I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14033

# The Effects of Free Secondary School Track Choice: A Disaggregated Synthetic Control Approach

Aderonke Osikominu
Gregor Pfeifer
Kristina Strohmaier

JANUARY 2021

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14033

# The Effects of Free Secondary School Track Choice: A Disaggregated Synthetic Control Approach

**Aderonke Osikominu**
*University of Hohenheim, CEPR, CESifo and IZA*

**Gregor Pfeifer**
*University of Sydney and University of Hohenheim, CESifo and IZA*

**Kristina Strohmaier**
*University of Tuebingen*

JANUARY 2021



ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

## The Effects of Free Secondary School Track Choice: A Disaggregated Synthetic Control Approach*

We exploit a recent state-level reform in Germany that granted parents the right to decide on the highest secondary school track suitable for their child, changing the purpose of the primary teacher's recommendation from mandatory to informational. Applying a disaggregated synthetic control approach to administrative district-level data, we find that transition rates to the higher school tracks increased substantially, with stronger responses among children from richer districts. Simultaneously, grade repetition in the first grades of secondary school increased dramatically, suggesting that parents choose school tracks also to align with their own aspirations – resulting in greater misallocation of students.

**JEL Classification:** C21, C46, I21, I28, J24

**Keywords:** school tracking, student performance, synthetic control method, treatment effect distributions, treatment effect heterogeneity

**Corresponding author:**
Kristina Strohmaier
University of Tuebingen
Department of Economics
Nauklerstrasse 47
72074 Tuebingen
Germany

E-mail: k.strohmaier@uni-tuebingen.de

* We thank Stefan Bauernschuster, Christian Dustmann, Nadja Dwenger, Bernd Fitzenberger, Jean Hindriks, Stefan Hoderlein, Patrick Puhani, Nadine Riedel, Dominik Sachs, Guido Schwerdt, and Stephan Thomsen, as well as seminar and conference participants at Ifo, ZEW, SSES, IAAE, IIPF, EALE, VfS, LAGV, the AEA, Berlin, Passau, and CReAM/UCL for valuable remarks. Financial support from the German Research Foundation (DFG grant no. PF-912/1-1) is gratefully acknowledged.

# 1 Introduction

While education systems in many advanced countries incorporate some form of school tracking (i.e., the allocation of students to different school types according to their abilities), the general trend in the last decades has been to postpone the tracking towards the end of secondary education, at ages 14 to 16 (e.g., Leschinsky and Mayer, 1990; Betts, 2011).[1] The main concern associated with an early-tracking system is insufficient information on a child's academic potential at young ages. In fact, some children may still experience a developmental surge at later ages. Therefore, early tracking may lead to a misallocation of students to school tracks. Moreover, it may increase social inequalities because parental aspirations and beliefs become even more important if the child's true academic potential is still uncertain. The disadvantages of early tracking may thus outweigh the advantages seen in efficiency improvements through more targeted teaching.

In line with this reasoning, existing empirical research does not find clear evidence in favor of either early or late tracking (see, Hall, 2012, on Sweden; Kerr et al., 2013, on Finland; Malamud and Pop-Eleches, 2011, on Romania; and Dustmann et al., 2017 and Matthewes, 2020, on Germany). Yet, occasional positive effects of postponing or abolishing school tracking tend to be concentrated among youths from disadvantaged backgrounds.[2] Similarly, some studies point to a strong association between parental background and student achievement in education systems with early school tracking (Dustmann, 2004, on Germany; as well as Hanushek and Wössmann, 2006, and Waldinger, 2007, in cross-country comparisons).[3]

[1]Whereas in most European countries tracking historically takes place across schools, in the United States and Canada students are mainly sorted within schools (see, for example, Card and Giuliano (2016), for a recent evaluation of a within-school tracking program in the United States).

[2]Recent contributions for the United States studying the effects of attending a magnet secondary school with a special curriculum and competitive admission process document basically no causal effects on student achievement (e.g., Abdulkadiroglu et al., 2014). Studies by Canaan (2020) for France and Meghir and Palme (2005) for Sweden evaluate reforms that involve multiple components in addition to postponing or abolishing tracking (i.e., increased compulsory schooling, changed curricula and teacher quality). They document beneficial effects on educational attainment and earnings especially for youths from disadvantaged backgrounds.

[3]The study by Waldinger (2007) challenges the view that early school tracking causes social inequality by arguing that the inequalities exist already in grades in which the tracking has not yet taken place.

In this paper, we aim to contribute to a better understanding of both the role of parents in secondary school track choice and the potential relationship between early tracking and social inequalities. Granting parents influence on track choice may, on the one hand, improve the assignment of students to tracks if parents have superior information than teachers on a child's ability and if they use this information as a basis for their decision. On the other hand, parents may base their decision on other criteria beyond ability, potentially overturning a track assignment that was based purely on ability. Parents tend to have aspirations for their child that depend on their own achievement rather than their child's ability. They often desire for their child to achieve a similar socioeconomic status (SES) to what they have. Alternatively, parents could hold unrealistic or biased beliefs about their child's ability that reflect their own achievement more than their child's academic potential. Thus, high SES parents are more likely to think that their children are smarter than they really are compared to low SES parents. In a school system with tracking, such parental aspirations and beliefs may contribute to fostering social inequalities and decreasing the efficiency of tracking if parental aspirations and beliefs do not conform to a child's ability.

We evaluate the effects of a recent state-level reform in Germany that changed the way in which the tracking is decided. We study the impact of changing the purpose of teachers' track recommendations from mandatory to informational on transition rates to secondary school tracks and grade repetition in the first two grades of secondary school as a measure of misallocation.[4] In Germany, school tracking into one of three different tracks takes place at the beginning of secondary school. The academic (high) track prepares for higher education, whereas the two vocational (intermediate and low) tracks prepare for apprenticeships and other forms of vocational training at the secondary level.[5] In grade four, teachers evaluate the performance and potential of each student and officially recommend—in their view—the

[4]A recent study by Bach and Fischer (2020) examines the effects of teacher recommendations on student achievement in *primary* school. They document a slightly better performance of students when teacher recommendations are binding.

[5]In some federal states, there also exist comprehensive schools in which students are tracked within schools as an alternative to the classical tiered schools.

most suitable secondary school track for the student. In some federal states, these teacher recommendations are mandatory, whereas in others they are purely informational. It is important to note, however, that in states with mandatory teacher recommendations, parents still have the right to send their child to a lower secondary school track than recommended by the teacher (i.e., downgrading is always possible).

We can think of two reasons why a change from a regime in which teachers decide to one in which parents decide may impact on transition rates and grade repetition rates. On the one hand, it may be that parents can correct the teacher's recommendation so as to send their child to a higher-than-recommended track that better matches their child's ability. In this case, we expect to find positive impacts on the transition rates to the high and intermediate tracks but no increases in grade repetition rates at the high track and likely neither at the intermediate track because the students who choose a higher track than the teacher recommended are not systematically worse in terms of academic ability than those students who attend the higher tracks as recommended. Transitions to the low track decline as a consequence of the reform, and the grade repetition rate potentially increases at this track as average student ability declines.

On the other hand, parents might have biased beliefs or may be primarily led by their aspirations. If this is the case, we expect to see heterogeneous effects on transition rates depending on the SES of families. If parents desire for their child an SES similar to what they have, high SES families have a stronger incentive to deviate from a lower track recommendation. The same holds true if high-SES parents are more likely to overestimate their child's academic ability than low-SES parents. Misallocation of high-SES children to higher tracks leads to a deterioration of average student ability at all tracks, which should increase grade repetition rates at all tracks. This reasoning implies that we expect the reform to have positive, heterogeneous effects on transition rates to the high and intermediate tracks and positive effects on grade repetition rates at all three tracks, if parental aspirations or upward biased beliefs rather than the child's ability are important for track choice.

Our analysis uses administrative data on two of the largest federal states, which cover around 30 percent of the German population: Baden-Wuerttemberg and Bavaria.[6] The federal state of Baden-Wuerttemberg abolished mandatory teacher recommendations in 2011, such that parents were granted full rights to choose the school track for their child. Bavaria, a comparable federal state with regard to size as well as political and economic characteristics, continues to rely on mandatory teacher recommendations until today.[7]

The empirical analysis unfolds in two steps. We start with a descriptive event study to illuminate the various responses to the reform, as well as the underlying channels and effect drivers. In order to examine the impacts of the reform and their heterogeneity causally, we then apply a disaggregated version of the Synthetic Control Method (SCM) (originated in Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). For each district in the treatment state of Baden-Wuerttemberg, we form a synthetic comparison district as an optimally weighted average of districts in the control state Bavaria. Under the conditions of the SCM approach, we thus obtain an unbiased estimate of the reform effect for every district in Baden-Wuerttemberg. Pooling the district-level treatment effects, we get a distribution of effects, which is ideal for studying effect heterogeneity. We use in-time permutations of treatment assignment to construct the distribution of treatment effects under the sharp null hypothesis of zero effects. Specifically, we apply analogous SCM procedures to the districts of the treated state in time periods in which the reform was not yet in place. This placebo distribution of treatment effects is then compared against either the treatment distribution as a whole or against a (weighted) average treatment effect in order to construct exact $p$-values.

Our findings suggest that, after the reform, parents of children with a recommendation for the two lower tracks increasingly overturned the teacher's advice, sending their child to the next higher track instead. Similarly, parents of children with a recommendation for

[6] Baden-Wuerttemberg and Bavaria together have about 24.15 million inhabitants (Statistisches Bundesamt, 2016).

[7] Track assignment in grade five usually persists through grade ten. In fact, there is only limited and mostly downward mobility between school tracks until the end of lower secondary education (e.g., Tamm, 2008; Dustmann et al., 2017).

the high track downgraded less often than before the reform. The behavioral responses of parents stand in contrast to the behavior of teachers who continued to recommend the same proportions of students to the three tracks after the reform as before.

