

Social Mobility in Germany*

Majed Dodin, Sebastian Findeisen, Lukas Henkel,
Dominik Sachs & Paul Schüle

April 30, 2021

Preliminary Draft

Abstract

We characterize intergenerational social mobility in Germany using census data on the educational attainment of 526,000 children and their parents' earnings. Our measure of educational attainment is the A-Level degree, a requirement for access to university and the most important qualification in the German education system. On average, a 10 percentile increase in the parental income rank is associated with a 5.2 percentage point increase in the probability to obtain an A-Level. Despite a marked increase in the A-Level share from 1997-2018, this parental income gradient remained unchanged for the affected cohorts. At the regional level, there exists substantial variation in mobility estimates. We find suggestive evidence that place effects, rather than sorting of households into different regions, account for most of these geographical differences.

JEL-Codes: I24, J62, R23

Keywords: Social Mobility, Intergenerational Mobility

*Dodin: University of Mannheim (majed.dodin@gess.uni-mannheim.de); Findeisen: University of Konstanz (sebastian.findeisen@uni-konstanz.de); Henkel: European Central Bank (lukas.henkel@ecb.europa.eu); Sachs: LMU Munich (dominik.sachs@econ.lmu.de); Schüle: ifo Munich and LMU Munich (paul.schuele@econ.lmu.de). We thank Klaus Adam, Timm Boenke, Winfried Koeniger, Fabian Kosse, Moritz Kuhn, Marc Piopiunik, Uta Schönberg, Jan Stuhler, Ludger Wößmann, Larissa Zierow and seminar participants in Lisbon, Mannheim, Munich, Tübingen and Vienna for helpful comments. We thank Lea Fricke and Victoria Szabo for great research assistance. Financial support by the Joachim Herz Foundation is gratefully acknowledged. The findings and conclusions expressed are solely those of the authors and do not represent the views of the foundation. This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

1 Introduction

Social mobility is an important indicator for both fairness and economic efficiency in a society. Besides the fairness aspect, low mobility can lead to the misallocation of resources, as talented children from disadvantaged families are impeded from realizing their full potential. Despite its importance, reliable mobility estimates are lacking for many countries. Measuring mobility is challenging as it requires data that link outcomes of parents to a measure of opportunity for children. Household panel data sets may provide this opportunity but are typically too small to deliver estimates for single years or regions with a sufficient degree of statistical confidence (Lee and Solon, 2009; Mazumder, 2018). While in some countries researchers were able to obtain high-quality estimates by using administrative tax records (e.g. Chetty et al., 2014), tax data linking children and parents are often not available. Germany is no exception here, so that to date no precise estimates of intergenerational mobility across time and space exist.

This paper uses census data to document time trends and regional differences in social mobility in Germany at a higher level of detail than previously possible. We measure mobility by the association between the educational attainment of a child and parental income. Our measure of educational attainment, the A-Level (Abitur), is the highest secondary schooling degree in Germany and typically obtained when children still live at home. We can therefore estimate mobility statistics using administrative census data, where children can only be linked to their parents when living in the same household. Our data covers one percent of the German population in every year from 1997 to 2018, providing detailed information on the educational attainment of 526,000 children aged 17 to 21 and their parents socioeconomic status.

We present three main findings. First, mobility at the national level has remained constant for recent cohorts. On average, a 10 percentile increase in the parental income rank was associated with a 5.2 percentage point increase in the probability to obtain an A-Level degree. This parental income gradient has not changed for the birth cohorts 1980-1996, despite a marked increase of more than one third in the share of children obtaining an A-Level. There is also a high degree of stability over time when estimating mobility for population subgroups. These results confirm the view that increased access to higher education alone may not be sufficient to effectively mitigate existing disparities in the intergenerational transmission of resources.

Second, we document substantial geographical variation in intergenerational mobility across local labor markets. For example, moving up 10 percentile ranks in the national parental income distribution increases the probability to graduate with an A-Level degree by 4.6 percent-

age points in Hamburg but 6.8 percentage points in Leipzig. While mobility is lower in East Germany, most of the variation originates from within-state differences. This is remarkable, as education policies, which prior literature has suspected to be a key determinant of mobility, vary mainly on the state level in Germany.

Third, we show that household characteristics can only explain a small fraction of these mobility differences. Spatial variation in mobility estimates can either arise due to place effects or due to systematic sorting of households into different local labor markets. Which answer prevails has important implications about the usefulness of place-based policies. For the US, Rothbaum (2016) and Gallagher et al. (2018) suggest that a substantial share of the geographical variation in the intergenerational mobility measures reported in Chetty et al. (2014) can be explained by differences in household characteristics across commuting zones. However, both articles can only indirectly assess the role of household characteristics for mobility estimates, as the administrative tax data used by Chetty et al. (2014) contain little covariates. By contrast, the census data employed in this paper contain rich information on family background. We can therefore directly test whether mobility differences across local labor markets are reduced when controlling for an extensive set of household characteristics. We find that the mobility ranking between local labor markets is largely unchanged when controlling for household characteristics, indicating that rather place based features account for the geographical variation.

We then explore which regional characteristics are most predictive for areas of high or low mobility. This evidence does not imply causality but can help to augment our understanding on potential mechanisms affecting mobility. Drawing on a rich set of regional indicators, find that local labor market conditions and social characteristics are most predictive of our mobility estimates. For example, mobility is low in regions with a high unemployment rate and a large prevalence of child poverty, and high in regions with a large share of married individuals. Regarding the relevance of the education system, the evidence is not clear cut. On the one hand, the structure and financing of the German education system should ensure that differences in school quality are modest, especially when considering within-state variation. On the other hand, we find that the school drop out rate is by far the most predictive indicator for low mobility, even when conditioning on parental income and labor market conditions. We also find suggestive evidence that labor markets with enhanced possibilities to obtain an A-Level degree on vocational schools are characterized by high mobility.

Apart from providing a comprehensive characterization of social mobility in Germany, this paper contributes to the literature on the measurement of intergenerational social mobility. The current gold standard in measuring intergenerational mobility is to estimate the expected income rank of children conditional on the income rank of their parents using administrative tax data (Dahl and DeLeire, 2008; Chetty et al., 2014). While administrative tax data have key advantages like a large sample size and a high data quality in the upper half of the income distribution, they come with their own limitations¹ and are often not available to researchers due to tax and data protection laws.

Our approach to measure mobility as the association between the parental income rank and the educational attainment after secondary school is an appealing alternative for several reasons. Most importantly, it allows to measure mobility using census data, where children can only be linked to their parents as long as they live in the same household. Census data also improve over standard household surveys in terms of sample size, are representative by construction and – as opposed to administrative tax data – may contain detailed information on the socio-economic situation of households. Moreover, we can obtain mobility statistics for relatively recent cohorts, which is not possible when estimating income mobility.²

Why is the receipt of an A-Level degree a compelling measure of opportunity in the German context? Unlike in the US, where almost all children attend high school and eventually receive a high school degree, Germany is a country with a rigorous early-age tracking system. Only completion of the highest track grants direct access to the tuition-free national university system, opening up the full range of career prospects. As a result, the average A-Level wage premium is as high as 42%, comparable to recent estimates for the college wage premium in the US (Ashworth and Ransom, 2019). Beyond the economic aspects of equality of opportunity, having obtained an A-Level is also an important sign of social distinction in the otherwise rather egalitarian German society. More broadly, a large literature shows that educational attainment has intrinsic value and predicts a wide range of non-pecuniary outcomes (Oreopoulos and Salvanes, 2011; Lochner, 2011). Educational attainment as a measure of op-

¹Tax return data are not representative, as they capture only the tax-paying part of the population. Another important limitation is the absence of information on household characteristics or other outcomes of interest like education or occupation. See Mazumder (2016; 2018) for a discussion.

²A well established literature documents life cycle bias in mobility estimates (Solon, 1992; Haider and Solon, 2006). Due to heterogeneity in life cycle earnings profiles, estimates obtained when children are still in their twenties tend to be downward biased. Haider and Solon (2006) suggest to measure the income of children around the age of 40, when this bias is minimized. This implies that even the most recent evidence on income mobility applies to children born 40 years ago. It is then an open question to what extent the insights retrieved for this generation are still relevant for children born today.

portunity is thus no second best solution in the German context, but, as we believe, a strong and comprehensive indicator for the opportunities of an individual. Beyond Germany, this approach to measure mobility could turn out useful also in other countries where the highest secondary school degree plays a similarly important role in shaping future career options.

The remainder of this paper is organized as follows. Section 2 discusses related literature and summarizes relevant features of the German education system. Section 3 outlines our approach to the measurement of social mobility and how it fits into the German institutional framework. Section 4 describes the data and provides details on the sample construction. In Section 5, we report our results on the national level, while Section 6 shows how mobility has evolved over time. Finally, Section 7 documents the geographical variation in mobility. Section 8 concludes.

2 Related Literature and Institutional Background

2.1 Related Literature

The study of intergenerational social mobility requires a characterization of the joint distribution of parental outcomes and child opportunities. Since economic and social opportunities are difficult to measure, empirical studies of social mobility have generally aimed at the joint distribution of outcomes. The standard approach in the literature has focused on the association of child and parent income, summarized by the intergenerational elasticity of earnings (IGE) (Zimmerman 1992; Solon, 1992; Mazumder, 2005) or more recently by rank-rank correlations (Dahl and DeLeire, 2008; Chetty et al., 2014), which are more robust to non-linearities and measurement issues at the bottom of the income distribution. Both approaches require the researcher to observe child earnings and are therefore only feasible for relatively old birth cohorts. The early canonical studies of intergenerational mobility thus mostly relied on panel data with a long time dimension like the Panel Study of Income Dynamics (PSID) in the US, which allow to observe both parents and children in the age range 30-40.

Given the low sample size of typical household panel data sets like the PSID, estimates are only representative at the national level and it is difficult to detect significant differences between birth cohorts or regions (Mazumder, 2018). A major step forward in terms of data quality has been achieved by Chetty et al. (2014), who use linked administrative tax data to study the association of parent and child incomes. The sample size in this study is an order of

magnitude higher than with any household panel and has allowed to obtain reliable estimates of the geographical variation in mobility, opening up the field of research for a better understanding of the causes of mobility (e.g. Chetty and Hendren, 2018a; 2018b). In the following, a number of studies have applied this concept and measurement framework to other countries, like Acciari et al. (2019) for Italy, Chuard and Grassi (2020) for Switzerland, Corak (2020) for Canada and Deutscher and Mazumder (2020) for Australia. However, for most countries no study of this kind has been forthcoming, as linked administrative tax data are not available everywhere.

Germany is no exception here. As German law forbids to link any kind of tax returns, most empirical evidence on social mobility is based on the German Socio Economic Panel (GSOEP), the German counterpart of the PSID. Like the PSID, the GSOEP is great in providing detailed information about child outcomes and parental background but suffers from a low sample size. In particular, using the GSOEP it is not possible to document time trends or more fine-grained geographical variation in mobility with a sufficient degree of confidence. Schnitzlein (2016) shows that existing mobility estimates based on the GSOEP are sensitive to small variations in sampling criteria, resulting in a wide range of plausible estimates. It is therefore not surprising that the empirical evidence regarding the level of intergenerational mobility in Germany is mixed. Studies that have investigated intergenerational income mobility in the GSOEP include Eisenhauer and Pfeiffer (2008), Riphahn and Heineck (2009), Eberharter (2013) and Bratberg et al. (2017). These studies typically find a higher level of mobility than in the US and lower mobility in East than in West Germany, albeit with high statistical uncertainty.

