

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| OLS | -0.0224*** | -0.0222*** | -0.0216** | -0.0230*** |
| | (0.00837) | (0.00836) | (0.00835) | (0.00823) |
| IV, three inst. | 0.0129 | 0.0112 | 0.0117 | -0.122 |
| | (0.131) | (0.130) | (0.127) | (0.107) |
| IV, all inst. | 0.0314 | 0.0296 | 0.0275 | -0.107 |
| | (0.123) | (0.122) | (0.120) | (0.102) |
| Observations | 7866 | 7866 | 7866 | 7866 |
| Country × cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic × wave | | | YES | YES |
| Country × cohort trend | | | | YES |

Standard errors in $(\cdot)$, clustered on birth cohort and country

$^{*}$ = p < 0.05, $^{**}$= p < 0.01, $^{***}$ = p < 0.001

Figure 16: Heterogeneity by gender, Hamburg

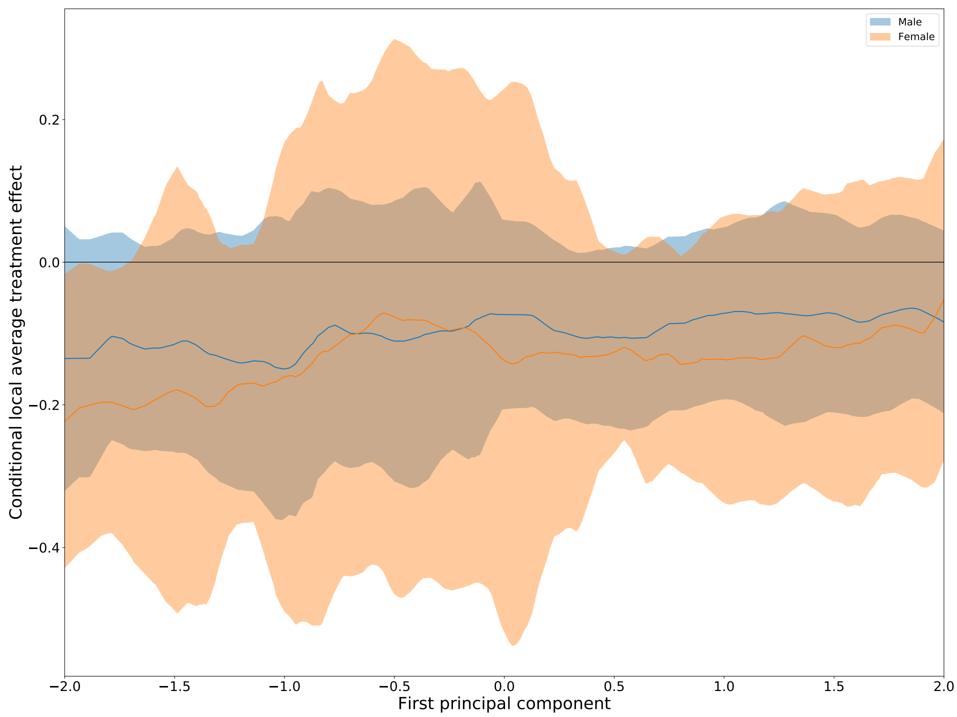