According to the disaggregated SCM estimates, the abolition of mandatory teacher recommendations led to an overall increase in the *transition rate* to the highest school track by about 8 percent in the first year after the reform. For the lowest school track, the effect is even more pronounced with a decline in the transition rate by 29 percent. Moreover, it seems that the effect magnitude in the first post-reform year is particularly dependent on the average income of private households in a district: the better off a district, the stronger the reaction to the policy reform. In subsequent years, these effects stay constant or even grow stronger, suggesting that students and their parents learn about the institutional change, and eventually update their behavior over time. Although these results might be seen as a desirable development towards more and higher quality education, this is only one part of the story. The other part concerns the overall *repetition rate* in grade five, for which we find a significantly positive effect: it increased by approximately 85 percent compared to the pre-intervention repetition rate, that is, the absolute number of repeaters almost doubled as a consequence of abolishing mandatory teacher recommendations. This translates into 500 additional children who failed and had to repeat grade five in the school year 2012/13 alone. We can, on top of that, present evidence that repetition rates also increased in grade six, suggesting that the effects of this misallocation might be even more pronounced and pass through to higher grades. Hence, while the reform may have eased the pressure on students, parents, and teachers in primary school, our findings suggest that this pressure may have just been shifted to (the beginning of) secondary school. Moreover, our findings are in line with the hypothesis that the socioeconomic status of parents is an important determinant of their school track choice, even if this means for their child to repeat a grade in secondary school. Hence, our results illustrate how early tracking may contribute to foster social inequalities.

The remainder of the paper proceeds as follows: In Section 2, we describe the relevant aspects of the German education system and introduce the data. In Section 3, we present first descriptive evidence on the consequences of the reform. In Section 4, we explain our implementation of the disaggregated SCM approach. In Section 5, we present our empirical findings on the causal effects of the reform in Baden-Wuerttemberg. In Section 6, we conclude. Appendix A gives detailed information on the data set, including a description of all variables and data sources. Appendix B provides further descriptive statistics, as well as appendant materials from the estimation.

# 2 Background and Data

## 2.1 The German Education System

In Germany, the federal states are primarily responsible for the education system and related policies. Although there exists no federal legislation for education, the assembly of the state ministers of education and research (called "Kultusministerkonferenz"), in which the federal minister also participates, uses thematic committees that work on the harmonization of education policies across states and formulates state treaties that are then passed in all federal states.[8] As a consequence, the state education systems share the same basic structure, and certificates and degrees awarded in one state are recognized in all other states. However, some education policies may differ across states or are implemented at different points in time. From the primary until the tertiary level, education is mostly public and financed by the federal states.

After four years of comprehensive primary school, students continue on a secondary school where they are grouped according to their academic abilities.[9] These schools are usually locally separated and each offers one of three main tracks that differ in their length, academic standard, and educational orientation. The two lower track schools, called "Hauptschule" and

[8]According to §30 of the Basic Law, cultural sovereignty rests with the 16 federal states, which means that laws in the fields of education, science, and culture are implemented at the state level.

[9]In Berlin and Brandenburg, primary schools last six years.

"Realschule," are academically less demanding and prepare for apprenticeship and other forms of vocational education at the upper secondary level. They end with a final examination after grades nine and ten, respectively. The highest track in terms of academic standard, "Gymnasium," prepares for university studies and ends after grade twelve or thirteen with the university entrance qualification.[10] Teachers have different qualifications in the different secondary school tracks. The education of teachers for the vocational tracks puts a stronger emphasis on pedagogy and vocational preparation of students, whereas the curriculum for teachers at the academic track is more subject-oriented and academically advanced.

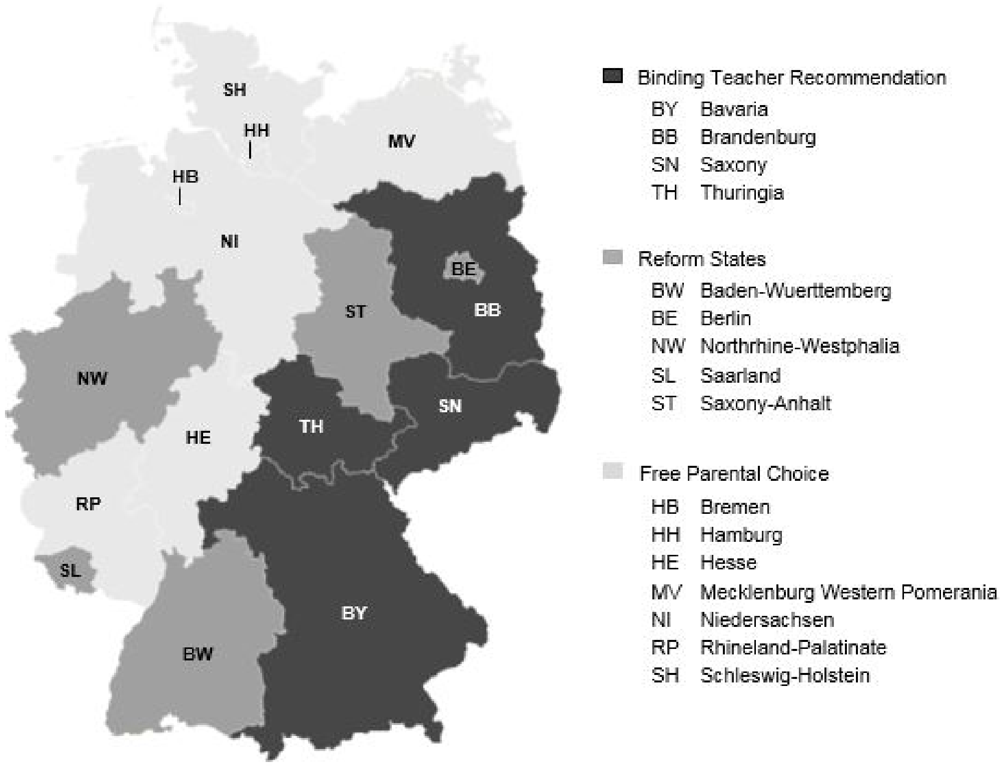
The transition from primary to secondary school marks an important milestone in the educational career of a child. In grade four, teachers evaluate the performance and potential of each student and recommend the highest possible secondary school track suitable for the student. In some federal states, these teacher recommendations are mandatory, whereas in others they are purely informational (see Figure 1 for an overview of the regulations on teacher recommendations across the German states).

In states with non-binding teacher recommendations, parents can enroll their child at any track they like, even against the teacher's advice. In states with mandatory teacher recommendations, transition to a higher-than-recommended track is either not possible or requires additional efforts involving, for instance, entrance examinations.[11] These are, however, very rarely pursued. Parents always have the right to send their child to a lower secondary school track than recommended by the teacher. Track choice targeting grade five is usually permanent; however, in principle, and depending on the performance, students can switch tracks during lower secondary education. However, Dustmann et al. (2017) document that less

[10]The East German states traditionally have a partially integrated secondary school system in which the two lower tracks are offered at the same schools. In some federal states, in addition to the tiered schools, there exist comprehensive schools in which the ability grouping is organized within school. In Baden-Wuerttemberg and Bavaria, the two federal states we focus on, comprehensive schools play only a negligible role. Baden-Wuerttemberg has introduced comprehensive schools only in the school year 2012/2013, while in Bavaria only four schools of this type exist (out of more than 1750 at the secondary school level).

[11]Some minor differences across states exist with respect to the evaluation criteria and procedures in the case of disagreement between teacher recommendation and parental preferences, but these are only marginal and not relevant for our study.

Figure 1: Cross-State Variation in the Binding Nature of Teacher Recommendations




*Note:* This figure gives an overview of the institutional design of the tracking system (for the year 2013). It distinguishes between states that always had binding recommendations (dark), states with free parental choice (light), and states that experienced a reform (grey).
*Source:* Own depiction based on state-level laws.

than 2 percent do so in practice.[12] Only after a successful completion of one of the two lower tracks changes are common, with the best graduates from a lower track continuing on a higher track.

Baden-Wuerttemberg changed the purpose of teacher recommendations from mandatory to informational in November 2011. The reform was implemented swiftly after a center-left coalition, consisting of Greens and Social Democrats, won the majority in the state elections in 2011. It was the first time in the post-War era that a center-left coalition came to power in Baden-Wuerttemberg. With the reform, the center-left government aimed to ease the pressure on primary school students and to decrease social inequalities. As such,

[12]In Figure 5 in Section 3 we present evidence for grade five that only a very small number of students decide to voluntarily repeat the grade while changing to the next higher track.

being the consequence of shortly preceding elections with unanticipated outcome, the reform constitutes a natural experiment that ideally lends itself for studying the role of parents in secondary school track choice. We compare the situation in Baden-Wuerttemberg to that in Bavaria, a politically, geographically, and economically similar state in which teacher recommendations have always been mandatory.

## 2.2 Data

Our main administrative data source is the school statistics collected by the Statistical Offices of the Federal States.[13] They contain the complete records of all secondary schools in a state. However, some breakdowns vary across states because education systems and statistics are regulated at the state and not at the federal level. We use student counts at the district level for both Baden-Wuerttemberg and Bavaria, covering the school years beginning in fall 2005 through fall 2016. Importantly, during this time frame, there were no other educational policy interventions that might invalidate our identification strategy.[14]

With 44 districts in Baden-Wuerttemberg and 94 districts in Bavaria[15], we have a balanced panel with 1,656 observations in total. Because the reform of secondary school tracking in Baden-Wuerttemberg took place in November 2011, we are left with seven pre-intervention periods and five post-intervention periods.

Our first set of outcome variables is the transition rates to the three secondary school tracks. We compute them as the number of incoming students in grade five in a particular track divided by the total number of incoming students in grade five. Thus, these rates are not conditioned on teachers' recommendations, unless not stated otherwise. We calculate the transition rates for all available years also to study the dynamics of the reform impacts.

Our second set of outcome variables refers to grade repetition in grade five and six as a

[13]Appendix A.2 provides further details on the data sources used.

[14]The last major educational reform in Germany was the shortening of the duration of the academic school track from nine to eight years, which came into effect in fall 2004.

[15]Note that we decided to merge Bamberg district and Bamberg city into a single district because these districts share common schools. The same rationale holds true for Schweinfurt. We therefore end up with 94 instead of the official number of 96 districts in Bavaria. See Appendix A.3 for more details.

measure of the academic performance of students. Note that grade retention is a result of a (very) weak overall grade point average (across all subjects taught by different teachers). A retained student has to repeat the respective grade in the following academic year at the same or at a lower school track. As dependent variables, we use the overall repetition rate, as well as the track-specific repetition rates. The track-specific repetition rates are constructed as the number of retained students in a specific grade-track combination divided by the total number of students in that respective grade-track pair. The overall repetition rate does not condition on the tracks and divides the number of retained students in a specific grade by the total number of students in that grade. Note that the district-level repetition rates only count students who repeat the grade at the same school track as the one at which they were retained. In Figure 5 in Section 3, using aggregate state-level data, we show that these repetition rates somewhat underestimate the corresponding repetition rates that also include repeaters who repeat the grade at a lower track, i.e., providing a rather lower bound estimate.[16] We provide further details on how we construct our dependent variables in Sections A.1.1 and A.3 in the Appendix.