In addition, several studies have investigated educational mobility in Germany. Closest to our analysis are Riphahn and Trübwetter (2013) and Klein et al. (2019) who draw on census data to investigate differences in educational mobility between East and West Germany. While we also use a (more comprehensive) version of this data, we depart from their analysis in considering the joint distribution of child education and parent income. Furthermore, we are able to capture regional differences in much greater detail. Methodologically, our study is closest to Hilger (2015), who for the US also uses the association of child education and parental income to leverage the potential of census data in the analysis of intergenerational mobility. This means that, as in our case, child outcomes are only observed if children are still co-resident with their parents, posing the risk of a selection bias. However, Hilger (2015) measures educational attainment as years of schooling, which typically manifest much later

in life than a secondary school degree like the A-Level. As a consequence, the author focuses on children aged 26-29, where most children have already left the parental household. This requires an imputation of educational attainment for these children. In our setting, this is not necessary as more than 80% of children in the age range 17-21 still live at home and the move-out decision is uncorrelated with parental income. As we argue in the following, the German education system therefore is – compared to the US – much better suited for this type of analysis.³

2.2 The German Education System

The German institutional framework is particularly suited to study social mobility through the lens of educational opportunities. We utilize the fact that Germany has a three-track system of secondary education, where only the highest track allows to access the university system. When children finish the four-year⁴ elementary school around the age of 10, they are allocated into three tracks. Only successful completion of the high track, the Gymnasium (grades 5-12/13), eventually results in the award of an A-Level degree (Abitur). The other two tracks last five (grades 5–9) and six years (grades 5–10) and focus on preparing students for an apprenticeship. The rigor of the tracking system is mediated by the possibility to switch tracks. In particular, it is common that talented students from the medium track later switch to the general high track or to attend a specialized high track. This entails that students "misallocated" after elementary school still have the possibility to correct for this later in their school time.⁵

The responsibility for the education system falls under the jurisdiction of the German states and not under the jurisdiction of the federal government. While the exact implementation of the school system can hence vary across states, these differences do not extend to the A-Level degree. The Standing Conference of State Education secretaries has the stated goal to ensure a high degree of comparability of educational qualifications across German States and there are no legal differences between the A-Level degrees issued from different states. An A-Level

³ A small but growing literature on educational mobility in developing countries also relies on children's co-residency with their parents (Asher et al., 2020; Alesina et al., 2021; Muñoz 2021). These studies typically try to minimize selection bias by focusing on simple measures of educational attainment that can already be observed early in life, like having obtained an elementary school degree.

⁴In the states of Berlin and Brandenburg, elementary school lasts six years.

⁵A more detailed overview on the tracking system and track switching is provided in Dustmann et al. (2017) and Biewen and Tapalaga (2017).

degree grants access to the free national university system.⁶ As schools, universities are state-financed, mostly based on student headcounts. This generates a comparatively large equality in the endowment and quality between different universities.

In consequence, the A-Level degree is by far the most important qualification in the German education system and individuals who obtain it enjoy substantially above-average economic outcomes. Using data on full-time workers aged 30-45, we find an A-Level wage premium of 42% for monthly net income.⁷ This estimate mirrors Schmillen and Stüber (2014) who report a 44% A-Level wage premium for total gross lifetime earnings. In sum, the A-Level wage premium is as high as recent estimates for the college wage premium in the US (Ashworth and Ransom, 2019). An A-Level degree is also associated with a lower risk of being unemployed (Hausner et al., 2015) and a higher life expectancy (Gärtner, 2002). Furthermore, it constitutes a beneficial factor for obtaining vocational training in certain white-collar occupations (Klein et al., 2019) and marks an important sign of social distinction in the otherwise rather egalitarian German society. All of this illustrates that an A-Level degree is indeed a meaningful measure of opportunity and strongly predictive for future economic well-being. Finally, one should not forget that many people value education as a worth in itself, regardless of the associated economic benefits.

3 Measuring Intergenerational Mobility

We define two sets of mobility statistics with the aim to distinguish between two mobility concepts: relative and absolute mobility. Measures of relative mobility compare the outcomes of children from low-income families to those of children from high-income families, whereas measures of absolute mobility ask about the absolute level of opportunity for children from low-income families.

⁶For students with low parental income, there even exists a very generous funding scheme (BAföG) to cover living expenses. This need-based state student finance has been in existence since 1971. Half of the financial support is usually a grant, the remaining half an interest-free loan. It is therefore fair to say that from the financial side there exist no binding constraints for taking up university studies in Germany.

⁷We use the waves 1997-2018 of the German Mikrozensus (described below) and compute the A-Level wage premium by regressing net monthly personal income of full-time workers aged 30-45 on an A-Level dummy, controlling for a full set of age and year fixed effects.

Relative Mobility. Relative mobility is concerned with the difference in opportunities between children from low-income families relative to children from high-income families. In the context of our analysis, this can be summarized by the question:

"How does the probability of obtaining an A-Level degree differ between children from low-income households and children from high-income households?"

To answer this question, we estimate the slope coefficient β in a regression of our binary outcome variable Y_i , indicating whether a child has obtained an A-Level degree, on the parental income rank R_i :

$$Y_i = \alpha + \beta R_i + \varepsilon_i \quad (1)$$

If the relationship between parental income rank and the probability of obtaining an A-Level degree is well described by a linear relationship (we provide evidence that this is indeed the case later on), the slope coefficient of this regression provides a parsimonious statistic of relative mobility. The probability of obtaining an A-Level degree for a child from a given income percentile r is then given by:

$$E(Y_i|R_i = r) = \alpha + \beta r \quad (2)$$

and the difference in probabilities between two parental income percentiles r and q , $r > q$, is given by:

$$E(Y_i|R_i = r) - E(Y_i|R_i = q) = \beta(r - q) \quad (3)$$

The slope coefficient thus summarizes the difference in the probability of obtaining an A-Level degree between households from different income percentiles. For better readability, we always multiply β by 100. In what follows, we refer to $\beta \times 100$ as the parental income gradient, our main measure for relative mobility. A high value of the parental income gradient implies a low degree of mobility.

Next to the parental income gradient, we estimate a second statistic of relative mobility, the Q5/Q1 ratio. It is defined as the share of children who obtain an A-Level degree from the top quintile of the parental income distribution divided by the share of children who obtain an A-Level degree from the bottom quintile of the parental income distribution:

$$\text{Q5/Q1} = \frac{E(Y_i|R_i > 80)}{E(Y_i|R_i \leq 20)} \quad (4)$$

This measure can be interpreted as the relative likelihood of obtaining an A-Level degree between both groups. A ratio of $Q5/Q1 = 2$ would mean that children from the top quintile of the income distribution are twice as likely to obtain an A-Level degree as children from the bottom quintile of the income distribution. A high value of the $Q5/Q1$ ratio thus implies a low degree of relative mobility. Conceptually, the $Q5/Q1$ ratio differs from the parental income gradient in that it is sensitive to the absolute probability of obtaining an A-Level in both quintiles.

Absolute Mobility. Next to relative mobility, one may also be interested in the following question:

"What is the probability of obtaining an A-Level degree for a child from a given (poor) household income level?"

The focus of this question is different from the one before. The question posed here is concerned with the absolute outcomes of disadvantaged children, regardless of the outcomes of children from more advantaged households. We measure absolute mobility by the probability of obtaining an A-Level degree for a child from the bottom quintile of the national income distribution:

$$Q1 = E(Y_i|R_i \leq 20) \quad (5)$$

We refer to this statistic as the Q1 measure. A high value of the Q1 measure implies high absolute mobility.

4 Data

We use the German Microcensus (Mikrozensus, hereafter MZ), an annual representative survey of the German population, administered by the Statistical Office of Germany. The survey is comparable to, but more detailed than, the American Community Survey and constitutes the largest survey program of official statistics in Europe. Each year 1% of the population living in Germany is randomly selected to participate in the survey. It covers a wide range of topics, including family status, citizenship, living conditions, labor market participation, income and educational attainment.⁸ In West Germany, the first MZ was administered in 1957,

⁸For more details on the survey and the sample design see <https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Bevoelkerung/mikrozensus-2017.html>.

in East Germany in 1991. In our analysis, we use the waves 1997 to 2018 of the MZ, where a consistent definition of the education outcomes of interest is available.

The MZ allows to directly observe family ties within each household. Household members are obliged by law to provide information on every person registered at their respective household. Consequently, we are able to match children to their parents as long as they are still registered at their parents' household. Table 1 displays the share of children living with their parents by age of the child, calculated from our data. Virtually all children younger than 15 still live with their parents. From then on, the share of children co-residing with their parents is decreasing with child age. While 92% of 18 year olds are living with their parents, only 44% of 23 year olds still live at home.

Table 1: Co-residence Rate by Child Age

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

Notes: This Table reports separately by age the fraction of individuals living with their parents in the MZ waves 1997 to 2018.

Measuring Educational Attainment of Children. The primary measure for educational attainment of children in our analysis is a binary variable Y_i that is equal to one if a child has obtained a school degree that qualifies for university entrance or if a child is on track to obtain such a degree, and zero otherwise. Specifically, our child outcome variable is equal to one if (i) a child has obtained a secondary school degree that qualifies for tertiary education⁹ or if (ii) a child is enrolled in the last 2-3 years of a track which leads to such a degree at the successful completion of school.¹⁰ We refer to this outcome as an A-Level degree. We choose this definition in order to also include younger children in our sample, for which move-out of home is less of an issue than for older children (see Table 1). Furthermore, using this broader definition reduces measurement error in our outcome measure for children aged around 18 to 19. A-Level degrees are usually awarded in the second quarter of the calendar year, while answers to the MZ are collected on a rolling basis. Hence, we have a non-negligible amount of

⁹This includes *Allgemeine Hochschulreife* (*Abitur*), *Fachgebundene Hochschulreife* and *Fachhochschulreife*.

¹⁰This definition subsumes all students on *Allgemeinbildende Schulen* enrolled in the *Gymnasiale Oberstufe* as well as students from specialized tracks like the *Berufliches Gymnasium* or the *Fachoberschule* which award an A-Level degree.

children in our sample that are surveyed before their graduation, but eventually graduate with an A-Level degree in the same year. Back of the envelope calculations suggest that, if we only count children who have already obtained an A-Level degree, we would miss-measure our outcome variable for around 40% of the graduating cohort in the survey year.¹¹ On the other hand, the share of children failing the final examination in a given year is very low (around 3 percent on average Germany-wide in 2014).¹²

Measuring Parental Income. We measure parental income as the monthly net household income, excluding the monthly net income of all dependent children. This income measure includes all sources of income, including labor income, firm profits and social security transfers received. Our data contain a continuous measure of household income, which we use as our main income measure. This income measure is not asked directly in the survey, but imputed by the Statistical Office. Respondents in the MZ report both their personal and their household income in 24 predefined bins. The Statistical Office then transforms the personal binned income into a continuous variable, essentially assuming that individuals are equally distributed within each bin. In a second step, these values are summed up to a continuous measure of household income. We thus measure parental income with some imprecision. As this measurement error is independent from the educational outcomes of children, it could theoretically lead to a small attenuation bias. We would then slightly underestimate the influence of parental background (overestimate mobility) in all of our measures. However, Nybom and Stuhler (2017) show that rank-based mobility measures like the ones employed in this paper are rather robust to attenuation bias. To account for differences in need and standard of living by household composition we scale all household incomes by the modified OECD equivalence scale. Appendix A shows that the choice of a scaling factor is not influential for our results. Finally, following Dahl and DeLeire (2008) and Chetty et al. (2014), we assign every household its percentile rank in the income distribution relative to all other households within our sample, separately for every year. We refer to this variable as the parental income rank R_i .

¹¹Consider the following example. A child is currently enrolled in the last year of school of the high track. The child graduates from the high track in July, obtaining an A-Level degree. The household of the child is surveyed in the MZ in February, when the child has not graduated yet. Hence, if we only take into account completed A-Level degrees, we would miss-measure the A-Level degree for around 40% of the children graduating from the

Table 2: Summary Statistics by Age of Children

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

Notes: This Table reports separately by age summary statistics for children living with their parents in the MZ waves 1997 to 2018. As in our baseline estimates, parental income is equivalized using the modified OECD scale. A summary statistic that systematically changes with age is correlated with the move-out of children, while a statistic that does not change is uncorrelated.

Sample Definition and Summary Statistics. In our primary sample, we restrict our analysis to children aged 17 to 21, living in the same household as at least one parent. The age cut-offs of 17 and 21 are chosen to balance the following trade-off: For older children, our child outcome variable is measured more precisely, i.e. we do not need to rely on enrollment in the high track but are more likely to observe the completed degree. However, the older the child, the higher the fraction that already moved out of the parental household. We therefore choose the higher cut-off at 21 because move-out increases strongly after this age. The lower cut-off is chosen as children enrolled in an A-Level track are typically at least 17 years old. But even with this conservative age range, we could under- or overestimate mobility with our measurement approach if the decision to move-out were systematically correlated with parental income and the educational attainment of children. Table 2 suggests that this is not the case by showing how key time-constant characteristics of the children in our sample change with age. If move-out would be random, we should not see systematic changes in these statistics for older children. Unsurprisingly, move-out is not random and stratified along a range of different characteristics like gender. However, there is no significant change in the mean parental

high track in a given year. This calculation is based on the fact that data collection for the MZ is distributed evenly over the year.