Figure 28: Heterogeneity by gender with a 9 year cutoff

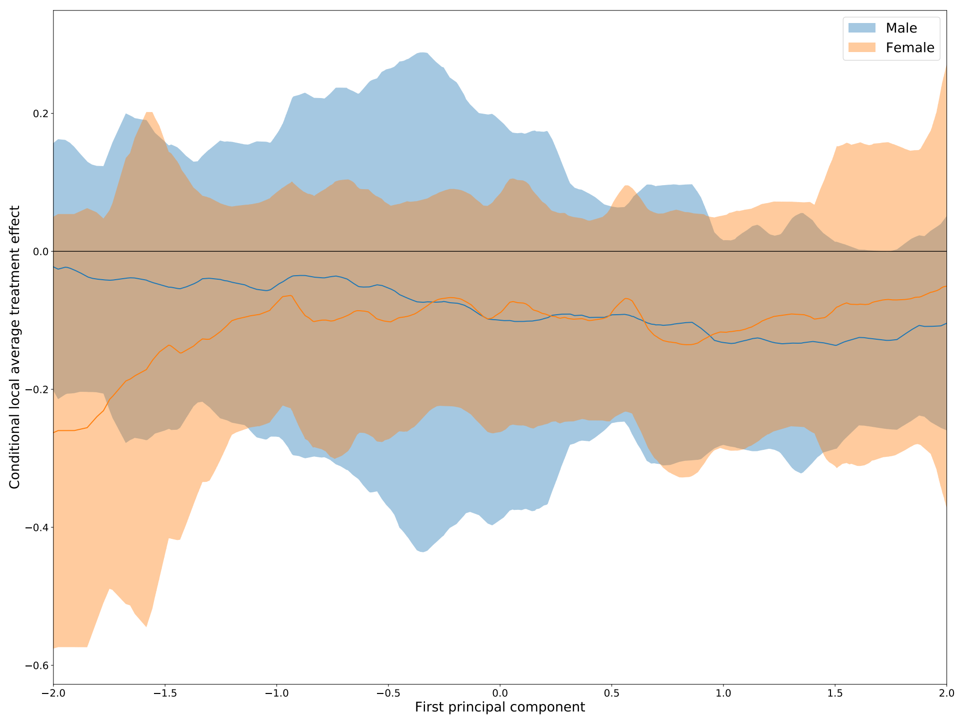


Figure 17: Heterogeneity by gender, Hessen

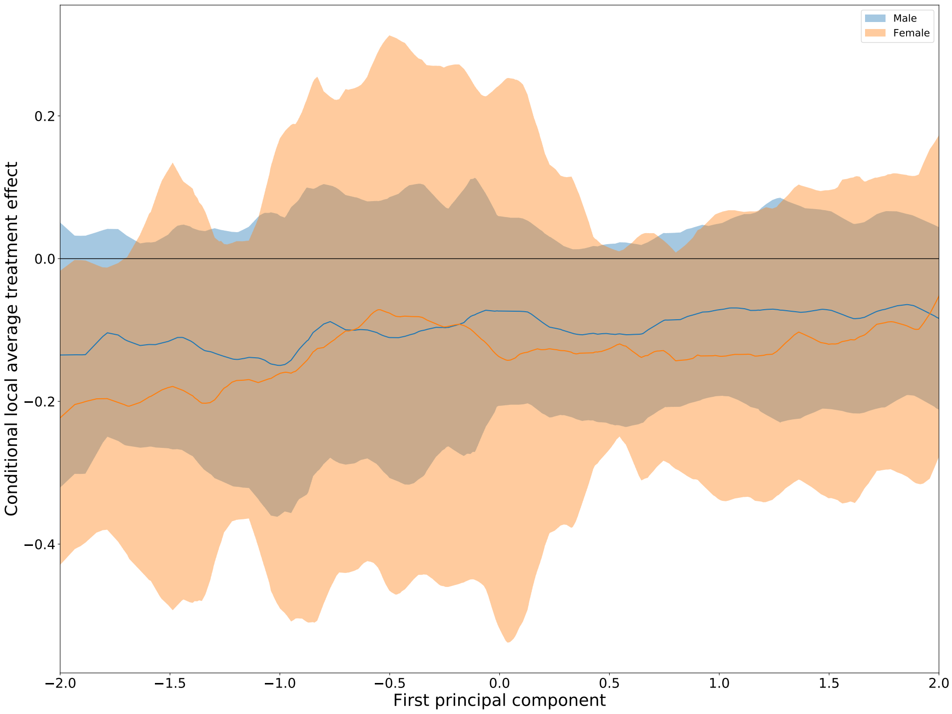



Figure 18: Heterogeneity by gender, Baden-Würtemberg

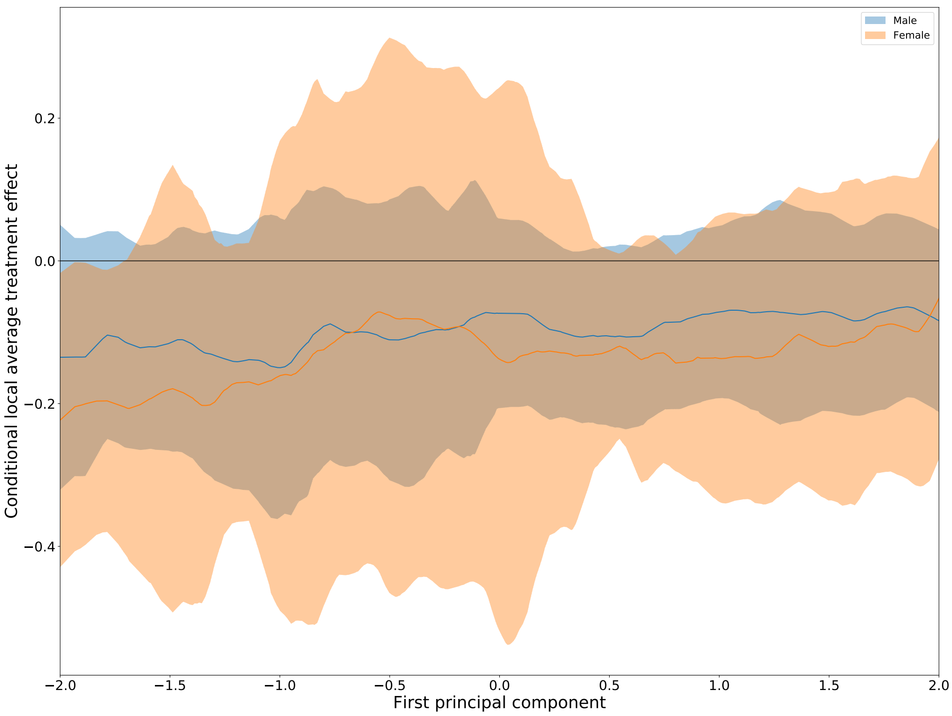


Figure 19: Heterogeneity by gender, Bayern

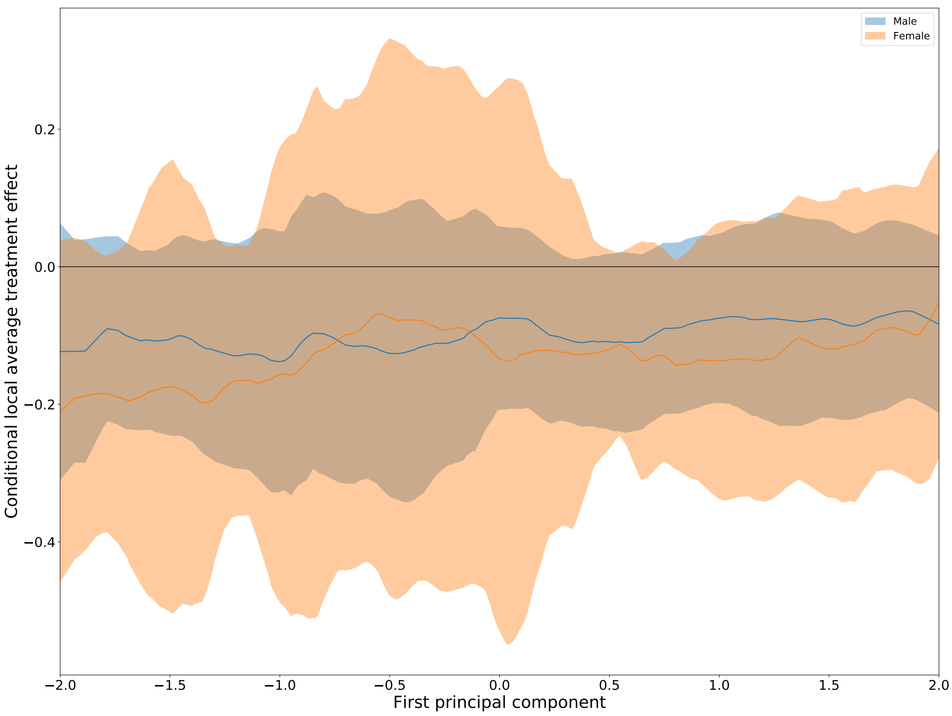



Figure 20: Heterogeneity by gender, Niedersachsen