We supplement the panel data on transition and grade repetition rates with other data from the Federal Employment Agency and District Statistics on national accounts and budgets to construct control variables capturing the economic and financial condition of a respective district. Specifically, we consider the following control variables: population (in 1,000 inhabitants), population density, unemployment rate, GDP per capita, and average primary household income (both in 1,000 Euro). We construct further controls from the school statistics on the student-teacher ratio, the share of foreign students aged 11-15 and the number of academic/intermediate track schools (per 100,000 inhabitants) in a district. More information on these variables and a comprehensive list of data sources are given in Appendix A.

[16] As we have the information on repeaters who downgrade the track merely at the level of the state for Baden-Wuerttemberg, we cannot use it to construct our district-level outcome variables on grade repetition.

# 3 Descriptive Evidence

Table 1 shows descriptive statistics (i.e., the mean, standard deviation, minimum, and maximum) of all variables used in the empirical analysis. We present unweighted averages across districts separately for Baden-Wuerttemberg and Bavaria. The values of the outcome variables refer to the pre-treatment year 2011, while the other variables are averaged across the whole pre-treatment period. In addition, the last two columns report the $t$-statistic of the difference-in-means test and an additional indicator comparing the support of the respective variable between Baden-Wuerttemberg and Bavaria. The latter will be discussed in Section 5.3, where we assess the matching quality of the SCM estimations.

In Baden-Wuerttemberg in 2011, 40.39 percent of all students transitioned to the highest school track, which is similar to 37.90 percent in Bavaria. However, there is a slight difference in the allocation of students to the two lower tracks. While 24.19 percent of all students transition to the lowest school track in Baden-Wuerttemberg, the corresponding figure is 10 percentage points higher in Bavaria. As apparent from this table, repetition rates are also generally higher in Bavaria than in Baden-Wuerttemberg. In 2011, for example, 0.5 percent of all students in Baden-Wuerttemberg had to repeat grade five, while the respective repetition rate is 1.1 percent in Bavaria. The $t$-statistics for the test of equality of means suggest that both states are similar in terms of average GDP per capita, population density, household income, unemployment, and student-teacher ratios, which is a first indication that the Bavarian districts are a suitable comparison group for the districts in Baden-Wuerttemberg.

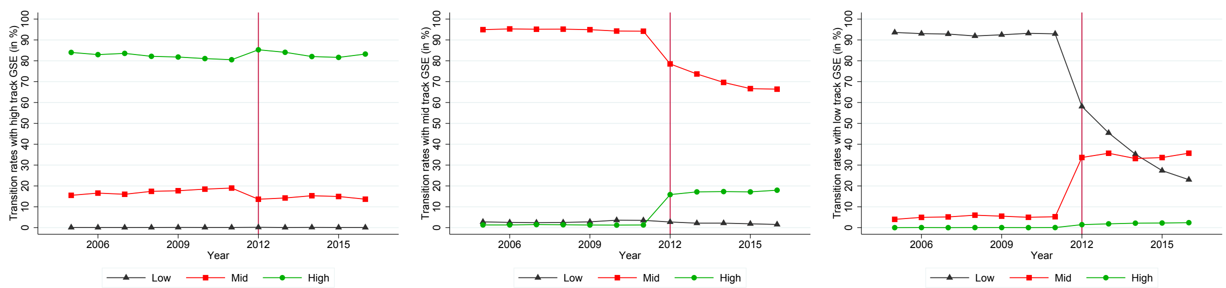
Table 1: Descriptive Statistics

| | Baden-Wuerttemberg | | | | Bavaria | | | | $t$-Stat. | Support |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max | | |
| Transition rate, high track | 40.39 | 7.65 | 29.86 | 65.53 | 37.90 | 12.33 | 9.71 | 66.17 | 1.45 | 1 |
| Transition rate, mid track | 34.12 | 5.52 | 15.73 | 42.27 | 25.35 | 7.16 | 6.99 | 46.63 | 7.88 | 1 |
| Transition rate, low track | 24.19 | 4.76 | 9.42 | 32.60 | 33.76 | 9.91 | 9.98 | 57.76 | -7.66 | 1 |
| Rep. rate in 5, all | 0.51 | 0.25 | 0.15 | 1.31 | 1.11 | 0.60 | 0.13 | 3.40 | -8.20 | 1 |
| Rep. rate in 5, high track | 0.26 | 0.26 | 0.00 | 1.31 | 0.60 | 0.49 | 0.00 | 1.90 | -5.27 | 1 |
| Rep. rate in 5, mid track | 0.61 | 0.37 | 0.00 | 1.33 | 1.73 | 1.42 | 0.00 | 7.14 | -7.14 | 1 |
| Rep. rate in 5, low track | 0.80 | 0.55 | 0.00 | 2.68 | 1.10 | 0.94 | 0.00 | 4.81 | -2.34 | 1 |
| Rep. rate in 6, all | 0.86 | 0.36 | 0.27 | 2.06 | 1.85 | 0.82 | 0.39 | 4.65 | -9.85 | 0 |
| Rep. rate in 6, high track | 0.75 | 0.52 | 0.00 | 2.61 | 1.76 | 0.94 | 0.00 | 4.00 | -8.13 | 1 |
| Rep. rate in 6, mid track | 1.15 | 0.65 | 0.24 | 3.23 | 2.65 | 1.47 | 0.47 | 7.76 | -8.32 | 0 |
| Rep. rate in 6, low track | 0.64 | 0.50 | 0.00 | 1.74 | 1.11 | 1.05 | 0.00 | 5.67 | -3.54 | 1 |
| Population, in 1000 | 243 | 129 | 54 | 598 | 133 | 140 | 39 | 1320 | 4.56 | 1 |
| Population density, in km$^2$ | 526 | 595 | 103 | 2882 | 424 | 636 | 69 | 4251 | 0.92 | 1 |
| GDP p.c., in 1000€ | 34.34 | 10.76 | 22.61 | 70.82 | 32.55 | 15.07 | 16.36 | 91.70 | 0.80 | 1 |
| Unemployment rate | 5.30 | 1.30 | 3.51 | 9.14 | 5.40 | 1.79 | 2.44 | 10.09 | -0.36 | 1 |
| Household income, in 1000€ | 25.20 | 2.88 | 21.07 | 38.19 | 24.19 | 3.83 | 18.77 | 41.96 | 1.71 | 1 |
| Stud./teacher, high track | 17.84 | 0.68 | 16.43 | 19.68 | 18.01 | 0.94 | 15.97 | 20.09 | -1.25 | 1 |
| Stud./teacher, mid track | 20.72 | 0.77 | 19.04 | 22.05 | 20.81 | 1.21 | 16.13 | 22.69 | -0.53 | 1 |
| Share foreign stud. aged 11-15 | 11.02 | 5.64 | 4.17 | 28.73 | 7.37 | 5.12 | 1.54 | 25.56 | 3.65 | 0 |
| High track schools per 100' | 4.23 | 1.16 | 2.41 | 9.20 | 3.04 | 1.62 | 0.55 | 9.67 | 4.92 | 1 |
| Mid track schools per 100' | 4.66 | 0.92 | 3.26 | 6.85 | 2.26 | 0.88 | 0.73 | 4.84 | 14.45 | 0 |
| Districts | 44 | | | | 94 | | | | | |

*Note:* This table reports descriptive statisitcs (i.e., mean, standard deviation, minimum and maximum) for all variables included in the estimation by treatment state. Transition and repetition rates, the unemployment rate and the share of foreign students aged 11-15 are all measured in percent. Population is the number of people living in the respective district (measured in 1000 inhabitants). GDP p.c. and household income are reported in 1000€. The number of high-track and intermediate-track schools per capita is given per 100,000 inhabitants. The last two columns report the $t$-statistic of the difference-in-means test and an additional variable indicating whether the support of the variable in the treated sample is a subset of the support in the comparison sample. Specifically, the indicator equals one if $\min(x_{BW}) \geq \min(x_{BY}) - 0.1 \cdot \mathrm{sd}(x_{BY})$ and $\max(x_{BW}) \leq \max(x_{BY}) + 0.1 \cdot \mathrm{sd}(x_{BY})$.
*Source:* Own calculations based on various data sources. See Appendix A.

Figure 2 depicts the time series of the aggregate transition rate *conditional on the recommendation type* for the state of Baden-Wuerttemberg over the sample period 2005-2016, which allows a first assessment of the potential reform effects.[17] If parents and teachers have diverging views on the appropriate school track for a child, we would expect that the 2011 reform in Baden-Wuerttemberg that abolished mandatory teacher recommendations would impact transition rates in the sense that transitions to higher school tracks increase.

Figure 2: Transition Rates Conditional on Recommendation Type




(a) Rec. to high track (b) Rec. to mid track (c) Rec. to low track

*Note:* The graphs show the transition rates by recommendation type over the sample period 2005-2016. Subfigure (a) shows the transition rates to the three school tracks for all students with a high-track recommendation. Subfigures (b) and (c) show the graphs for students with an intermediate- and low-track recommendation, respectively. In all three graphs, the green-dotted line represents the share of students attending the highest school track, while the red-squared and the black-triangulated lines show the respective time series for the intermediate and low track.
*Source:* Own calculations based on school statistics of Baden-Wuerttemberg.

Subfigure (a) shows the transition rates to the three different school tracks for all students who obtained a high-track recommendation at the end of primary school. The green-dotted line represents the conditional share of students transitioning to the highest school track, while the red-squared and the black-triangulated lines show the respective time series for the mid and low track. Subfigures (b) and (c) show the time series for students with a mid- and low-track recommendation, respectively.[18] Three things are worth noting. First, prior to the reform, the recommendation was indeed binding, as more than 80 percent (95%,

[17]Note that these conditional transition rates are available on an aggregate level only (i.e., not on district level), such that we cannot use them for our disaggregated analysis in Section 5.

[18]Note that these three conditional transition rates do not necessarily have to add up to 100 percent. For example, in 2011, about 95 percent of the students attended one of the three regular secondary school tracks, which are the focus of our analysis. Hence, a small share of students attend other institutions, such as special needs schools or Waldorf schools, which are not included in Figure 2.