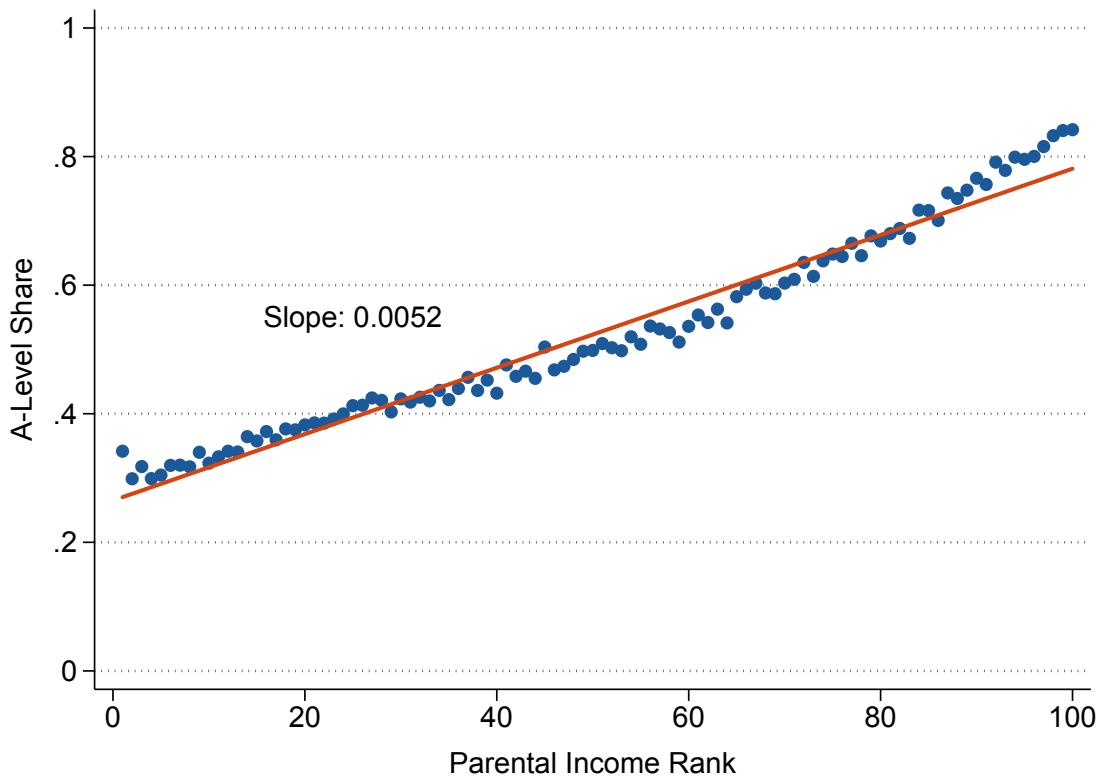
¹²For an overview over the share of children failing the final examination see <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html>

income rank from age 17, where 97% of children still live at home, to age 21, where only 62% of children still live with their parents. This suggests that sample selection is not a major concern for our analysis. More importantly, we demonstrate in the next section that choosing an alternative age range is not influential for our results.

5 National Estimates

In this section, we characterize social mobility at the national level. Figure 1 shows how the share of children with an A-Level degree varies with the position in the parental income distribution.¹³ The graph reveals an essentially linear relationship. As a consequence, the empirical

Figure 1: Social Mobility at the National Level



Notes: This figure shows a non-parametric binned scatter plot of the fraction of children aged 17-21 that are either enrolled in the last two/three years of the A-Level track or already completed the A-Level degree by percentile rank of their parents in the national income distribution in the period 2011-2018. The parental percentile rank is computed within each child cohort. The OLS slope of 0.0052 reported in the figure is estimated using the underlying micro data.

¹³For the national estimates, we pool our data in the period 2011-2018 to achieve a high sample size for the most recent years. As shown later in this paper, the year range chosen here is not influential as the parental income gradient remained remarkably stable for the time of our analysis.

conditional expectation function of obtaining an A-Level is well approximated by the slope coefficient as estimated in Equation 1. We find a slope coefficient of $\beta = 0.0052$, which translates into a parental income gradient of $\beta \times 100 = 0.52$. This means that a 10 percentile increase in the parental income rank is associated with an increase in the probability of obtaining an A-Level by 5.2 percentage points.

As suggested in Section 3, we also compute quintile-based measures of intergenerational mobility. We find $Q1 = 0.34$, meaning that in the bottom quintile of the parental income distribution, 34% of children eventually obtain an A-Level degree. This compares to 52% on average and 72% in the top quintile of the parental income distribution ($Q5 = 0.72$). Accordingly, the $Q5/Q1$ ratio, our second measure of relative mobility, is $0.72/0.34 = 2.13$: a child in the top quintile of the parental income distribution is more than twice as likely to graduate with an A-Level than a child in the bottom quintile. Table 3 shows that these estimates are robust to choosing an alternative age range: If we only select children aged 17-19 into our sample, the estimates of the unconditional A-Level share, the $Q1$ and $Q5$ measures and the gradient change only in the third decimal place. This shows that the move-out of many 20 and 21 year olds (compare Table 1) does not introduce a significant bias. This is also confirmed by the estimates for children aged 19-21, which again differ only marginally. Once we select children aged 16-22, we underestimate the $Q5$ measure. This is due to the fact that many 16 year olds – mostly from high-income households – are still in the school year prior to the A-Level track and are thus misclassified by our measure. Yet, also in this sample we estimate an only marginally lower parental income gradient of $\beta \times 100 = 0.49$.

Table 3: Robustness of National Estimates

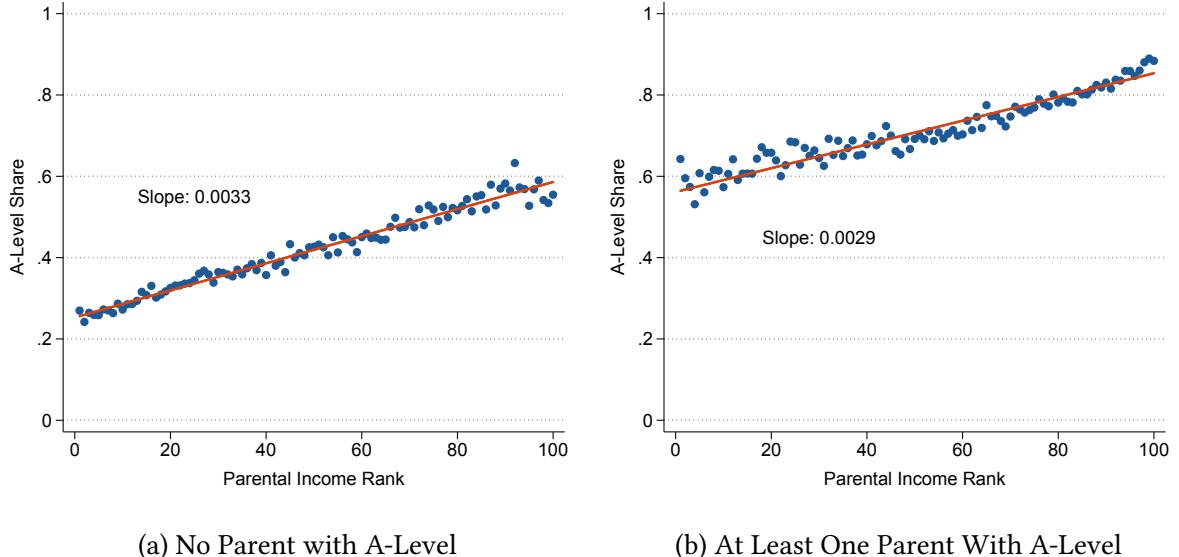
Age Range	A-Level Share	Q1	Q5	Q5/Q1	Gradient	N
17-21	0.52	0.34	0.72	2.13	0.52	224,017
16-22	0.48	0.31	0.67	2.18	0.49	307,050
17-19	0.52	0.34	0.72	2.16	0.52	147,035
19-21	0.53	0.35	0.72	2.09	0.51	122,345

Notes: This table displays mobility statistics at the national level for MZ waves 2011-2018. The results in the first row (age range 17-21) are our baseline results. The bottom three rows show how these statistics change once we select a different age range for the children in our sample.

Do these estimates depict Germany as a country of high or low mobility? A cross-country comparison of our results is not straightforward, as the institutional setting exploited in this paper is quite unique. For the US, we are aware of two studies which investigated similar

relationships. Closest is Hilger (2015), who uses data from the Census 2000. For children aged 19-21, he finds that a 10 percentile increase in parental income rank is associated with an increase in the share of children who attend college by 3.6 percentage points. A much higher number is reported in Chetty et al. (2014), who find an association between parental income rank and college enrollment of 6.7 percentage points. This difference may arise due to issues with sample selection in the study of Hilger (2015). Both estimates compare to the 5.2 percentage points for the share of children obtaining the possibility to attend college in our setting. Under the assumption that the likelihood to attend college conditional on having obtained an A-Level degree is weakly increasing in the parental income rank, we would expect an association between parental income rank and college enrollment of at least 5.2 percentage points in Germany. Therefore, the mobility estimates documented for Germany in this article are not far from recent evidence for the US, where researchers have for a long time stressed the difficulties of children from low-income households to access college.

Figure 2: Differences by Parental Education



Notes: This figure shows a binned scatter plot of the fraction of children aged 17-21 where none of the parents has obtained an A-Level degree (panel a) or where at least one of the parents has obtained an A-Level degree (panel b) that are either enrolled in the last two/three years of the A-Level track or already completed the A-Level degree by percentile rank of their parents in the national income distribution in the period 2011-2018. The parental percentile rank is computed within each child cohort and the OLS slopes reported in the figure are estimated using the underlying micro data.

Subgroup Estimates. Next, we analyze how mobility varies between different population subgroups. We continue to rank parents according to their position in the aggregate income

distribution. Figure 2 displays the parental income gradient separately for children from households where no parent has an A-Level degree (panel a) and for children from households where at least one parent has an A-Level degree (panel b). As expected, children with low parental education have a much lower baseline probability to obtain an A-Level. Remarkably though, this difference is close to constant across the whole income distribution and the parental income gradients in both groups are very similar: $\beta \times 100 = 0.33$ for children without parental A-Level degree and $\beta \times 100 = 0.29$ for children with parental A-Level degree. Furthermore, the probability of obtaining an A-Level still rises linearly in the parental income rank in the two groups defined by our conditioning variable.

Table 4: Intergenerational Mobility for Subgroups

		A-Level Share	Q1	Q5	Q5/Q1	Gradient	N
Gender	male	0.47	0.29	0.68	2.35	0.53	119,975
	female	0.58	0.40	0.77	1.93	0.50	104,042
Parental education	no A-Level	0.39	0.29	0.49	1.73	0.33	141,148
	A-Level	0.75	0.61	0.83	1.36	0.29	82,869
Migration status	native	0.54	0.32	0.73	2.25	0.55	163,018
	migrant	0.48	0.36	0.68	1.90	0.47	60,908
Parenting status	single parent	0.47	0.34	0.65	1.92	0.50	49,397
	two parents	0.54	0.34	0.73	2.16	0.55	174,620
Parental marital status	not married	0.47	0.33	0.65	1.97	0.46	51,018
	married	0.54	0.35	0.73	2.11	0.54	172,999
Region	East Germany	0.51	0.31	0.77	2.49	0.60	28,872
	West Germany	0.52	0.35	0.72	2.08	0.51	195,145
Siblings	Yes	0.52	0.35	0.74	2.13	0.55	151,793
	No	0.52	0.32	0.70	2.21	0.49	72,224
Birth order	first child	0.53	0.34	0.73	2.13	0.51	160,884
	second child	0.51	0.34	0.71	2.10	0.52	54,978
	after second child	0.45	0.32	0.62	1.97	0.57	8,155

Notes: This table shows differences in mobility estimates for selected subgroups in the MZ waves 2011-2018.

Next to parental income, we also check for other differences between subgroups of the German population, summarized in Table 4. The baseline probability to obtain an A-Level is lower for males, migrants and children late in the birth order, as well as for children where only one parent is present or the parents are not married. The parental income gradient differs

mainly by migration status and the parental marital status. For children where parents are not married, the association between parental income and the probability to obtain an A-Level is particularly strong. The opposite holds true for migrants, where parental income is comparatively less important. Interestingly, the Q1 share is larger for migrants than for natives. In addition, relative mobility is lower in East Germany than in West Germany. We investigate this regional pattern in more detail in Section 7.