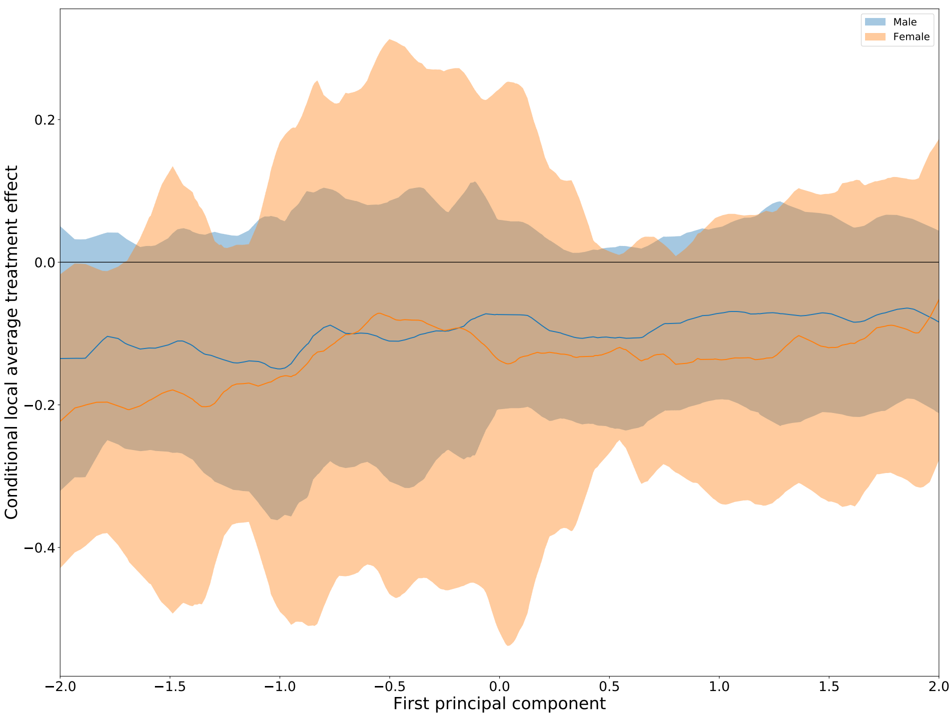


Figure 21: Heterogeneity by gender, Nordrhein-Westphalia

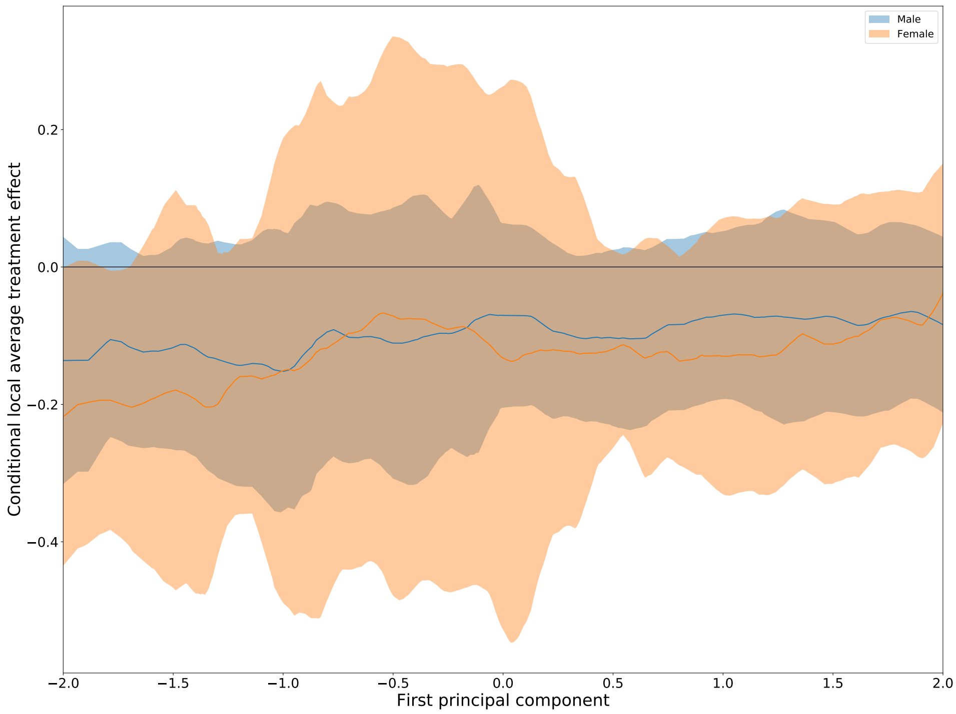



Figure 22: Heterogeneity by gender, Rheinland-Pfalz

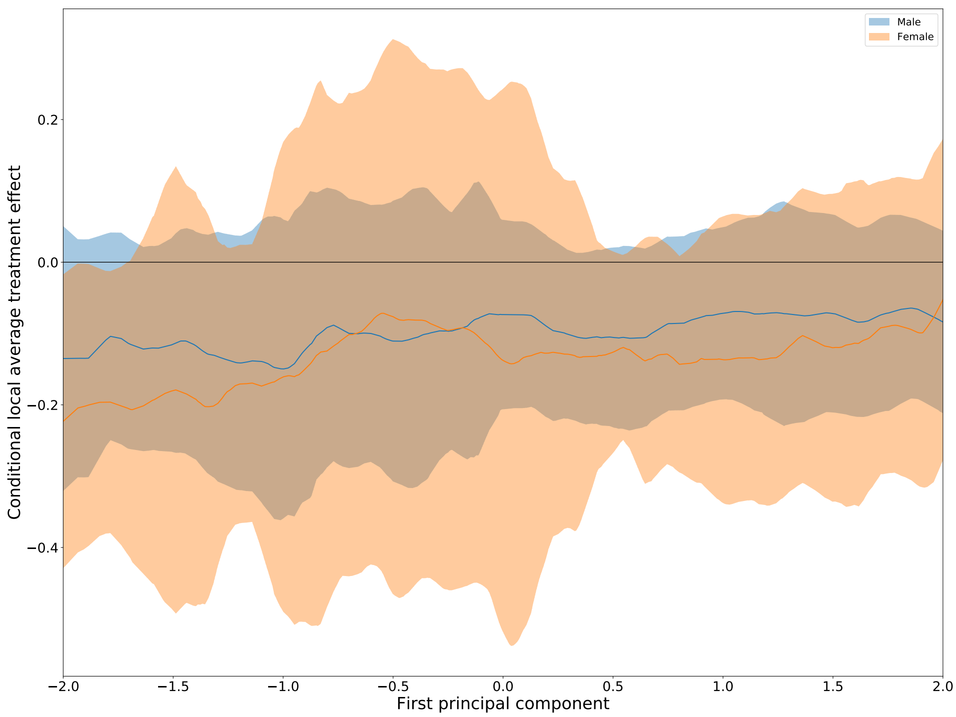


Figure 23: Heterogeneity by gender, Saarland

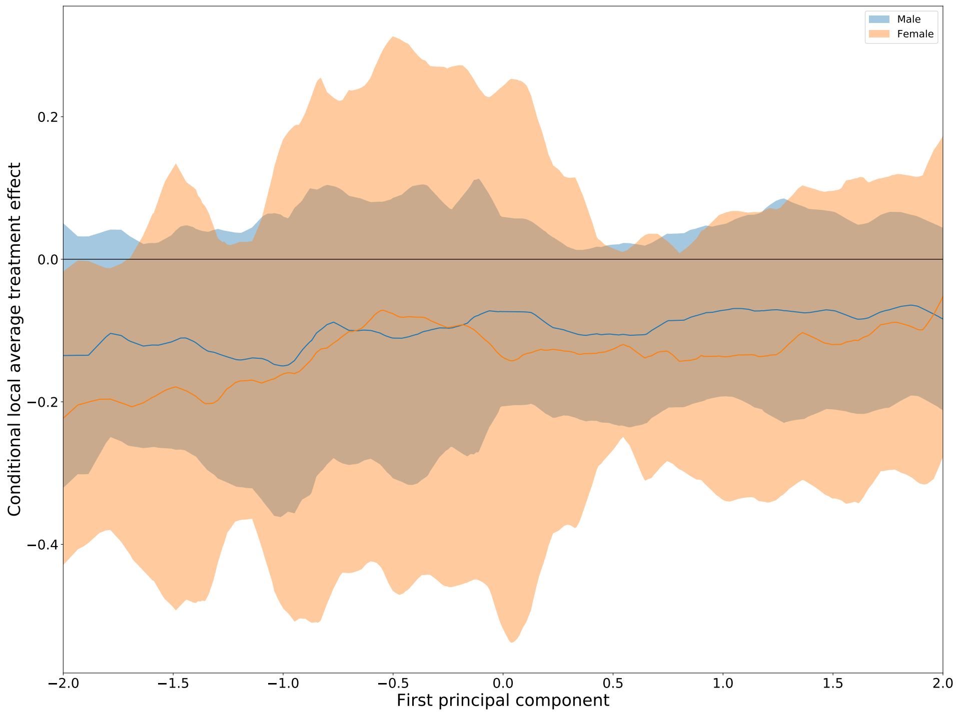



Figure 24: Heterogeneity by gender, Schleswig-Holstein