92%) of all students with a high (intermediate, low) recommendation actually attended the recommended track. Second, while upgrading was almost impossible, around 15-20 percent of the students with a high-track recommendation decided to attend the lower intermediate school track. Third, Figure 2 provides prima facie evidence that the reform in 2011 had a clear impact on the unconditional transition rates to the different tracks. After the reform, students with a recommendation for the academic track were less likely to attend the lower intermediate track (Subfigure (a)), whereas students with a recommendation for the intermediate track were more likely to attend the higher academic track (see the jump of the green-dotted line in Subfigure (b)). Even among students with a low-track recommendation, a few chose to attend the high track after the reform (Subfigure (c)). Taken together, these reactions increase the overall transition rate to the high track. Similarly, students with a recommendation for the low track increasingly attend the intermediate track after the reform. Conditional on having a recommendation for the low track, the transition rate to this track decreased from more than 90 percent to less than 50 percent within two years. Overall, this led to a decline in the transition rate to the low track. Finally, for the intermediate school track, there were "incoming" students who would have attended the low track without the reform (see the jump of the red-squared line in Subfigure (c)), but this school track also lost students with a mid-track recommendation to the highest school track (see the jump in the green-dotted line in Subfigure (b)). The overall effect on the transition rate to the mid track is, thus, ambiguous. Eventually, there is almost no change in the behavior of teachers regarding their recommendations (see Figure 9 in Appendix B).

One may wonder whether the substantive shifts in track attendance affect student-teacher ratios. As Figure 10 in Appendix B suggests, however, student-teacher ratios do not appear to be affected by the reform. The student-teacher ratios at the academic and intermediate track exhibit a constant, slightly declining trend, whereas the student-teacher ratio at the low track (including primary schools) shows a slight U-shaped pattern with an increasing trend from 2012 onwards. Overall, this pattern of rather constant student-teacher ratios over time can partly be explained by the fact that teachers in the intermediate and low track have the same education and are typically state civil servants, which implies that they can be relocated to another intermediate- or low-track school.

In a second step, making use of our district-level data for Baden-Wuerttemberg, we assess potential effect drivers. Specifically, we regress the change in the respective transition rate from 2010 to 2011 (i.e., the two pre-treatment years) and from 2011 to 2012 (i.e., a pre- and a post-treatment year) on selected district characteristics, including population density, average household income, and GDP per capita. Conditional on population density, which captures differences between more urban and more rural districts, and GDP per capita, which reflects the overall economic condition of a district, we interpret the average household income as a proxy for the average socioeconomic status of families in a district. The results are shown in Table 2.

Table 2: Relationship between Change in Transition Rates and District Characteristics

| | High Track | | Mid Track | | Low Track | |
|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post |
| Population density, in km$^2$ | 0.0010 | 0.0004 | -0.0006 | 0.0001 | -0.0005 | 0.0004 |
| | (0.0010) | (0.0013) | (0.0008) | (0.0012) | (0.0004) | (0.0011) |
| Household income, in 1000€ | 0.0408 | 0.1858** | -0.0681 | 0.0196 | 0.0478 | -0.1855** |
| | (0.0893) | (0.0785) | (0.0835) | (0.0915) | (0.0395) | (0.0718) |
| GDP p.c., in 1000€ | -0.0109 | 0.0073 | 0.0394 | -0.0177 | -0.0235 | 0.0004 |
| | (0.0436) | (0.0683) | (0.0406) | (0.0687) | (0.0190) | (0.0430) |
| Constant | -0.9012 | -2.7039 | 0.9176 | 3.0700 | -0.6613 | -3.1316 |
| | (2.6283) | (3.0647) | (2.2790) | (3.2312) | (1.2874) | (2.5952) |
| Observations | 44 | 44 | 44 | 44 | 44 | 44 |

*Note:* The table shows the results from regressions of the change in the respective transition rate before the reform (from year 2010 to 2011, columns (1), (3) and (5)) and after the reform (from year 2011 to 2012, columns (2), (4) and (6)) on selected district characteristics using the 44 districts of Baden-Wuerttemberg. Heteroskedasticity-consistent standard errors are given in parentheses. ***,**,* indicate significance at the 1%, 5%, 10% level, respectively.
*Source:* Own calculations based on various data sources. See Appendix A.

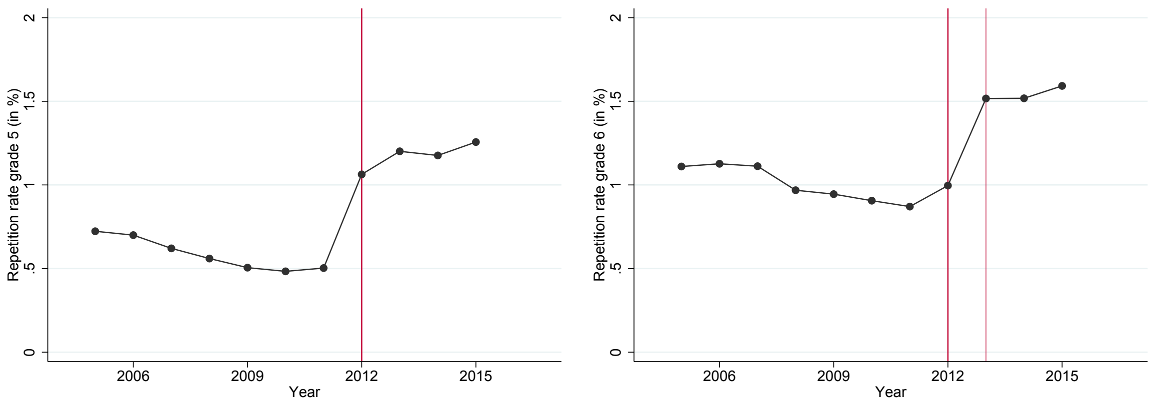
In line with our hypothesis that high SES families show a larger response to the reform, we find that household income is the only regressor that significantly correlates with the change in the transition rate for the high and the low secondary school track: the higher the household income in a given district, the larger the increase (decrease) in the transition rate for the high (low) school track, keeping constant population density and GDP per capita. There is no such correlation (i) in the pre-treatment case (see Columns (1), (3) and (5)) and (ii) for the intermediate school track. Again, the latter can be explained by mutually offsetting effects.

While the first piece of evidence suggests that the reform had the potential to increase the standard of secondary schooling and educational attainment of German students as they, on average, attend a higher school track after the reform, Figure 3 delivers a less optimistic assessment of the reform. Subfigure (a) shows the time series for the overall repetition rate in grade five for the school years beginning in 2005 through 2015. The overall repetition rate includes all students who are forced to repeat a grade because of an insufficient grade point average and who repeat the grade at the same track than the one at which they were retained.[19] The red vertical line denotes the first post-intervention school year beginning in 2012. The same time series for grade six is plotted in Subfigure (b), where the reform effects kick in one year later in 2013 (second vertical line in the graph) when the first cohort of students affected by the reform progresses to grade six. Both graphs show a dramatic increase in the repetition rates after the reform. After a slightly declining trend before 2011, the repetition rate in grade five jumps from around 0.5 percent in 2011 to more than 1 percent in 2012. Similarly, the repetition rate in grade six also jumps up from 1 percent in 2012 to around 1.5 percent in 2013. If we split up the repetition rates by school track (see Figure 4), it seems that the aggregate effect tends to be driven by increases in grade repetition at the intermediate track and to some extent also at the high track.

Finally, Figure 5 presents two further pieces of evidence on grade repetition: (i) on voluntary repetitions with the intention to upgrade to a higher track and (ii) on forced repetitions

[19] In rare cases, students decide to voluntarily repeat a grade and at the same time to change to a higher track, see Figure 5 below. Moreover, retained students may repeat the grade at a lower track. These cases involving track upgrading and downgrading are not included in our main measures of grade repetition.

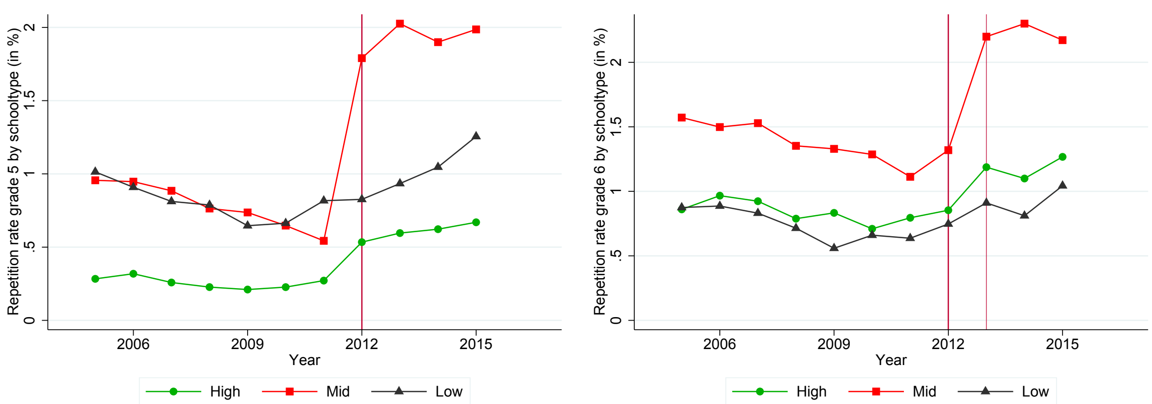
Figure 3: Repetition Rates over Time




(a) Grade 5 (b) Grade 6

*Note:* Subfigure (a) shows the time series for the overall repetition rate in grade five. The red vertical line denotes the first post-intervention year 2012. Subfigure (b) plots the time series for the repetition rate in grade six. As the first cohort of students affected by the reform progresses to grade six in 2013, the first post-intervention year is 2013 (second vertical line). The repetition rates are measured by the number of students repeating a grade divided by the total number of students in this grade.
*Source:* Own calculations based on school statistics of Baden-Wuerttemberg.