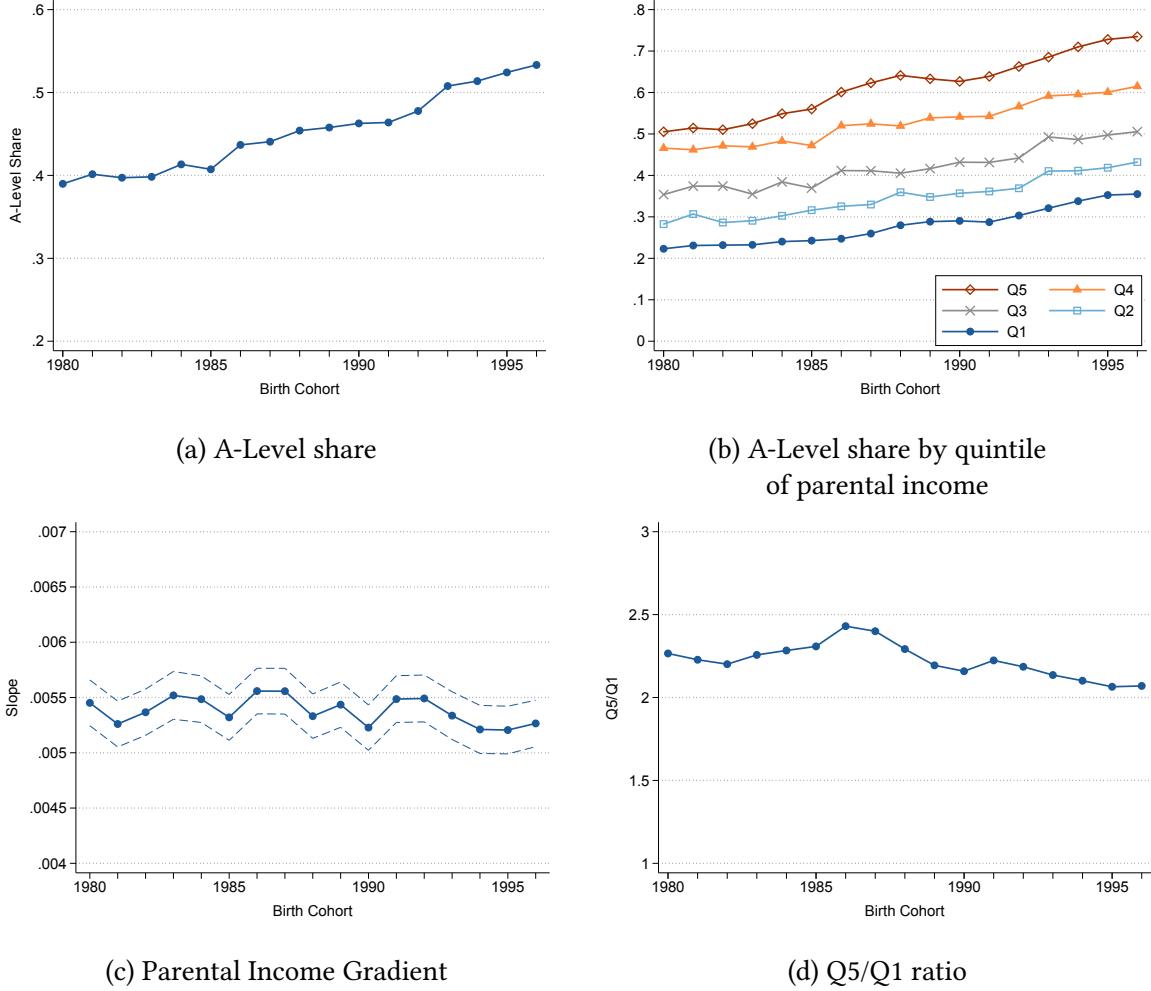
6 Time Trends

How have these mobility statistics developed over time? To answer this question, we draw on all MZ waves available in our data (1997-2018). We focus on a sample of 526,000 children in birth cohorts 1980-1996 for which we have a balanced age distribution.¹⁴ Compared to most other studies of intergenerational mobility, we thus study relatively recent cohorts. Children born during these years faced a marked increase in access to tertiary education. As illustrated in Figure 3, panel (a), the share of children with an A-Level degree increased from 39% for children born in 1980 to 53% for children born in 1996, implying an increase of about one third. Panel (b) shows that the absolute increase in the share of awarded A-Levels was more or less the same in each quintile of the parental income distribution. Although the share of children in Q5 increased somewhat above-average, there is no strong pattern indicating that the expansion of the A-Level share happened predominantly among high income households. In consequence, the parental income gradient remained remarkably stable over all cohorts (Figure 3, panel a). The Q5/Q1 ratio in contrast slightly decreased.

How can we rationalize these findings? To see how a roll-out of higher secondary education affects our mobility measures, let's assume for concreteness that the increase in the A-Level share was achieved by opening up one additional class in an A-Level track. This gives additional (marginal) children the opportunity to obtain an A-Level. If only one of those marginal children is from a family in the bottom quintile of the parental income distribution, this leads to an increase in absolute mobility as measured by the Q1 measure. Absolute mobility as measured in our context therefore tends to mechanically increase in the A-Level share.

¹⁴This means that for each of these cohorts, the relative share of 17, 18, 19, 20 and 21 year olds does not change. This way, we ensure to hold constant age when comparing two different cohorts. We do so to rule out that the time trends presented in this section may be driven by changes in the age-distribution across cohorts.

Figure 3: Time Trends



Notes: The top two panels of this figure show the development of the A-Level share in the aggregate sample (panel a) and separately for each quintile of the parental income distribution (panel b) for birth cohorts 1980-1996 in the MZ. At the bottom, panel (c) displays the development of the parental income gradient (including 95% confidence bands), and the Q5/Q1 ratio is depicted in panel (d).

The reaction of the two relative measures of mobility depends on whether the probability to enter this marginal class varies with parental income. Panel (b) suggests that selection in fact happened independently of parental income, that is we have equal shares of children from Q1, Q2, Q3, Q4 and Q5 in the class. This decreases the Q5/Q1 ratio, which is sensitive to changes in the absolute probabilities of obtaining an A-Level in the bottom and top quintile of the income distribution, but does not change the absolute difference between the income ranks and therefore does not affect the parental income gradient.

In terms of relative mobility, there are at least two readings possible. A positive view would emphasize that the influence of parental income is strongly attenuated for the marginal children and point to the decline in the Q5/Q1 ratio. A less optimistic view would highlight that

one should even expect that children from low income households catch up and are oversampled in the marginal class, which is something we do not observe: If ability is only weakly correlated with parental income, the share of talented children which is not yet enrolled in an A-Level track should be higher at the bottom of the income distribution. This means that if the receipt of an A-Level would only depend on ability, an expansion of the A-Level share should incentivize relatively more children from low-income households to enroll in an A-Level track. We do not take a stance here which reading is appropriate but emphasize that trends in mobility critically depend on the exact mobility statistic under consideration.

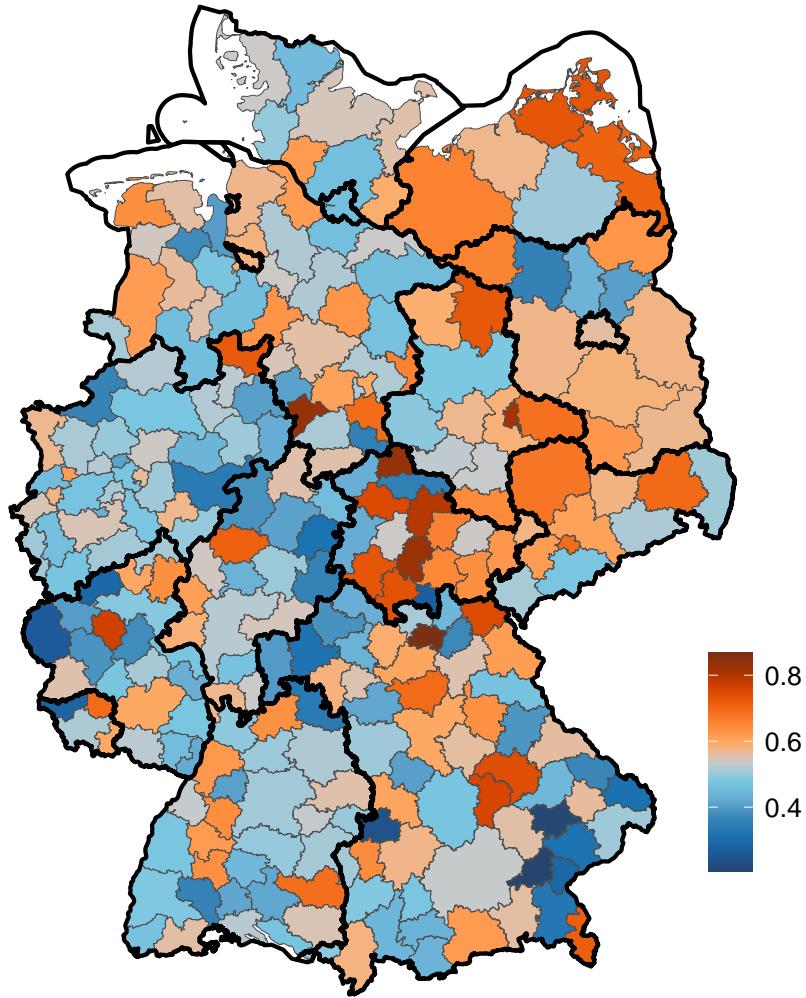
7 Geographical Variation

We now turn to a detailed characterization of regional differences in social mobility across Germany. The unit of analysis is the level of local labor markets (LLM). The 258 LLMs in Germany present an aggregation of counties based on commuting flows, comparable to the definition of Commuting Zones in the US. With the exceptions of the five local labor markets Bremen, Bremerhaven, Hamburg, Mannheim and Ulm, all are strict subsets of states. We assign households to the local labor market of their current place of main residence as reported in our data.¹⁵ The median number of children in our sample (observations) per LLM is 552 (mean: 895). The lowest number of observations across all LLMs is 100 (LLM Sonneberg) and the largest number of observations is 8159 (LLM Stuttgart). For each of these local labor markets, we separately compute the parental income gradient, the Q1 and Q5/Q1 measures and the unconditional A-Level share. Again, we continue to rank parents according to their position in the national income distribution.

Differences in Relative Mobility. Figure 4 presents a heat map of relative mobility. Blue areas represent areas of high mobility, whereas red areas indicate a high gradient and thus less mobile regions. While LLMs in the East tend to exhibit lower mobility on average, clusters of high and low mobility are spread out across all of Germany. In some rural local labor markets the parental income gradient ranges below 0.3, whereas in the least mobile areas the gradient exceeds 0.8. In these regions, a child born at the top of the income distribution has a 80 percentage point higher chance to obtain an A-Level degree than a child born at the

¹⁵Conceptually, this is very similar to the approach in Chetty et al. (2014), who use for most children the place of residence of a child at the age of 15 as their main geographical indicator.

Figure 4: Parental Income Gradient across Local Labor Markets

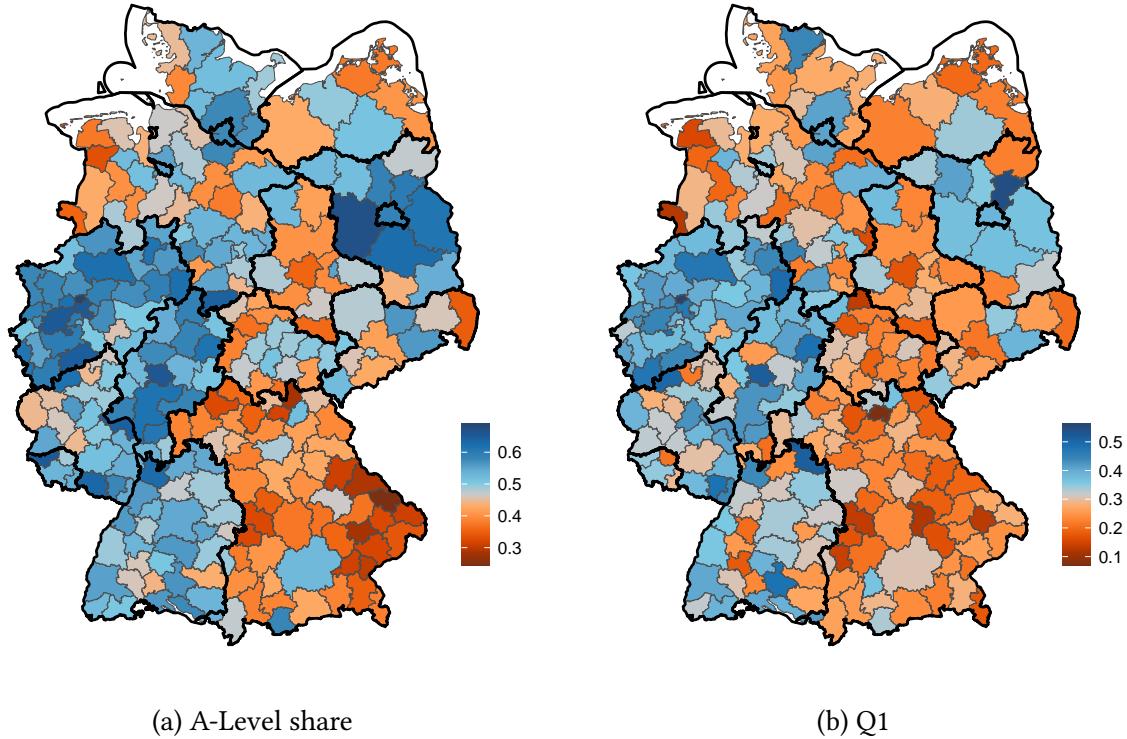


Notes: This figure shows the estimates of the parental income gradient, estimated separately for every LLM in Germany. MZ data, pooled over the period 2011-2018.

bottom of the distribution. The local labor markets with the highest gradient (Lichtenfels) and the lowest gradient (Mühldorf) are both located in Bavaria, highlighting the existence of substantial variation in mobility even within states. Indeed, we find that only 12% of the variation across LLMs can be explained by state differences.¹⁶ The corresponding heatmap for the Q5/Q1 ratio is displayed in Figure B.1 in the appendix.