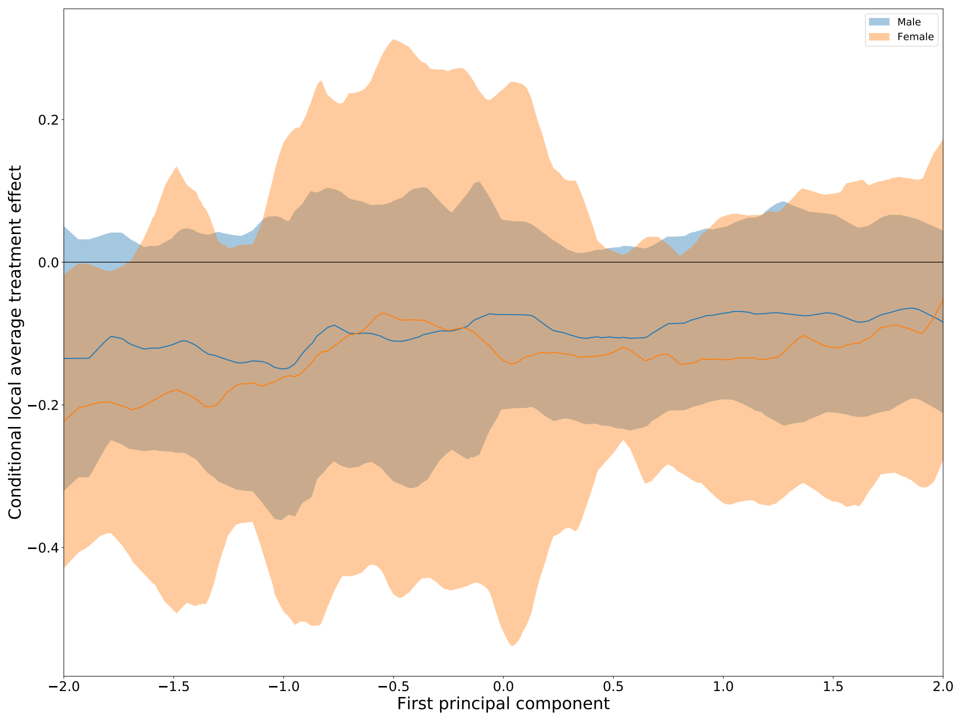


Table 11: Estimated models with a 5 year cutoff

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| OLS | -0.0245** | -0.0243** | -0.0231** | -0.0266*** |
| | (0.0102) | (0.0102) | (0.0102) | (0.00962) |
| IV, three inst. | -0.00803 | -0.00927 | -0.00873 | -0.177 |
| | (0.162) | (0.161) | (0.158) | (0.126) |
| IV, all inst. | -0.0841 | -0.0837 | -0.0872 | -0.207* |
| | (0.154) | (0.153) | (0.150) | (0.121) |
| Observations | 5273 | 5273 | 5273 | 5273 |
| Country $\times$ cohort | YES | YES | YES | |
| Age cubic | | YES | | |
| Age cubic $\times$ wave | | | YES | YES |
| Country $\times$ cohort trend | | | | YES |

Standard errors in $(\cdot)$, clustered on birth cohort and country

$^{*} = p < 0.05$, $^{**}= p < 0.01$, $^{***} = p < 0.001$