Figure 4: Repetition Rates over Time by School Type



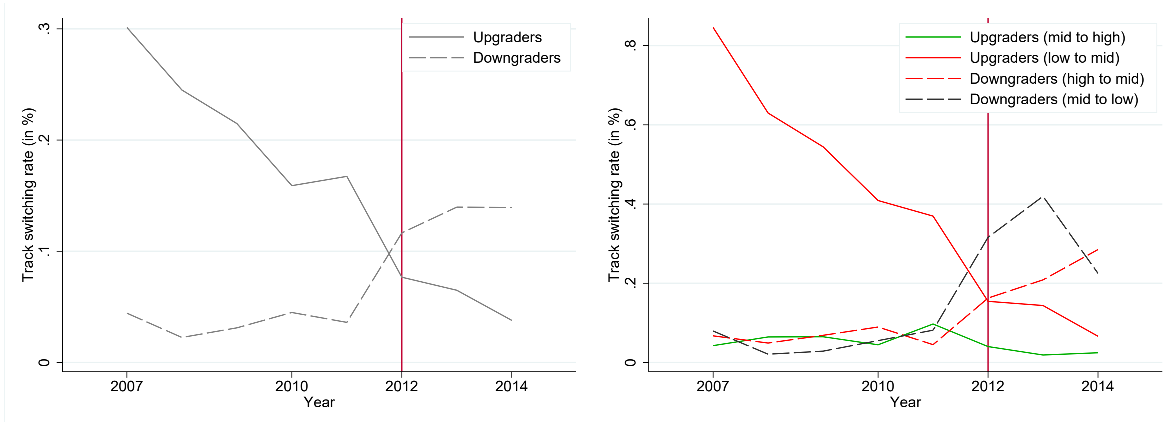

(a) Grade 5 (b) Grade 6

*Note:* This figure shows the time series for the repetition rates in grade five and six split by school type. In both figures, the green-dotted line represents the repetition rate in high school track, while the red-squared and the black-triangulated line show the respective time series for the mid and low track.
*Source:* Own calculations based on school statistics of Baden-Wuerttemberg.

involving a downgrade of track. As becomes apparent from Subfigure (a) of Figure 5, grade repetition involving a change of track is quantitatively much less important than grade repetition at the same track. In 2011 (2012), when the overall grade repetition rate in grade five was at 0.50 percent (1.06 percent in 2012, see Subfigure (a) of Figure 3), voluntary grade repetition with the intention to upgrade was at 0.17 (0.08) percent and grade repetition involving a downgrade was at 0.04 (0.12) percent. Subfigure (b) of Figure 5 shows the repetition rates involving track changes by origin and destination track. From 2012 on, when teacher recommendations became nonbinding, the share of students who voluntarily repeated a grade in order to attend a higher track declines and vanishes, both at the intermediate track (green solid line) and at the academic track (red solid line). In contrast, the share of repeating students with insufficient school marks who downgrade the track increases in the first two years after the reform. In the year 2012, the share of downgraders on the intermediate track coming from the academic track (red dashed line) is 0.16 percent, up from 0.04 percent in 2011. Similarly, the share of downgraders on the low track coming from the intermediate track is 0.08 percent in 2011 and 0.32 percent in 2012. Taken together, the evidence on downgrading suggests that the track-specific repetition rates for the high and intermediate tracks, displayed in Subfigure (a) of Figure 4, underestimate the full extent of grade retention in grade five at these tracks by 0.15 to 0.20 percentage points in the early years after the reform—i.e., they provide a rather conservative, lower bound.

In sum, the descriptive analysis generates the following insights: although we observe an increase in the transition rates to higher school tracks, the evidence hints at a potential misallocation of students to school tracks as repetition rates have significantly increased after the reform. In addition, the effects of the reform seem to be more pronounced for richer districts in terms of the average household income, which can be interpreted as districts with a higher average SES of households given that we also condition on other district indicators. In the remainder of the paper, we engage in a more formal analysis of the causal reform effects and compare the development in the districts of Baden-Wuerttemberg to that of matched districts in Bavaria.

Figure 5: Grade Repetition in Grade 5 Involving Track Changes




(a) Overall (b) By Track

*Note:* This figure shows the time series for the repetition rates in grade five that involve track changes. In Subfigure (b), the solid green line represents voluntary grade repetition of former intermediate track students who upgrade to the high school track; the dashed black line refers to forced grade repetition of former intermediate track students who downgrade to the low track; and the red lines refer to up- and downgrading students who repeat grade five at the intermediate track.
*Source:* Own calculations based on Table 3.7, Fachserie 11/1, Statistical Offices of the States.

# 4 Empirical Strategy

## 4.1 Disaggregated Synthetic Control Method

To evaluate the causal effects of the reform of secondary school tracking, we apply the synthetic control method (SCM). The approach was originally developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) for the quantitative analysis of comparative case studies based on longitudinal data on a single treated unit and a moderate number of control units. Recent contributions have extended this approach to evaluate treatments based on random samples with a large cross-section dimension (see Abadie and L'Hour (2019)). In our application, we consider an intermediate case: our data are derived from a finite population of 44 treated units and 94 comparison units.[20] Hence, we apply the SCM to each of the treated units and then study the reform effects at the aggregate level and their heterogeneity across treated units.

[20]See, for example, Acemoglu et al. (2016), Dickert-Conlin et al. (2019), and Dube and Zipperer (2015), who also use the synthetic control method with multiple treated units.

More specifically, our approach unfolds as follows. Suppose access to a panel data set with $J+1$ cross-section units (e.g., districts in our case) indexed by $i = 1, \ldots, J+1$ that are followed over $T$ time periods indexed by $t = 1, \ldots, T$. The goal is to evaluate the effect of a reform (i.e., the treatment) taking place between periods $T_0$ and $T_0+1$. Thus, we have a pre-treatment period with $t \leq T_0$ and a post-treatment period with $t > T_0$. Following Splawa-Neyman et al. (1990), Roy (1951), and Rubin (1974), we rely on the potential outcome framework for treatment evaluation and define $Y_{it}(1)$ as the potential outcome in the case of treatment and $Y_{it}(0)$ the potential outcome in the absence of treatment. Let $D_{it}$ denote an indicator for active exposure to the treatment.[21] Then, $Y_{it} \equiv (1 - D_{it})Y_{it}(0) + D_{it}Y_{it}(1)$ corresponds to the observed outcome. In periods $t \leq T_0$, we have $Y_{it} = Y_{it}(0)$ for all $i$. In periods $t > T_0$, we observe $Y_{it}(1)$ for units in the treatment group and $Y_{it}(0)$ for units in the control group.

To simplify, let us start with the case that only unit $i = 1$ receives the treatment such that $D_{it} = 1$ if $i = 1$ and $t > T_0$ and $D_{it} = 0$ else. We want to identify and estimate the treatment effect for unit $i = 1$ in periods $t > T_0$, which we denote as

$$\tau_{1t} = Y_{1t}(1) - Y_{1t}(0)\,,\ t > T_0\,, \tag{1}$$

where $Y_{1t}(1)$ is observed for unit $i = 1$ and we can replace $Y_{1t}(1)$ by $Y_{1t}$ in Eq. (1), but $Y_{1t}(0)$ is counterfactual. We can identify $\tau_{1t}$ under the following assumptions (see also Abadie et al., 2010).

First, the potential non-treatment outcome, $Y_{it}(0)$, follows a linear factor model given by

$$Y_{it}(0) = \boldsymbol{\theta}_t \mathbf{Z}_i + \delta_t + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{it}\,, \tag{2}$$

where $\mathbf{Z}_i$ is an $(R \times 1)$ vector of observable characteristics and $\boldsymbol{\theta}_t$ a conformable vector of

[21] Implicitly, we assume that active exposure to the treatment starts with the implementation of the reform (i.e., there are no anticipation effects).

unknown coefficients.[22] The term $\delta_t$ denotes an unobserved common time factor, $\boldsymbol{\lambda}_t$ a $(1 \times F)$ vector of unobserved time effects (i.e., factors) and $\boldsymbol{\mu_i}$ a $(F \times 1)$ vector of unobserved unit-specific effects (i.e., factor loadings). $\varepsilon_{it}$ is a unit- and time-specific error term. Equations (1) and (2) together imply that the observed outcome is given by[23]

$$Y_{it} = \tau_{it} D_{it} + \boldsymbol{\theta}_t \mathbf{Z}_i + \delta_t + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{it}\,.$$

Second, we assume that the unit- and time-specific error term, $\varepsilon_{it}$, is independent and identically distributed across $i$ and $t$ and that

$$\mathbb{E}(\varepsilon_{it} \,|\, \mathbf{Z}, \boldsymbol{\mu}) = 0\,, \tag{3}$$

with $\mathbf{Z} \equiv (\mathbf{Z}'_1, \ldots, \mathbf{Z}'_{J+1})'$ and $\boldsymbol{\mu} \equiv (\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_{J+1})'$.

Third, we assume that there exist weights $w_j^*$, $j = 2, \ldots, J+1$, with $w_j^* \geq 0$ and $\sum_{j=2}^{J+1} w_j^* = 1$, such that

$$\mathbf{Z}_1 = \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j \quad \text{and} \quad \boldsymbol{\mu}_1 = \sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu}_j \tag{4}$$

holds. Eq. (4) means that we can express the characteristics of the treated unit, $(\mathbf{Z}_1, \boldsymbol{\mu}_1)$, by a convex combination of the characteristics of the control units. For the case of an infinitely large population, Gobillon and Magnac (2016) show that Eq. (4) holds in expectation if the support of the characteristics of the treated unit is a subset of the support of the character-

[22] If these characteristics vary over time, the usual procedure is to use their averages over time. However, Klößner and Pfeifer (2015) develop an extended SCM approach allowing to incorporate whole time series of such characteristics.

[23] The SCM generalizes the Difference-in-Differences approach that does not allow for an interaction between unobserved time- and unit-specific effects but restricts $\boldsymbol{\lambda_t}$ to be constant over time. The SCM also has some similarity with the interactive fixed effects model studied in Bai (2009), for example. However, the SCM is more general than the interactive fixed effects model, as the SCM does not estimate Eq. (2) but Eq. (6), which would also be unbiased if the underlying model for $Y_{it}(0)$ included no factors but potentially lagged values of the dependent variable (Abadie et al., 2010). Moreover, the SCM with cross-section data at the individual level can also be viewed as a generalization of regression and matching estimators that recover causal effects under common support and conditional independence assumptions (see Abadie and L'Hour (2019)).

istics of the comparison units (see Gobillon and Magnac, 2016, Lemma 1).[24] In practice, the factor loadings, $\boldsymbol{\mu}$, are unobserved and the task consists in finding weights $w_j^*$, with $w_j^* \geq 0$ and $\sum_{j=2}^{J+1} w_j^* = 1$, satisfying

$$\mathbf{Z}_1 = \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j \,, \ Y_{11} = \sum_{j=2}^{J+1} w_j^* Y_{j1} \,, \ldots, \text{ and } Y_{1T_0} = \sum_{j=2}^{J+1} w_j^* Y_{jT_0} \,. \tag{5}$$

Because matching on pre-treatment outcomes is not the same as matching on $\boldsymbol{\mu}$ directly, it is important in applications to select the units to be included in the donor pool carefully so that each comparison unit is likely to have similar values of the observed and the unobserved characteristics as the treated unit (Abadie et al., 2010).