¹⁶This calculation is based on a regression of the slope in each LLM on a set of state fixed effects. This analysis uses only LLMs that are strict subsets of states, i.e. the 5 LLMs that cross state boundaries are excluded from this analysis. The R-squared of this regression is $R^2 = 0.1218$, meaning that roughly 12% of the variation arises from state-level differences.

Figure 5: A-Level Share and Q1 Measure across Local Labor Markets



Notes: This figure shows the estimates of the A-Level share (panel a) and the Q1 measure (panel b), estimated separately for every LLM in Germany. MZ data, pooled over the period 2011-2018.

Differences in absolute mobility. These large within-state differences are astonishing, as the unconditional A-Level share is clearly clustered at the state level. As illustrated in Figure 5, panel (a), the share of children who eventually obtain an A-Level is systematically higher in Northrhine-Westfalia, Hessia and the city states Berlin, Bremen and Hamburg. Similar patterns are also visible for our measure of absolute mobility in panel (b). Again, red areas indicate regions of low and blue values regions of high mobility. Absolute mobility is systematically lower in Bavaria and systematically higher in Northrhine-Westfalia compared to the German average. Just visually, it hence becomes evident that the unconditional A-Level share is closely linked to absolute but not to relative mobility. Overall, we also find substantial variation in absolute mobility. In some regions, less than 15% of children from the bottom quintile of the national income distribution obtain an A-Level degree, whereas in other regions this number exceeds 50%. We find that 42% of the variation in the Q1 measure and 57% of the variation in the A-Level share can be attributed to state level differences.

Table 5: Mobility in the 15 largest Local Labor Markets

Rank	Name	Population	Gradient	Q5/Q1	Q1	A-Level share
1	Hamburg	2911254	0.46	1.81	0.41	0.58
2	Düsseldorf	1552097	0.47	1.81	0.45	0.65
3	Münster	810377	0.48	1.62	0.47	0.62
4	Gelsenkirchen	1150983	0.50	1.82	0.40	0.57
5	Stuttgart	2522246	0.50	2.16	0.34	0.55
6	Bonn	924546	0.51	1.86	0.44	0.65
7	Duisburg	1170198	0.51	1.86	0.42	0.58
8	Frankfurt/Main	2274907	0.52	1.94	0.43	0.62
9	München	2764232	0.53	2.27	0.31	0.53
10	Dortmund	1159719	0.54	1.93	0.40	0.59
11	Köln	1830947	0.55	2.13	0.38	0.60
12	Hannover	1152675	0.56	2.49	0.30	0.53
13	Berlin	3613495	0.56	2.16	0.39	0.59
14	Nürnberg	1094554	0.60	2.91	0.23	0.43
15	Leipzig	1037782	0.68	3.15	0.25	0.48

Notes: This table reports estimates of absolute and relative mobility for the 15 largest LLMs in Germany (measured by total population in 2017). Local labor markets are sorted in descending order by the relative mobility rank, as measured by the parental income gradient. A higher rank thus indicates higher relative mobility.

Mobility in the Largest Local Labor Markets. Table 5 presents estimates of absolute and relative mobility for the 15 largest local labor markets in Germany. Labor markets are ranked based on the parental income gradient, which varies between 0.46 in Hamburg and 0.68 in the LLM Leipzig. The probability to obtain an A-Level is thus substantially less dependent on parental income for children raised in Hamburg compared to children raised in Leipzig. In terms of absolute mobility, the variation across LLMs is even larger: the probability to obtain an A-Level for a child in the bottom 20% of the income distribution is nearly twice as high in Hamburg (41%) as in Leipzig (25%) or Nürnberg (23%). Table 5 also illustrates that a high degree of absolute mobility does not necessarily imply a high degree of relative mobility. Stuttgart for example, which is ranked 5th in terms of relative mobility would only be ranked as number 11 when considering absolute mobility. The opposite is true for Berlin, which improves from rank 13 to 9 when moving from relative to absolute mobility. Appendix Table C.1 complements this evidence by showing mobility estimates for all German states. Children born in the state of Hamburg face the highest degree of social mobility. With the exception of Bremen, the least mobile states are all located in East Germany.

Table 6: Correlation between Mobility Statistics

Measure	Corr.	A-Level	Q1	Q5/Q1	Gradient
A-Level	ρ	1	-	-	-
	r	1	-	-	-
Q1	ρ	0.75	1	-	-
	r	0.77	1	-	-
Q5/Q1	ρ	-0.33	-0.73	1	-
	r	-0.36	-0.79	1	-
Gradient	ρ	-0.01	-0.46	0.68	1
	r	-0.08	-0.49	0.76	1

Notes: This Table reports the correlation between estimates of different measures of intergenerational mobility across LLM's in Germany. ρ denotes the Pearson correlation coefficient of two measures across LLMs, r denotes the Spearman rank correlation coefficient.

Correlation between the measures. Table 6 compares how these different mobility statistics relate to each other. As expected, the A-Level share is closely correlated with the Q1 measure: in LLMs with a high A-Level share, also children in the bottom quintile have a comparatively high likelihood to complete an A-Level degree. This in turn also means that the Q5/Q1 ratio is negatively correlated with both the A-Level share and the Q1 measure. In contrast, there exists no systematic association between the A-Level share and the parental income gradient, because the gradient is not sensitive to the baseline probability of obtaining an A-Level degree. For this reason, the correlation between the parental income gradient and the Q5/Q1 ratio is – although clearly positive – not close to one. The correlation between the parental income gradient and the Q1 measure is even below 0.5, meaning that a high level of absolute mobility in a given LLM does not always imply a high level of relative mobility. Altogether, these cross-sectional correlations mirror our considerations regarding the time trends in the previous section. While the A-Level share is strongly correlated with the quantile measures, its correlation with the income gradient is negligible.

7.1 Places versus Sorting

What drives these large mobility differences across local labor markets? One explanation are place effects. An active literature argues that places can shape economic outcomes and that place-based policies can help to improve local conditions (Kline and Moretti, 2014; Neumark and Simpson, 2015). However, regional differences can also arise if household systematically

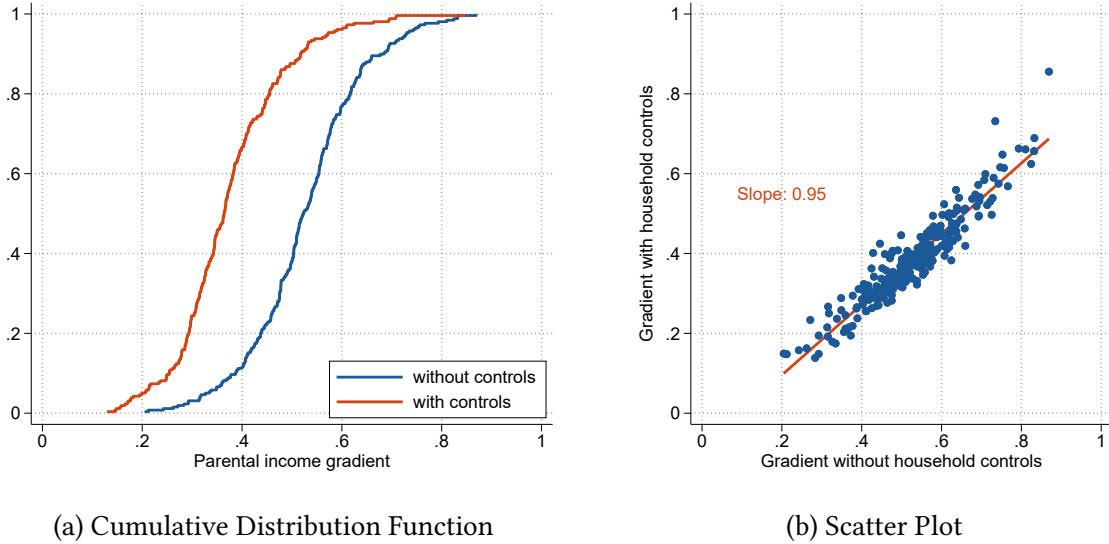
sort into local labor markets. If, for example, (i) parental education differs across areas because households sort into different local labor markets based on their education status and (ii) parental income and education are correlated, spatial mobility differences would arise even in absence of any place-based effects. For the US, Rothbaum (2016) and Gallagher et al. (2018) suggest that a substantial share of the geographical variation in the intergenerational mobility measures reported in Chetty et al. (2014) can be explained by differences in household characteristics across commuting zones. However, this is challenging to verify with the data used by Chetty et al. (2014), as administrative tax data provide only limited information on household characteristics.

Our data, in contrast, contain a rich set of covariates and we can directly test to what extent differences in household characteristics can account for the geographical variation in mobility. To proceed, we recompute the parental income gradient for every local labor market l but control for a comprehensive set of household characteristics X_i , including age and gender of the child, age and marital status of the parents, the number of siblings, a dummy for single parents and the highest parental education level in four categories. This results in the following model:

$$Y_{i,l} = \alpha_l + \beta_l R_l + \gamma X_i + \varepsilon_{i,l} \quad (6)$$

We constrain the effects of additional household characteristics to be equal in all local labor markets, as γ does not have a local-labor market specific subscript. By comparing the conditional parental income gradient ($\beta_l \times 100$) from this equation to the unconditional gradient estimated so far, we can assess the influence of household characteristics on geographical mobility differences. Figure 6, panel (a) plots the Cumulative Distribution Function (CDF) of both gradients across all local labor markets. The CDF of the unconditional gradient first order stochastically dominates the CDF of the conditional gradient, which means that in all local labor markets, the parental income gradient is reduced when controlling for household characteristics. This is intuitive, as especially controlling for parental education reduces the association between our outcome of interest and the parental income rank (compare Figure 2). At the same time, the variance of the distribution (second moment) is not reduced when moving from the unconditional to the conditional slope and, critically, the ranking between local labor markets is largely unaffected: Figure 6, panel (a) shows a scatter plot of the unconditional and the conditional gradients. The estimated linear fit is 0.95, demonstrating that controlling for

Figure 6: Adjustment for Household Characteristics



Notes: Panel (a) plots the Cumulative Distribution Function (CDF) for the parental income gradient estimated with and without household controls across local labor markets in Germany. Panel (b) shows a scatter plot of LLM-specific parental income gradients estimated when not controlling for additional household characteristics (x-Axis) and LLM-specific Parental Income Gradients estimated when controlling for additional household characteristics (y-Axis).

household characteristics affects the estimates in most local labor markets equally. The correlation coefficient is 0.91 and the Spearman rank correlation amounts to 0.89. We conclude that differences in household characteristics seem to explain only a small fraction in mobility differences between local labor markets.

7.2 Place Based Predictors of Mobility

As household characteristics cannot account for the large spatial variation in mobility, it must be that place based characteristics play an important role in shaping these disparities. While our setting does not allow to identify the causal determinants of these differences, we can examine which regional characteristics are most predictive for spatial differences in mobility. This can give a first hint which factors may matter for mobility and guide future research. For this aim, we collect a comprehensive database of 71 regional indicators with information on labor market participation, economic conditions, infrastructure, demographics, housing and living conditions, the education structure and social characteristics. Appendix D provides more details on the regional indicators.