Under the assumptions stated in Eq. (2) through (4), an unbiased estimator of the treatment effect of unit one is given as

$$\hat{\tau}_{1t}^* = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \quad \text{for} \quad t > T_0 \tag{6}$$

(Abadie et al., 2010). For the implementation of the SCM estimator, we define a $(K \times 1)$ column vector $\mathbf{X}_1 = (\mathbf{Z}_1', \overline{Y}_1^1, \ldots, \overline{Y}_1^M)'$ containing the (average) values of the observable characteristics and $M$ linear combinations of the pre-intervention outcomes for the treated unit.[25] Each of these $M$ linear combinations is computed as $\overline{Y}_1^m = \sum_{t=1}^{T_0} l_t^m Y_{1t}$, where $l_t^m$ represents a weight and $m = 1, \ldots, M$. We construct an analogous $(K \times J)$ matrix $\mathbf{X}_0$ for the control units. Specifically, we use for $\mathbf{Z}_i$ pre-treatment averages of the population size and density, the unemployment rate, GDP per capita, household income (in 1,000 Euro), the student-teacher ratios in the academic and intermediate tracks, the share of foreign students aged 11-15, and the number of academic-track and intermediate-track schools per 100,000

[24] Similarly, Abadie and L'Hour (2019) impose a nested support condition stating that the probability measure of the treated characteristics is absolutely continuous with respect to that of the comparison units in the population.

[25] For a discussion on the use of pre-treatment outcomes in the SCM context, see Kaul et al. (2015), who show that the weights for the observable characteristics will be zero if all pre-intervention outcomes are included separately.

inhabitants. Moreover, we use the last pre-treatment value of all outcome variables (i.e., all three transition rates and all track-specific repetition rates), i.e., $M = 1$ and weights $l_t = 0$ for all $t < T_0$ and $l_{T_0} = 1$ for all outcome variables.

The standard SCM method then finds the vector $\mathbf{W} \equiv (w_2, \ldots, w_{J+1})'$ that minimizes $\sqrt{(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})'\mathbf{V}(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})}$, where $\mathbf{V}$ is a positive semi-definite diagonal matrix with dimension $(K \times K)$ (see Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015).[26] The elements of $\mathbf{V}$ are weights chosen to reflect the predictive power of the corresponding variable in $\mathbf{X}_1$.[27]

In our setup, we do not just have a single treated unit but 44 treated units in Baden-Wuerttemberg that are contrasted with the 94 control units in Bavaria. For the estimation, this means that we repeat the procedure outlined above for every treated unit in Baden-Wuerttemberg so as to obtain 44 estimates of $\hat{\tau}_{it}$ per post-reform year, where $i$ indexes the districts in Baden-Wuerttemberg. This allows us to derive a distribution of treatment effects in our treated federal state as well as an average treatment effect on the treated.

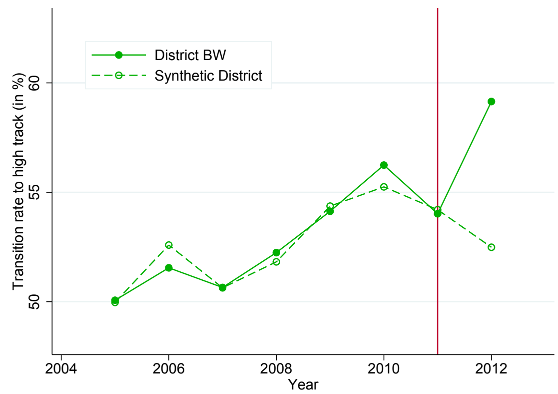
The idea of this disaggregated SCM is illustrated in Figure 6, showing in Subfigure (a) the results for the transition rate to the highest track in the first post-treatment year 2012 in Freiburg, a randomly picked district in Baden-Wuerttemberg. The synthetic control for Freiburg, constructed from the 94 districts in Bavaria, almost perfectly fits the pre-treatment time series of the transition rate to the highest track. In the post-treatment period, the actual and the synthetic transition rates for Freiburg diverge. We repeat the procedure for the other 43 districts in Baden-Wuerttemberg to obtain a distribution of treatment effects as shown in Subfigure (b) of Figure 6. Although the distribution resembles a normal one (depicted in dashed grey), our method does not hinge on any distributional assumptions.

To obtain an estimate of the average effect in the treated state of Baden-Wuerttemberg, we follow Acemoglu et al. (2016) and calculate a weighted average of the district-level ef-

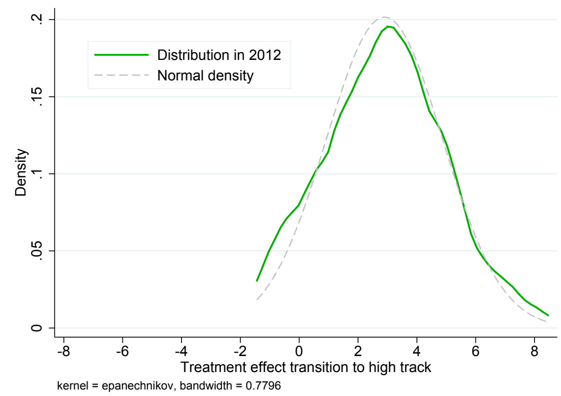
[26]For this purpose, we use the Stata command ‘synth’ that accompanies Abadie et al. (2010) (see also Abadie et al., 2011).

[27]The most common approach is to choose the values of $\mathbf{V}$ that maximize the in-sample fit for the pre-intervention outcomes of the treated unit.

Figure 6: Illustration of the Disaggregated SCM Approach




(a) One district




(b) Distribution of treatment effects

*Note:* Subfigure (a) displays the time series for the transition rate to the highest school track for one exemplary district in Baden-Wuerttemberg (Freiburg) and its synthetic control. Subfigure (b) plots the distribution of treatment effects as well as the normal density. BW is the abbreviation for Baden-Wuerttemberg. *Source:* Own calculations based on various data sources. See Appendix A.

fects, assigning a higher weight to districts with a better pre-treatment fit between actual and synthetic outcome values. This makes sense as the post-treatment difference between the actual and synthetic unit is more informative on the causal treatment effect if we are better able to predict the outcome of interest during the estimation frame.[28] This weighted treatment effect on the treated (WTT) is given as

$$\widehat{\mathrm{WTT}}_t = \sum_{i=1}^{S} \hat{\tau}_{it} \frac{1/RMSPE_i}{\sum_{s=1}^{S} 1/RMSPE_s}, \tag{7}$$

where $S$ denotes the number of treated districts and $RMSPE$ denotes the root mean squared prediction error (RMSPE) in the pre-treatment period. The higher the RMSPE, the lower the weight of the respective district-level treatment effect. Analogously, we compute the weighted standard deviation of the treatment effects.

Our disaggregated approach that computes synthetic controls at the district level has several advantages over an aggregate approach that first averages across districts and then

[28]This is similar in spirit to a matching approach that calculates average treatment effects using only the observations in the area of common support.

finds synthetic controls at the state level. Zooming in at the district level, we can exploit the larger number and heterogeneity of districts in Bavaria to obtain for each of the 44 treated districts in Baden-Wuerttemberg a sparse synthetic control consisting of control districts that are each very similar to the treated district under consideration. This should reduce the bias and improve the precision of our estimates. At the state level, we would have at most four states that could serve as the donor pool for the synthetic control (see Figure 1). While Bavaria is very similar to Baden-Wuerttemberg in economic and political terms, the three other comparison states are located in East Germany, and one may doubt whether they are sufficiently similar to Baden-Wuerttemberg. A further advantage of the disaggregated approach is that we can use the district-level treatment effects to study distributional effects and effect heterogeneity. As a potential disadvantage, the risk of overfitting and bias increases with the number of control units (Abadie et al., 2010). In fact, interpolation bias may be large if the synthetic control for a treated unit is constructed from control units that are individually very dissimilar (Abadie, 2019; Abadie and L'Hour, 2019). This could be an issue at the disaggregate level if the matching of control units to treated units by the SCM algorithm is not monitored by the researcher. However, in our case, the number of treated districts is still small enough in order to verify for all cases (44 treated districts times 11 outcome variables) that the solution of the SCM algorithm is reasonable in the sense that the synthetic control for a treated district is a sparse combination of comparison districts with similar values of the predictor variables. Furthermore, we examine the support of the predictor variables and check the pre-treatment fit of the outcome trajectory (i.e., verify that Eq. (5) holds approximately). We discuss the supporting evidence on matching quality in Section 5.3.

## 4.2 Inference

For inference, we randomly reassign the treatment across pre-treatment periods, an approach that Abadie et al. (2010, 2015) term *in-time placebo tests*. In our setup, relying on in-time permutations has the advantage that they correspond to classical randomization inference (Fisher, 1971; Imbens and Rubin, 2015).[29] The basic idea of this inferential framework is that, under random assignment of the treatment across time periods and under the sharp null hypothesis of a zero treatment effect for every unit in every time period subject to the random assignment, all time periods are exchangeable and the potential treatment and non-treatment outcomes are observable for all units. Thus, one can permute the treatment status across time periods in order to generate the exact finite-sample distribution of treatment effects under the null. If the treatment effects computed under the actual treatment assignment lie in the tails of the placebo distribution of treatment effects generated from the in-time permutations of treatment status, then this is evidence against the null hypothesis.

We prefer the in-time permutations over permutations in space because, in our study, we have a clustered design in which the treatment is assigned at the state and not at the district level. Therefore, we do not believe that permuting treatment status across districts, regardless of the state to which they belong, would be the right way to conduct inference.

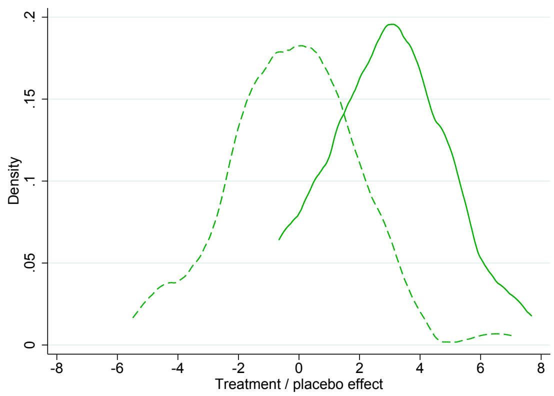
Hence, we permute the treatment status for all districts in Baden-Wuerttemberg as a whole across pre-treatment periods $t = 1, \ldots, T_0$. To generate the placebo distribution of treatment effects in Baden-Wuerttemberg, we compute the synthetic control estimates of the treatment effects for the pre-treatment school years beginning in fall 2007 through fall 2011.[30] The distribution of treatment effects for the respective post-reform school year can be compared to the placebo distribution of treatment effects generated using the pre-treatment years as shown in Figure 7 for the first post-reform year 2012. Further, we compute the Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1948) to test for equality of the actual and placebo treatment effect distributions.