To study the association between local characteristics and intergenerational mobility, prior articles investigating the geography of intergenerational mobility have typically relied on bi-

ivariate correlations or estimated multiple linear models via OLS (Chetty et al., 2014; Corak, 2020). Both approaches have clear disadvantages. As socio-economic characteristics are highly correlated at the regional level, correlation coefficients are often spurious and may add little information. While this remedy is overcome in a multiple linear OLS regression, these models are prone to heavily overfit the data in high-dimensional data sets (Babyak 2004). This may hamper the generalizability of the estimates. To address this, it is necessary to achieve dimensionality reduction via some form of variable selection. Belloni and Chernozhukov (2013) suggest to preselect covariates via Lasso before estimating a multiple linear model. This approach has for example been applied by Finkelstein et al. (2016) to explain geographical variation in health care utilization in the US.

We take a similar two-step approach, but preselect variables using a Random Forest variable importance measure instead of preselecting variables with a Lasso regression. This is because we find that a linear Lasso model only poorly fits our data due to non-linear interactions between the variables.¹⁷ The Random Forest algorithm in contrast is fully non-parametric and can capture higher-order interactions in the data. After fitting the Random Forest, we can rank covariates according to their predictive power and thus obtain a measure of variable importance.¹⁸

Variable Selection. In a first step, we thus fit a Random Forest based on all 71 regional indicators. We then compute a measure of variable importance and rank predictors accordingly. The 15 most informative predictors are displayed in Table 7.¹⁹ Overall, mainly social characteristics, the local organization of the education system and labor market conditions seem to matter for mobility. The column to the right displays the sign of the bivariate correlation between each variable and the parental income gradient. A positive sign indicates that the

¹⁷To compare the out-of-sample performance of this algorithm against an implementation of a Lasso and an Elastic Net regression with $\alpha = 0.5$, we split our data in a training and test data set (75-25 split). The Random Forest algorithm predicts 38% of the variation in the test sample ($R^2 = 0.38$), whereas the predictive power of Lasso ($R^2 = 0.11$) and Elastic Net ($R^2 = 0.17$) is much lower. The results for Lasso and Elastic Net are based on λ chosen by 5-fold cross-validation. For the Random Forest, we fit 1000 trees and randomly select $72/3 = 24$ variables for each split.

¹⁸There are several ways to compute a Random Forest based measure of variable importance. We choose the implementation proposed by Strobl et al. (2008), which computes a conditional permutation importance measure that accounts for the dependence structure between the predictors. The Random Forest variable importance measure used here is hence conceptually similar to a variable importance ranking provided by Lasso - with the addition that it is fully flexible to account for non-linear interactions between covariates.

¹⁹The exact ranking of predictors especially after rank 5 varies for different implementations of the Random Forest algorithm. We are therefore cautious not to over-interpret the ranking between single predictors. As described above, this ranking only serves to reduce dimensionality of our covariates.

Table 7: The 15 Most Predictive Regional Indicators

Rank	Variable	Importance Measure	ρ
1	school dropout rate	1.043e-07	+
2	share married	5.349e-08	-
3	students	4.551e-08	-
4	teenage pregnancies	2.615e-08	+
5	broadband availability	1.715e-08	+
6	distance to next college	1.465e-08	-
7	median income vocational qualification	1.225e-08	-
8	unemployment rate	1.207e-08	+
9	gender wage gap	1.204e-08	+
10	share without vocational qualification	9.343e-09	-
11	Gini parental income	6.757e-09	-
12	share on vocational A-Level track	6.356e-09	-
13	voter turnout	5.134e-09	-
14	child poverty	4.061e-09	+
15	share children 0-2 in childcare	4.052e-09	+

Notes: This table shows the 15 most predictive indicators for explaining variation in the parental income gradient across local labor markets in Germany. See text for details on the implementation via a Random Forest variable importance measure. The last column shows the sign of the bivariate correlation between each variable and the parental income gradient.

indicator is predictive for low mobility (a high gradient). This is true for the school dropout rate, which is by far the most informative indicator: LLMs with a high prevalence of school dropouts are associated with low relative mobility. The same applies for teenage pregnancies, child poverty and the share of individuals which are dependent on social assistance. All these indicators point to comparatively disadvantaged social contexts in these labor markets, which is consistent with social capital based explanations of regional disparities in mobility. A high share of married individuals in contrast is a signal for high mobility. Other variables like the access to broadband Internet or the distance to the next elementary school are less straightforward to interpret, but apparently predictive for mobility.

Regression Estimates. In a second step, we regress the gradient on these 15 most predictive variables in a multiple linear model. All right-hand side variables are standardized so that the coefficients report the association between a one standard deviation change in the covariate and the absolute change in the gradient. The results are reported in Table 8. The signs of the coefficients generally match those from the bivariate correlations in Table 7. For example,

Table 8: Intergenerational Mobility and Regional Characteristics

	(1)	(2)	(3)	(4)
school dropout rate	0.0450*** (0.0112)	0.0442*** (0.0114)	0.0550*** (0.0158)	0.0543*** (0.0158)
share married	-0.0195*** (0.0084)	-0.0267*** (0.0085)	-0.0165 (0.0102)	-0.0206*** (0.0103)
students	-0.0228* (0.0128)	-0.0248* (0.0126)	-0.0271* (0.0160)	-0.0301* (0.0159)
teenage pregnancies	0.0305 (0.0234)	0.0266 (0.0233)	0.0249 (0.0261)	0.0194 (0.0271)
broadband availability	0.0218*** (0.0108)	0.0236*** (0.0107)	0.0203* (0.0111)	0.0230*** (0.0112)
distance to next college	-0.0043 (0.0074)	-0.0063 (0.0078)	-0.0048 (0.0075)	-0.0079 (0.0080)
median income vocational qualification	-0.0168 (0.0114)	-0.0144 (0.0115)	-0.0171 (0.0142)	-0.0138 (0.0151)
unemployment rate	0.0168 (0.0327)	0.0144 (0.0324)	0.0271 (0.0416)	0.0201 (0.0417)
gender wage gap	-0.0045 (0.0143)	-0.0065 (0.0144)	0.0061 (0.0177)	0.0005 (0.0182)
share without vocational qualification	0.0062 (0.0162)	0.0067 (0.0162)	0.0176 (0.0220)	0.0158 (0.0222)
Gini parental income	-0.0222 (0.0143)	-0.0187 (0.0144)	-0.0051 (0.0196)	0.0013 (0.0205)
share on vocational A-Level track	-0.0159* (0.0093)	-0.0160* (0.0091)	-0.0201* (0.0102)	-0.0207*** (0.0101)
voter turnout	0.0199 (0.0122)	0.0228* (0.0123)	0.0254 (0.0168)	0.0314* (0.0172)
child poverty	-0.0415 (0.0319)	-0.0260 (0.0332)	-0.0496 (0.0385)	-0.0223 (0.0444)
share children 0-2 in childcare	-0.0440*** (0.0183)	-0.0430*** (0.0181)	-0.0351 (0.0261)	-0.0396 (0.0261)
Additional controls	-	✓	-	✓
State Fixed Effects	-	-	✓	✓
<i>N</i>	258	258	252	252
<i>R</i> ²	0.259	0.279	0.293	0.306

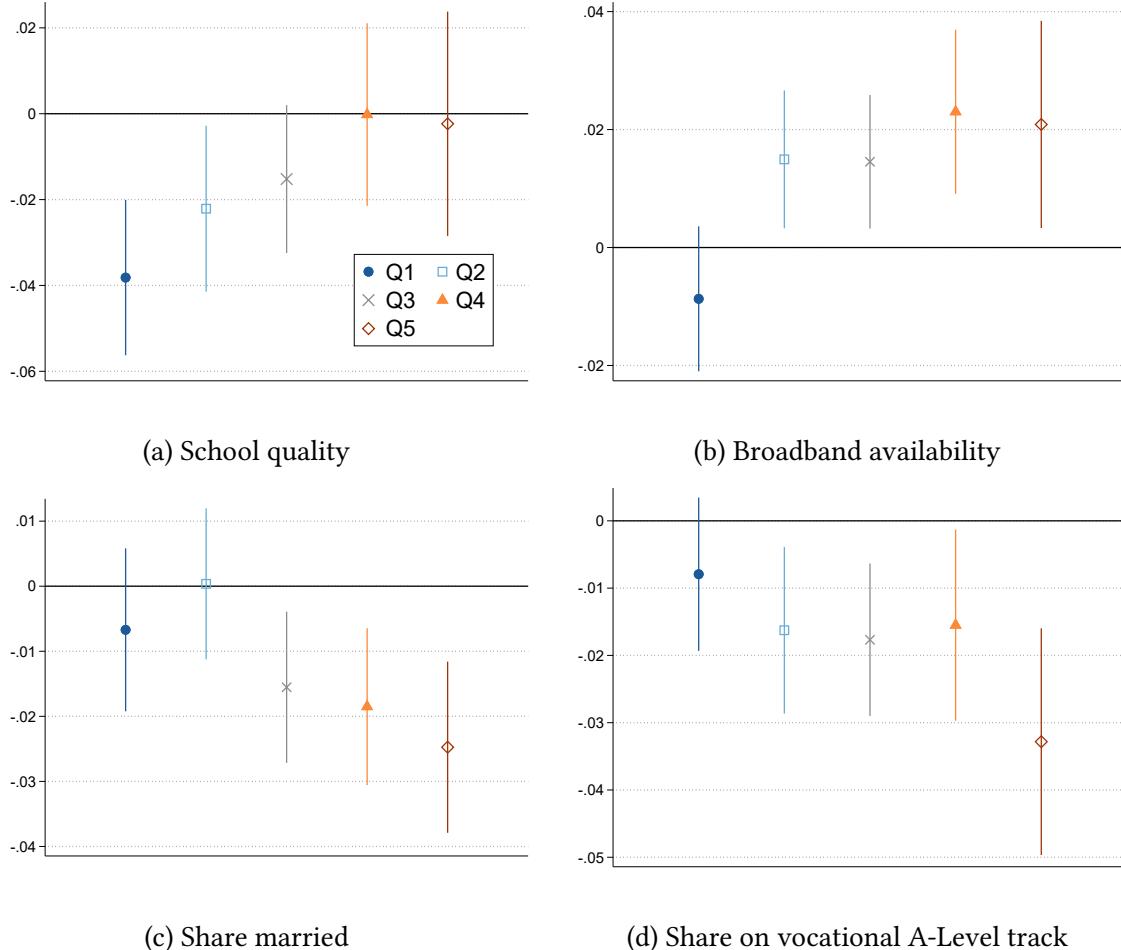
Notes: Each column of this table reports coefficients from a linear regression with robust standard errors reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable in all columns is the parental income gradient. The independent variables are those selected by the Random Forest (see text for details) and are standardized to have mean 0 and standard deviation 1. Columns (3) and (4) contain state fixed effects, for which we have to drop the five LLMs crossing state borders and the LLM of Berlin. In columns (2) and (4), we additionally control for population, population density and the region type (rural, urban or mixed) to test if coefficients of the regional indicators are affected by structural differences in mobility between more rural or urban LLMs.

a one standard deviation increase in the school dropout rate is associated with a 4.5 percentage point higher parental income gradient. This association becomes even stronger adding state fixed effects. A high gradient also aligns with a high number of teenage pregnancies, enhanced child poverty, a higher unemployment rate, a high voter turnout and a high share of households with broadband Internet access. A negative association with the parental income gradient arises for the share of married individuals, the distance to the next college, the median income for individuals with a recognized vocational qualification, the share of children on a vocational A-Level track and for the share of children aged 0-2 in childcare. However, due to the limited sample size, we lack the power to precisely estimate most coefficients. Exceptions are the school dropout rate and the broadband availability which are both associated with lower mobility as well as the share of married individuals and the share of children on a vocational A-Level track, which are associated with higher mobility.

Graphical Evidence. To better understand the relationship between relative mobility and these statistically significant indicators, we separately regress the A-Level share in each quintile of the parental income distribution on each indicator and plot the estimates in Figure 7. From this graphical representation we can learn whether for example a positive association between the parental income gradient and an indicator is driven by a lower A-Level share of children from low-income households or by a higher A-Level share of children from high-income households.