Similarly, we compute the WTT for each of the pre-reform years 2007 through 2011, as

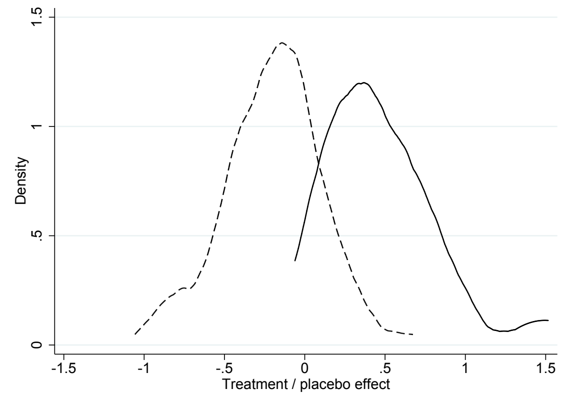
[29] See also Abadie (2019) for a discussion of different inferential approaches that can be used with the synthetic control method.

[30] Recall that the reform was passed in late 2011 and our observation period starts in the school year beginning in fall 2005. We cannot use all pre-treatment periods for the permutation exercise, as one needs to have two periods predating the respective placebo treatment period to run the SCM.

Figure 7: Treatment and Placebo Distributions for Transition and Repetition Rates, 2012




(a) Transition rate to high track




(b) Overall grade repetition

*Note:* The figure displays the kernel density of the treatment effects under the actual treatment assignment (solid line) and the placebo treatment assignment (dashed line) for selected outcomes of the reform in Baden-Wuerttemberg in school year 2012. Subfigure (a) refers to the transition rate to the high track (scaled between 0 and 1) and Subfigure (b) to the overall repetition rate in grade five.
*Source:* Own calculations based on various data sources. See Appendix A.

well as for one of the post-reform years and determine the percentile rank of the post-reform WTT in this distribution.[31] If the WTT in the respective post-reform year is positive and has percentile rank 0.917, which is the highest possible percentile, we can reject the exact null hypothesis of a zero effect for every unit in every year at the 10 percent significance level in a one-sided test (i.e., the $p$-value is 0.083). As an alternative inferential exercise for the WTT, we also compute the $p$-value for a two-sided test of the null hypothesis of a zero WTT under the assumption that the district-level treatment effects are normally distributed, which implies a $t$-distribution for the standardized estimator of the WTT.

[31] We compute the percentile rank of a unit as $(i - 0.5)/N$, where $i$ denotes the rank of the unit when the observations are sorted in ascending order and $N$ the total number of observations.

# 5 Causal Effects of School Tracking Reform

## 5.1 Main Results

Besides illustrating inference in our disaggregated SCM approach, Figure 7 also gives a first visual impression of the distribution of treatment effects for the school year beginning in fall 2012, which is the first post-reform year. Subfigure (a) refers to the transition rate to the high track, while Subfigure (b) corresponds to the grade repetition rate in grade five, the first grade of secondary school. For both outcomes, the treatment effect distribution lies clearly to the right of the placebo distribution generated from in-time permutations of treatment assignment, which suggests positive average treatment effects.

Table 3 shows the results on the weighted average treatment effect on the treated (WTT) (as defined in Eq. (7)) for the transition rates to the three school tracks. According to Panel (a), the reform increases the transition rate to the high track by 3.2 percentage points on average, which is an increase of 8 percent from the level in 2011, the last pre-treatment year. The $p$-value for a two-sided test of a zero WTT, assuming normally distributed district-level treatment effects, suggests that the WTT is statistically significant at any level. However, as normality is a strong assumption, we prefer to rely on the more conservative $p$-values for a one-sided test of the exact null hypothesis that the WTT is zero in every year generated from the in-time permutations of treatment assignment.

Table 3: Effects on Transition Rates

| | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|
| (a) High Track | | | | | |
| **Inference on the average treatment effect** | | | | | |
| WTT (N=42) | 3.150 | 3.646 | 3.121 | 2.699 | 2.484 |
| Standard deviation (weighted) | 1.901 | 2.190 | 3.251 | 2.986 | 2.958 |
| $p$-value, normally dist. district-level treatment effects | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $p$-value, in-time permutations | 0.083 | 0.083 | 0.083 | 0.083 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| (b) Mid Track | | | | | |
| **Inference on the average treatment effect** | | | | | |
| WTT (N=44) | 1.941 | 1.031 | -1.320 | -3.112 | -1.650 |
| Standard deviation (weighted) | 3.289 | 3.926 | 3.668 | 5.413 | 5.002 |
| $p$-value, normally dist. district-level treatment effects | 0.000 | 0.044 | 0.011 | 0.000 | 0.017 |
| $p$-value, in-time permutations | 0.083 | 0.417 | 0.417 | 0.250 | 0.250 |
| **Inference on the distribution of treatment effects** | | | | | |
| $p$-value, equality of distributions | 0.000 | 0.001 | 0.379 | 0.001 | 0.310 |
| (c) Low Track | | | | | |
| **Inference on the average treatment effect** | | | | | |
| WTT (N=44) | -6.968 | -11.201 | -12.432 | -13.603 | -16.114 |
| Standard deviation (weighted) | 1.953 | 2.639 | 3.500 | 3.344 | 3.820 |
| $p$-value, normally dist. district-level treatment effects | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $p$-value, in-time permutations | 0.083 | 0.083 | 0.083 | 0.083 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

*Note:* This table reports the results of the SCM estimation for the three transition rates. Starting with the first treatment year 2012, the table reports effects up to five years after the intervention. Rows labeled "WTT" show the weighted average treatment effect on the treated defined in Eq. (7), rows labeled "$p$-value, in-time permutations" the $p$-value from a one-sided test of the exact null hypothesis that the WTT is equal to zero based on in-time permutations of treatment status. The row labeled "$p$-value, equality of distributions" refers to the $p$-value of a Kolmogorov-Smirnov test of equality of actual and placebo treatment effect distribution.
*Source:* Own calculations based on various data sources. See Appendix A.

For the transition rate to the high track, this $p$-value amounts to 0.083, which is the lowest possible value given that we have five pre-treatment periods to conduct the in-time permutations. The two-sample Kolmogorov-Smirnov test for equality of the actual and placebo distributions of treatment effects also gives us a $p$-value of zero, which confirms a distinct difference between the distributions of actual and placebo treatment effects. A comparison of the estimates across post-reform years shows that the WTT on the transition rate to the high track increases from the first to the second post-reform year, declines thereafter but is with 2.5 percentage points still significantly positive five years after the reform (school year 2016/17). For the second outcome of interest, the transition rate to the intermediate track, shown in Panel (b) of Table 3, the WTT of the first post-treatment year is 1.9 percentage points with a $p$-value of 0.083 according to the in-time permutation test. This would translate, when compared with the pre-reform baseline, to a very small increase by 0.05 percent. However, as mentioned before, there are two offsetting changes occurring simultaneously at this school track in response to the reform, which can explain this rather small overall effect. As a result of the reform, students who would, without the reform, attend the intermediate school track now prefer the high track and, at the same time, students who would, without the reform, attend the low track now choose the intermediate track. The two-sample Kolmogorov-Smirnov test again confirms that actual and placebo distributions are distinct. The comparison of reform impacts across post-intervention years shows that the effect stays constant in 2013, albeit not significant anymore, and turns negative thereafter, with the different modes of inference showing mixed results.

Thirdly, as shown in Panel (c) of Table 3, the reform has a clear, negative effect on the transition rate to the low track. In the first post-treatment year the WTT is $-7.0$ percentage points, corresponding to a decrease on the pre-treatment level by 29 percent. The effect grows stronger in the following post-intervention years, suggesting a decrease by 46 and 51 percent in 2013 and 2014, respectively. This means that within three years after the policy change, the student body that flows from primary school to the lowest secondary school track is cut in half. All three modes of inference uniformly suggest that these effects are statistically significant.

Tables 4 and 5 display the WTTs on the repetition rates in grades five and six, the first two grades of secondary school. In both tables, Panel (a) displays the results for the overall repetition rate as well as the track-specific rates in Panels (b) to (d). Students who are retained at the academic or intermediate track may decide to change tracks and to repeat the grade at the next lower track. These cases are not included in our outcome measures. Therefore, the effects on grade repetition rates discussed here represent a lower bound of the effect on grade retention (also recall Figure 5 in Section 3).

In 2012, the first post-intervention year, the overall repetition rate in grade five increased by 0.43 percentage points, up from 0.51 percent in 2011. Thus, across all tracks, the incidence of grade repetition increased by 85 percent, i.e., it almost doubled. Moreover, this repetition rate remained similarly elevated in all post-intervention years, suggesting a permanent increase in the misallocation of students to tracks as a consequence of the abolition of mandatory teacher recommendations. These effects are highly significant across all modes of inference. Turning to the track-specific repetition rates, the results suggest that the increase in the overall repetition rate is driven by a higher incidence of grade repetition in the high and intermediate tracks, which is to be expected if, after the reform, parents make use of their right to overturn the teacher's recommendation and send their relatively low-ability child to a higher track than recommended. In relative terms, both repetition rates increase by around 110 percent from 2011 to 2012. While the increase is stable over time and clearly significant for the intermediate track, the pattern is more volatile for the high track.

Overall, these patterns are consistent with the effects on the transition rates. The high track receives students of worse academic ability who, without the reform, would attend one of the two lower tracks. This results in an increase in the repetition rate in grade five. However, this increase only constitutes a lower bound of the effect on grade retention in this track because those retained students who decide to move to a lower track and repeat the grade there are not included in the repetition rate of the high track. Similarly, the intermediate track now loses students to the high track and receives systematically weaker students from the low track. This seems to translate into a substantive increase in grade retention. The repetition rate for this track may still underestimate somewhat the effect on grade retention at this track, because repeaters who continue at the low track are not included.