We start with the school dropout rate. In the US context, Chetty et al. (2014) interpret the school dropout rate – adjusted by parental income – as an indicator of school quality and find a strong negative correlation with mobility. In close analogy, we regress the dropout rate on mean parental income, the Gini coefficient of parental income, the share of parents holding an A-Level and the unemployment rate and take the residuals to obtain a measure of school quality which is adjusted for parental income and labor market conditions. This indicator is still highly correlated with mobility. As depicted in Figure 7, panel (a), low school quality (a high value of the indicator) is associated with a lower probability to obtain an A-Level degree for children from low income households but does not seem to affect children in the top two quintiles of the parental income distribution. While this would be consistent with the

Figure 7: Predicting the A-Level Share by Parental Income Quintile



Notes: Each panel of this Figure reports coefficients from five separate linear OLS regressions with robust standard errors and 95% confidence bands. The dependent variable is the share of children which obtained an A-Level in the respective quintile of the parental income distribution. The independent variable is the school quality index (see text for details) in panel (a), the share of broadband connections per 100 inhabitants in panel (b), the share of married individuals in panel (c) and the share of students on a vocational (rather than general education) A-Level track (d). In addition, all regression include a set of state fixed effects and control for population, population density and the region type (rural, urban or mixed). Due to the inclusion of state fixed effects, we have to drop the five LLMs crossing state borders and the LLM of Berlin from the sample ($N=252$). All regressors are standardized to have mean 0 and standard deviation 1.

hypothesis that school quality is crucial in enhancing mobility, further information is needed to test this hypothesis in detail.²⁰

²⁰Most importantly, it remains open if the adjusted school drop out rate is indeed an appropriate proxy for school quality. In the US, Rothstein (2019) studies how closely the transmission of parental income to educational attainment and achievement (test scores) are correlated with income mobility at the commuting zone level. He finds income-income transmission to be closely connected to income-educational attainment transmission, but not to income-educational achievement transmission. Rothstein (2019) therefore finds little evidence that differences in the quality of secondary schooling are a key mechanism driving variation in intergenerational mobility, despite the high correlation of the school quality indicator in Chetty et al. (2014). However, the very different nature of the German secondary schooling system could lead to very different patterns in our data. Unfortunately, there exist no comparable data on student test scores in Germany, preventing us from investigating this issue further.

Panel (b) sheds light on the negative association between broadband availability and mobility. While broadband access is in general associated with a higher A-Level share, this is not true for children in Q1, for whom the relationship becomes negative. We can only speculate what is going on here. Broadband access is highly correlated with a number of factors that indicate dynamic and prosperous labor markets which exhibit above average inequality. For example, broadband access is associated with a high share of business creations and highly qualified individuals, but also with high levels of income inequality and a high share of people without any vocational degree. This suggests that broadband availability proxies urban environments in which most children profit from a dynamic and rewarding economic environment, but the kids at the bottom fail to take part in this development. However, it is also possible that there exists a genuine causal effect of broadband availability on social mobility. For the US, Dettling et al. (2018) document that increased broadband availability fosters access to college and find the effect to be concentrated among students with parents from high socio-economic status. Similarly, Sanchis-Guarner et al. (2021) report a causal (positive) impact of broadband access on student test scores in England but find comparatively lower effects for students eligible for free school meals. The correlation presented here would be in line with these findings.

The opposite pattern applies for the share of married individuals in panel (c): this statistic is related to higher mobility but a lower level of children obtaining an A-Level degree in general. Finally, panel (d) reports the association between the Q-measures and the share of children on a vocational, rather than general interest, A-Level track. There is reason to believe that the availability of such vocational tracks may dampen the influence of parental background for the opportunities of children. Typically, children in these tracks have obtained a degree from the medium track (Realschule) and now attend a specialized vocational school to obtain an A-Level degree on top. In that setting, vocational schools may especially foster the opportunities of children from low-income households initially "misallocated" into the the medium instead of the high track. Dustmann et al. (2017) show that vocational schools have the potential to fully offset adverse affects of early age tracking on long-term labor market outcomes, but cannot observe parental background.

Our correlational evidence shows that children from the bottom quintile are more likely relative to children from the top quintile to obtain an A-Level in local labor markets with a high prevalence of such schools. In addition, we find that at the national level the parental income rank is more predictive for the probability to attend the general high track (Gymnasium) at

the age of 13-14 than to obtain an A-Level degree later on (gradient of 0.55 versus 0.52), again suggesting that vocational schools may mediate the influence of parental background.

8 Conclusion

In this paper, we use census data to characterize social mobility in Germany. We measure mobility by the association between a child's probability of obtaining an A-Level degree and her parents income rank in the parental income distribution. At the national level, a 10 percentile increase in parent income is associated with a 5.2 percentage point increase in the A-Level share. This relationship remained stable for recent cohorts, despite a massive roll-out of higher secondary education in the period 1997-2018. This suggests that an expansion in access to higher education alone may not be sufficient to reduce the opportunity gap between children from high and low income households.

At the same time, mobility varies substantially at the regional level. Most of this spatial variance arises within states, indicating that the factors shaping mobility seem to be of very local nature. We show that sorting of households into different regions can only account for a small fraction of these differences but that instead place-based characteristics seem to matter more for mobility. We find that especially social characteristics, the local organization of the education system and labor market conditions are predictive for mobility differences. More research is needed to really understand structural determinants of mobility at the local level. The mobility statistics constructed here could serve as a starting point for this kind of analysis, as they provide variation in both time and space. Arguably exogenous events like state-level reforms of the education system or labor market specific shocks could be used to shed light on the determinants of mobility at the regional level.

The approach described in this paper provides a timely and feasible way to monitor the development of social mobility in Germany for recent cohorts. Future work could also investigate more closely the nexus between our mobility estimates and traditional measures like rank-rank correlations in income. Beyond the German setting, our approach to measure mobility could turn out useful in all countries where the highest secondary school degree has comparable importance than the Abitur in Germany. Many European countries have an education system that more closely resembles the German than the US one, in the sense that there exist a degree broadly comparable to the Abitur which is crucial for future career options. Ex-

amples are Italy (Maturità), Austria (Matura) and the UK (A-Level). If similar census data were available also in these countries, it would be interesting to see how estimates compare to the findings provided in this paper. Only if we have gathered enough reliable estimates of mobility, we can really move forward in trying to understand how to enhance social mobility in our societies.

References

- ACCIARI, P., A. POLO, AND G. L. VIOLANTE (2019): ““And Yet, It Moves”: Intergenerational Mobility in Italy,” NBER Working Paper No. 25732.
- ALESINA, A., S. HOHMANN, S. MICHALOPOULOS, AND E. PAPAIOANNOU (2021): “Intergenerational mobility in Africa,” *Econometrica*, 89(1), 1–35.
- ASHER, S., P. NOVOSAD, AND C. RAFKIN (2020): “Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities,” Discussion paper.
- ASHWORTH, J., AND T. RANSOM (2019): “Has the college wage premium continued to rise? Evidence from multiple US surveys,” *Economics of Education Review*, 69, 149–154.
- BABYAK, M. A. (2004): “What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models,” *Psychosomatic medicine*, 66(3), 411–421.
- BELLONI, A., AND V. CHERNOZHUKOV (2013): “Least squares after model selection in high-dimensional sparse models,” *Bernoulli*, 19(2), 521–547.
- BIEWEN, M., AND M. TAPALAGA (2017): “Life-Cycle Educational Choices in a System with Early Tracking and ‘Second Chance’ Options,” *Economics of Education Review*, 56, 80–94.
- BRATBERG, E., J. DAVIS, B. MAZUMDER, M. NYBOM, D. SCHNITZLEIN, AND K. VAAGE (2017): “A Comparison of Intergenerational Mobility Curve in Germany, Norway, Sweden, and the US,” *Scandinavian Journal of Economics*, 119(1), 72–101.
- CHETTY, R., AND N. HENDREN (2018a): “The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects,” *The Quarterly Journal of Economics*, 133(3), 1107–1162.
- (2018b): “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates,” *The Quarterly Journal of Economics*, 133(3), 1163–1228.
- CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014): “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *The Quarterly Journal of Economics*, 129(4), 1553–1623.

- CHUARD, P., AND V. GRASSI (2020): “Switzer-Land of Opportunity: Intergenerational Income Mobility in the Land of Vocational Education,” University of St. Gallen Economics Working Paper Series 2020-11.
- CORAK, M. (2020): “The Canadian geography of intergenerational income mobility,” *The Economic Journal*, 130(631), 2134–2174.
- DAHL, M., AND T. DELEIRE (2008): *The Association between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data*. University of Wisconsin-Madison, Institute for Research on Poverty.
- DETTLING, L. J., S. GOODMAN, AND J. SMITH (2018): “Every little bit counts: The impact of high-speed internet on the transition to college,” *Review of Economics and Statistics*, 100(2), 260–273.
- DEUTSCHER, N., AND B. MAZUMDER (2020): “Intergenerational mobility across Australia and the stability of regional estimates,” *Labour Economics*, 66, 101861.
- DUSTMANN, C., P. PUHANI, AND U. SCHÖNBERG (2017): “The Long-Term Effects of Early Track Choice,” *The Economic Journal*, 127, 1348–1380.
- EBERHARTER, V. V. (2013): “The Intergenerational Dynamics of Social Inequality—Empirical Evidence from Europe and the United States,” SOEPpaper No. 588.
- EISENHAUER, P., AND F. PFEIFFER (2008): “Assessing intergenerational earnings persistence among German workers,” *Journal of Labour Market Research*, (2-3), 119–137.
- FINKELSTEIN, A., M. GENTZKOW, AND H. WILLIAMS (2016): “Sources of geographic variation in health care: Evidence from patient migration,” *The Quarterly Journal of Economics*, 131(4), 1681–1726.
- GALLAGHER, R., R. KAESTNER, AND J. PERSKY (2018): “The Geography of Family Differences and Intergenerational Mobility,” *Journal of Economic Geography*.
- GÄRTNER, K. (2002): “Differentielle Sterblichkeit. Ergebnisse des Lebenserwartungssurvey des BiB,” *Zeitschrift für Bevölkerungswissenschaft*, 27(2), 185 – 211.
- HAIDER, S., AND G. SOLON (2006): “Life-cycle variation in the association between current and lifetime earnings,” *American Economic Review*, 96(4), 1308–1320.

HAUSNER, K. H., D. SÖHNLEIN, B. WEBER, AND E. WEBER (2015): “Qualifikation und Arbeitsmarkt: bessere Chancen mit mehr Bildung,” IAB-Kurzbericht.

HILGER, N. G. (2015): “The Great Escape: Intergenerational Mobility in the United States Since 1940,” NBER Working Paper No. 21217.

KLEIN, M., K. BARG, AND M. KÜHHIRT (2019): “Inequality of Educational Opportunity in East and West Germany: Convergence or Continued Differences?,” *Sociological Science*, 6, 1–26.

KLINE, P., AND E. MORETTI (2014): “People, places, and public policy: Some simple welfare economics of local economic development programs,” *Annual Review of Economics*, 6(1), 629–662.

LEE, C.-I., AND G. SOLON (2009): “Trends in intergenerational income mobility,” *The Review of Economics and Statistics*, 91(4), 766–772.

LOCHNER, L. (2011): “Nonproduction benefits of education: Crime, health, and good citizenship,” *Handbook of the Economics of Education*, 4, 183–282.

MAZUMDER, B. (2005): “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data,” *The Review of Economics and Statistics*, 87(2), 235–255.

——— (2016): “Estimating the Intergenerational Elasticity and Rank Association in the U.S.: Overcoming the Current Limitations of Tax Data,” *Research in Labor Economics*, 43, 83–129.

——— (2018): “Intergenerational mobility in the United States: What we have learned from the PSID,” *The Annals of the American Academy of Political and Social Science*, 680(1), 213–234.

MUÑOZ, E. (2021): “The Geography of Intergenerational Mobilityin Latin America and the Caribbean,” Discussion paper.

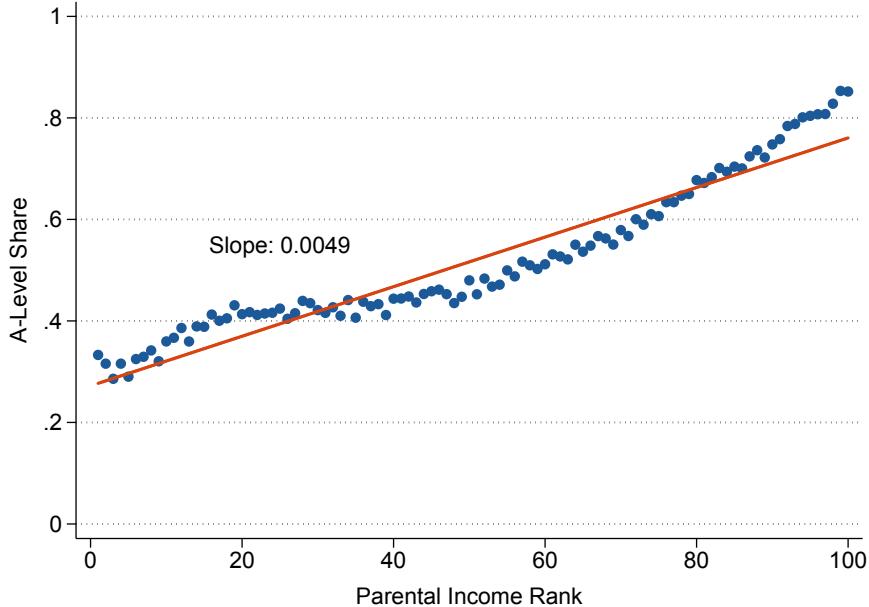
NEUMARK, D., AND H. SIMPSON (2015): “Place-based policies,” in *Handbook of Regional and Urban Economics*, vol. 5, pp. 1197–1287. Elsevier.

NYBOM, M., AND J. STUHLER (2017): “Biases in standard measures of intergenerational income dependence,” *Journal of Human Resources*, 52(3), 800–825.

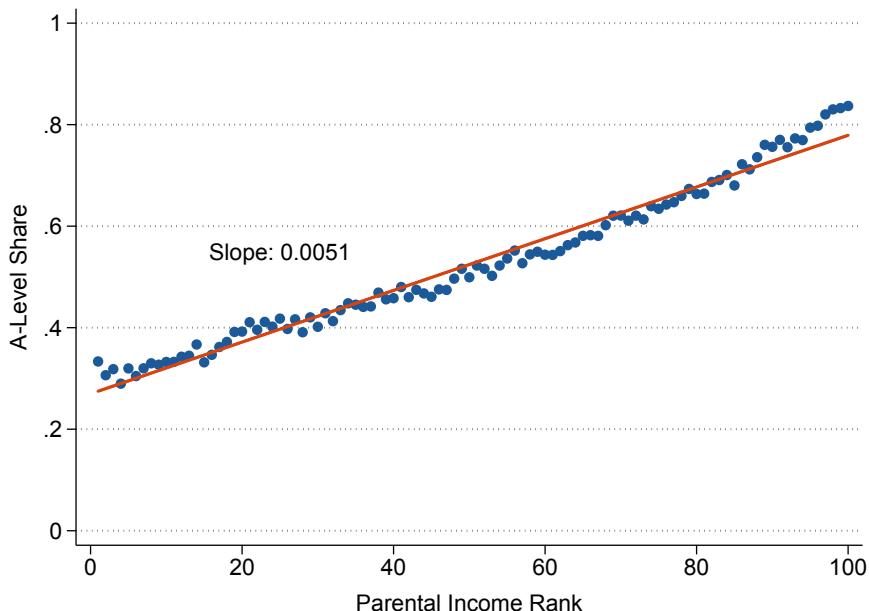
- OREOPOULOS, P., AND K. G. SALVANES (2011): “Priceless: The nonpecuniary benefits of schooling,” *Journal of Economic perspectives*, 25(1), 159–84.
- RIPHAHN, R., AND G. HEINECK (2009): “Intergenerational Transmission of Educational Attainment in Germany – The Last Five Decades,” *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 229(1), 36–60.
- RIPHAHN, R., AND P. TRÜBSWETTER (2013): “The Intergenerational Transmission of Educational Attainment in East and West Germany,” *Applied Economics*, 45(22), 3183–3196.
- ROTHBAUM, J. (2016): “Sorting and geographic variation in intergenerational mobility,” *Unpublished Manuscript*.
- ROTHSTEIN, J. (2019): “Inequality of educational opportunity? Schools as mediators of the intergenerational transmission of income,” *Journal of Labor Economics*, 37(S1), S85–S123.
- SANCHIS-GUARNER, R., J. MONTALBÁN, AND F. WEINHARDT (2021): “Home Broadband and Human Capital Formation,” *CESifo Working Paper*.
- SCHMILLEN, A., AND H. STÜBER (2014): “Lebensverdienste nach Qualifikation - Bildung lohnt sich ein Leben lang,” Discussion paper.
- SCHNITZLEIN, D. (2016): “A New Look at Intergenerational Mobility in Germany Compared to the U.S,” *The Review of Income and Wealth*, 62(4), 650–667.
- SOLON, G. (1992): “Intergenerational Income Mobility in the United States,” *American Economic Review*, 82(3), 393–408.
- STROBL, C., A.-L. BOULESTEIX, T. KNEIB, T. AUGUSTIN, AND A. ZEILEIS (2008): “Conditional variable importance for random forests,” *BMC bioinformatics*, 9(1), 307.
- ZIMMERMAN, D. J. (1992): “Regression Toward Mediocrity in Economic Stature,” *American Economic Review*, 82(3), 409–429.

A Alternative Equivalence Scales

Figure A.1: National estimates with different equivalization scheme



(a) no adjustment

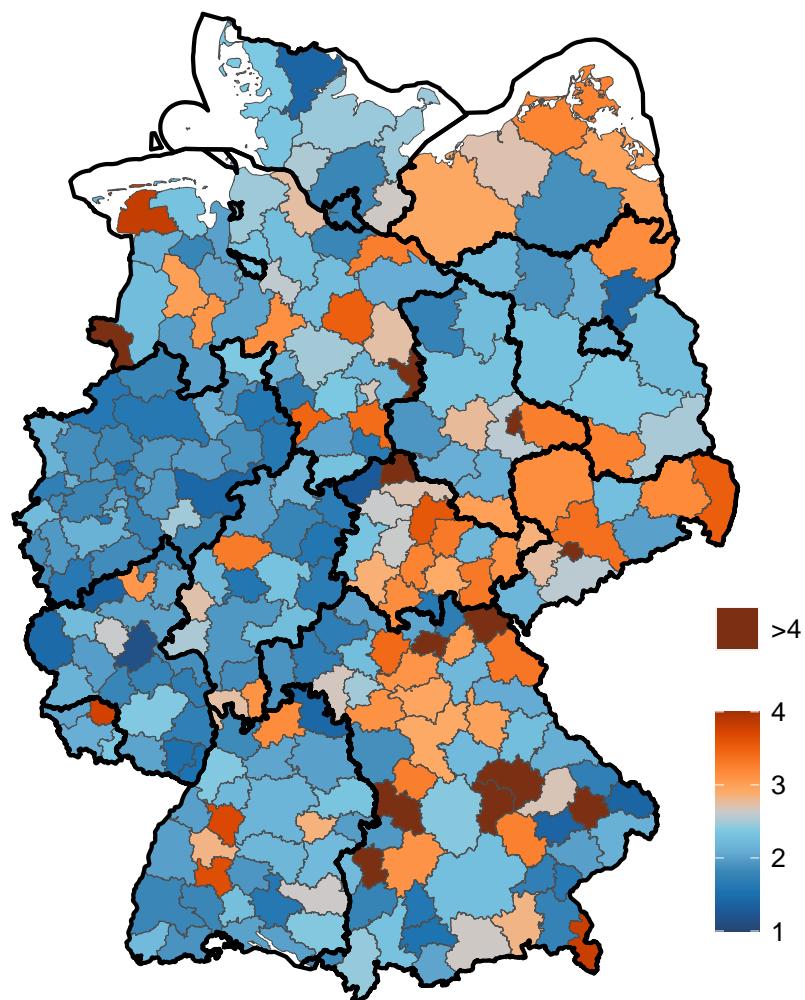


(b) per capita adjustment

Notes: This figure shows non-parametric binned scatter plots of the fraction of children aged 17-21 that are either enrolled in the last two/three years of the A-Level track or already completed the A-Level degree by percentile rank of their parents in the national income distribution in the period 2011-2018. In panel (a), parental income is not adjusted for household size, whereas in panel (b) we adjust income by dividing through the household size. The parental percentile rank is computed within each child cohort and the OLS slopes reported in the figure are estimated using the underlying micro data.

B Heatmap of the Q5/Q1 ratio

Figure B.1: Q5/Q1 ratio across Local Labor Markets



Notes: This figure shows the estimates of the Q5/Q1 ratio, estimated separately for every LLM in Germany. MZ data, pooled over the period 2011-2018.

C Mobility at the State Level

Table C.1: Mobility at the State Level

Rank	Name	Gradient	Q5/Q1	Q1	A-Level share
1	Hamburg	0.45	1.71	0.43	0.60
2	Rheinland-Pfalz	0.50	1.93	0.36	0.53
3	Nordrhein-Westfalen	0.51	1.88	0.41	0.59
4	Hessen	0.52	2.03	0.39	0.59
5	Baden-Württemberg	0.52	2.16	0.34	0.53
6	Saarland	0.53	2.23	0.33	0.54
7	Schleswig-Holstein	0.53	2.16	0.32	0.52
8	Niedersachsen	0.54	2.44	0.29	0.48
9	Bayern	0.55	2.57	0.25	0.42
10	Berlin	0.56	2.16	0.39	0.59
11	Brandenburg	0.57	2.32	0.36	0.60
12	Sachsen-Anhalt	0.57	2.44	0.25	0.43
13	Sachsen	0.61	2.83	0.27	0.48
14	Mecklenburg-Vorpommern	0.63	2.72	0.25	0.45
15	Bremen	0.64	2.51	0.32	0.55
16	Thüringen	0.64	2.86	0.25	0.46

Notes: This table reports estimates of absolute and relative mobility for all 16 German states. States are sorted in descending order by the relative mobility rank, as measured by the parental income gradient. A higher rank thus indicates higher relative mobility.

D Details on Regional Indicators

Table D.1: List of Regional Indicators

Category	Variable	Data source
Labor market	unemployment rate	INKAR
	share long term unemployed	INKAR
	share female employees	INKAR
	share part time employed	INKAR
	share without vocational qualification	INKAR
	share marginal employment	INKAR
	share employed in manufacturing sector	INKAR
	apprenticeship positions	INKAR
	apprentices	INKAR
	vocational school students	INKAR
	employees with academic degree	INKAR
	commuting balance	INKAR
	hours worked	INKAR
	A-Level wage premium	MZ
Education	students (before tertiary education)	INKAR
	students (all universities and colleges)	INKAR
	students (only universities of applied sciences)	INKAR
	school dropout rate	INKAR
	highly qualified persons	INKAR
	share children 0-2 in childcare	INKAR
	share children 3-5 in childcare	INKAR
	share enrolled in Gymnasium	Destatis
	distance to next college	HRK
	distance to next elementary school	INKAR
	share on vocational A-Level track	MZ
Income	share of parents with an A-Level	MZ
	median household income	INKAR
	median household income with vocational qualification	INKAR
	gender wage gap	INKAR
	child poverty	INKAR
	mean household income	INKAR
	Gini household income	MZ
	mean parental income	MZ
	Gini parental income	MZ
	ratio p85/p50 (household income)	MZ
	ratio p50/p15 (household income)	MZ

Economy	GDP per capita	INKAR
	municipal tax revenues per capita	INKAR
	municipal debt per capita	INKAR
	business creation	INKAR
Housing	construction land prices	INKAR
	new apartments	INKAR
	building permits	INKAR
	living area	INKAR
	share of apartment buildings	INKAR
	rent prices	INKAR
Infrastructure	physician density	INKAR
	broad band availability	INKAR
	passenger car density	INKAR
	hospital beds	INKAR
Demographics	average age	INKAR
	share female	INKAR
	share foreigners	INKAR
	share asylum seekers	INKAR
	total net migration	INKAR
	births net of deaths	INKAR
	fertility rate	INKAR
	teenage pregnancies	INKAR
	life expectancy	INKAR
	child mortality	INKAR
	population density	INKAR
	share single parents	MZ
	share married	MZ
	share divorced	MZ
Social	voter turnout	INKAR
	vote share CDU	INKAR
	vote share SPD	INKAR
	share social assistance	INKAR
	mean ISEI	MZ
	Gini ISEI	MZ

Notes: This Table displays all regional indicators considered for our analysis. The third column reports the data source, which is either the INKAR database (see description below), the German statistical office (Destatis), the Hochschulrektorenkonferenz (HRK) or the Mikrozensus (MZ)

Table D.1 displays all 71 regional indicators we use as predictors in the Random Forest algorithm. In a first step, we retrieve data from Federal Institute for Building, Urban Affairs and Spatial Research (BBSR), which maintains the INKAR database of regional indicators (<https://www.inkar.de/>). These data are collected from various government bodies

in Germany, including the German Statistical Office Destatis and the Institute for Employment Research IAB. We select all indicators which may potentially be relevant for mobility and are not clearly collinear to each other (for example, it would not make sense to include the unemployment rates among males and females at the same time). In a second step, we add data from Destatis publications with information on the share of Gymnasium students among all students and compute the distance of the geographical center of each LLM to the next college based on data from the Website of the Hochschulrektorenkonferenz (HRK; <https://www.hochschulkompass.de/hochschulen/downloads.html>). In a third step, we compute additional statistics on the LLM level in the MZ data, like the Gini coefficient in household income, the local A-Level wage premium or the ISEI (an international index of social status). We construct our final indicators as the time averages of variables over the years 2011 to 2018 at the local labor market level.