Table 4: Effects on Repetition Rates Grade Five

| | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|
| (a) All Tracks | | | | |
| **Inference on the average treatment effect** | | | | |
| WTT (N=44) | 0.433 | 0.445 | 0.428 | 0.632 |
| Standard deviation (weighted) | 0.298 | 0.391 | 0.469 | 0.475 |
| $p$-value, normally distributed district-level treatment effects | 0.000 | 0.000 | 0.000 | 0.000 |
| $p$-value, in-time permutations | 0.083 | 0.083 | 0.083 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 | 0.000 |
| (b) High Track | | | | |
| **Inference on the average treatment effect** | | | | |
| WTT (N=43) | 0.285 | -0.000 | -0.128 | 0.294 |
| Standard deviation (weighted) | 0.276 | 0.418 | 0.525 | 0.524 |
| $p$-value, normally distributed district-level treatment effects | 0.000 | 0.499 | 0.059 | 0.000 |
| $p$-value, in-time permutations | 0.083 | 0.750 | 0.250 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | |
| $p$-value, equality of distributions | 0.000 | 0.079 | 0.001 | 0.000 |
| (c) Mid Track | | | | |
| **Inference on the average treatment effect** | | | | |
| WTT (N=44) | 0.659 | 0.824 | 0.773 | 0.666 |
| Standard deviation (weighted) | 0.788 | 0.780 | 0.866 | 0.923 |
| $p$-value, normally distributed district-level treatment effects | 0.000 | 0.000 | 0.000 | 0.000 |
| $p$-value, in-time permutations | 0.083 | 0.083 | 0.083 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 | 0.000 |
| (d) Low Track | | | | |
| **Inference on the average treatment effect** | | | | |
| WTT (N=43) | 0.057 | 0.042 | 0.002 | 0.879 |
| Standard deviation (weighted) | 0.592 | 0.604 | 0.981 | 1.948 |
| $p$-value, normally distributed district-level treatment effects | 0.267 | 0.324 | 0.495 | 0.003 |
| $p$-value, in-time permutations | 0.250 | 0.250 | 0.417 | 0.083 |
| **Inference on the distribution of treatment effects** | | | | |
| $p$-value, equality of distributions | 0.785 | 0.666 | 0.009 | 0.018 |

*Note:* This table reports the results of the SCM estimation for the repetition rates in grade five. Starting with the first treatment year 2012, the table reports effects up to four years after the intervention. Rows labeled "WTT" show the weighted average treatment effect on the treated defined in Eq. (7), rows labeled "$p$-value, in-time permutations" the $p$-value from a one-sided test of the exact null hypothesis that the WTT is equal to zero based on in-time permutations of treatment status. The row labeled "$p$-value, equality of distributions" refers to the $p$-value of a Kolmogorov-Smirnov test of equality of actual and placebo treatment effect distribution.
*Source:* Own calculations based on various data sources. See Appendix A.

Table 5: Effects on Repetition Rates Grade Six

| | 2013 | 2014 | 2015 |
|---|---|---|---|
| (a) All Tracks | | | |
| **Inference on the average treatment effect** | | | |
| WTT (N=44) | 0.304 | 0.321 | 0.284 |
| Standard deviation (weighted) | 0.453 | 0.537 | 0.564 |
| $p$-value, normally distributed district-level treatment effects | 0.000 | 0.000 | 0.001 |
| $p$-value, in-time permutations | 0.071 | 0.071 | 0.071 |
| **Inference on the distribution of treatment effects** | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 |
| (b) High Track | | | |
| **Inference on the average treatment effect** | | | |
| WTT (N=44) | 0.057 | -0.078 | 0.092 |
| Standard deviation (weighted) | 0.518 | 0.596 | 0.568 |
| $p$-value, normally distributed district-level treatment effects | 0.236 | 0.195 | 0.144 |
| $p$-value, in-time permutations | 0.071 | 0.786 | 0.071 |
| **Inference on the distribution of treatment effects** | | | |
| $p$-value, equality of distributions | 0.000 | 0.004 | 0.000 |
| (c) Mid Track | | | |
| **Inference on the average treatment effect** | | | |
| WTT (N=44) | 0.419 | 0.554 | 0.151 |
| Standard deviation (weighted) | 0.768 | 1.021 | 0.895 |
| $p$-value, normally distributed district-level treatment effects | 0.000 | 0.000 | 0.134 |
| $p$-value, in-time permutations | 0.071 | 0.071 | 0.071 |
| **Inference on the distribution of treatment effects** | | | |
| $p$-value, equality of distributions | 0.000 | 0.000 | 0.000 |
| (d) Low Track | | | |
| **Inference on the average treatment effect** | | | |
| WTT (N=44) | 0.075 | 0.019 | 0.447 |
| Standard deviation (weighted) | 0.634 | 0.948 | 0.961 |
| $p$-value, normally distributed district-level treatment effects | 0.219 | 0.447 | 0.002 |
| $p$-value, in-time permutations | 0.357 | 0.357 | 0.214 |
| **Inference on the distribution of treatment effects** | | | |
| $p$-value, equality of distributions | 0.089 | 0.587 | 0.006 |

*Note:* This table reports the results of the SCM estimation for the repetition rates in grade six. Starting with the first treatment year 2013, the table reports effects up to three years after the intervention. Rows labeled "WTT" show the weighted average treatment effect on the treated defined in Eq. (7), rows labeled "$p$-value, in-time permutations" the $p$-value from a one-sided test of the exact null hypothesis that the WTT is equal to zero based on in-time permutations of treatment status. The row labeled "$p$-value, equality of distributions" refers to the $p$-value of a Kolmogorov-Smirnov test of equality of actual and placebo treatment effect distribution.
*Source:* Own calculations based on various data sources. See Appendix A.

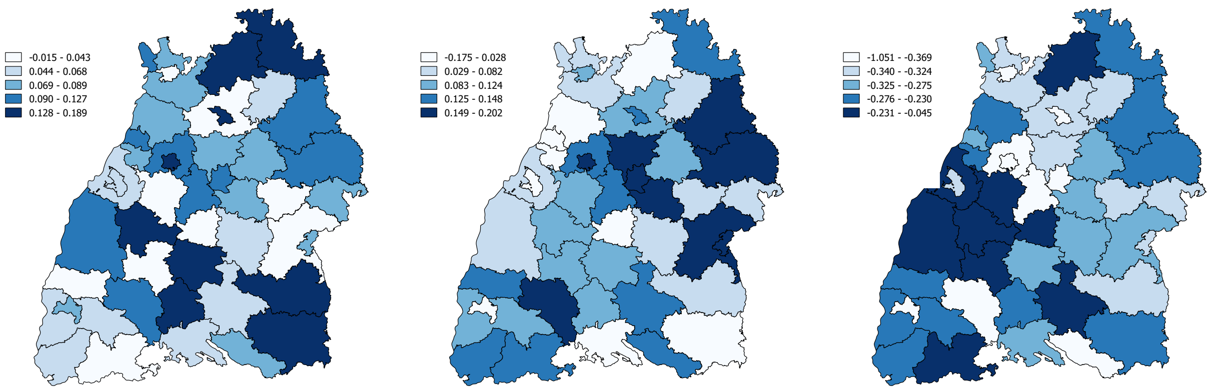
Finally, as apparent from Table 5, the reform also has an impact on grade repetition rates in grade six. In 2013, the year when the first cohort affected by the reform attended grade six, the overall repetition rate increased significantly by 0.30 percentage points, translating into a relative effect of 35 percent, and remained elevated in the two subsequent years. Again, the increase in overall grade repetition seems to have been driven by the high- and the mid-track schools.

In sum, the results of the estimations using the SCM corroborate the evidence laid out in the descriptive part. In fact, the primary school teacher's and the parents' assessment of the best-suited track for a child often diverge, with parents preferring a higher track than the teacher. With the abolition of mandatory teacher recommendations, parents increasingly make use of their right to send their child to a higher track than recommended by the teacher, which is reflected in the steep increase (decrease) in the transition rate to the high (low) school track. Although it may still be true that the children with a lower track recommendation benefit from attending the higher track in the long run, our results on the repetition rates in grades five and six point to a greater misallocation of students to the different tracks in the short run. Potential gains from school tracking stemming from more homogeneous peers and a more targeted teaching might therefore not realize fully in the lower grade levels of secondary schooling.

## 5.2 A Closer Look at Effect Heterogeneity

So far, we used the district-level effects to compute a (weighted) average treatment effect representative of Baden-Wuerttemberg as a whole. In this section, we now use the district-level treatment effects to examine effect heterogeneity in more detail. In this way, we aim to shed more light on the underlying mechanisms of the reactions to the reform. Figure 8, depicting the district-level treatment effects on the transition rates in 2012 (in percent of the level in 2011), suggests that the responses of the 44 districts to the reform are indeed heterogeneous. Compared with Figure 7, which already gave a first impression of treatment effect heterogeneity, we can now see the treatment effect distribution across space.

Figure 8: 2012 Treatment Effect Heterogeneity across Districts




(a) Transition rate high track (b) Transition rate mid track (c) Transition rate low track

*Note:* Subfigures (a) to (c) display the percentage treatment effect (as compared to pre-treatment baseline) from our SCM analyses across the 44 districts in Baden-Wuerttemberg for the transition rates to the high, intermediate, and low secondary school track, respectively.
*Source:* Own calculations based on various data sources. See Appendix A.

The heat maps visually confirm that the abolition of mandatory teacher recommendations had a varying impact on the transition rates across treated districts and also suggest a certain degree of regional clustering. Except for the transition rate to the high track, this can also formally be confirmed when we test for constant treatment effects (see Tables 3 to 5).

In order to better understand the pattern of effect sizes across districts, we regress the estimated SCM treatment effects on the same covariates we used for the descriptive analysis in Table 2. For inference, we rely again on in-time permutations of treatment assignment (as outlined in Section 4.2), which provides us with $p$-values for the exact null hypothesis of zero effects in a one-sided test. According to our hypothesis that the reform effects vary depending on the socioeconomic status of families, we should see stronger effects on the transition rates to secondary school tracks in districts with a higher average household income. The results of this exercise, as shown in Table 6, corroborate our earlier descriptive finding. Conditional on GDP per capita and population density, districts that have a higher average household income, i.e., socioeconomic status, show a stronger shift towards higher tracked schools.

Although there is no significant impact for the intermediate track, we also see a stronger reduction in the transition rate to the low track in districts with a higher average household income. Overall, this piece of evidence, again, confirms the hypothesis that higher SES families have a stronger incentive to deviate from the teacher's recommendation by choosing a higher track for their child than recommended.

Table 6: Regressions of District-Level Treatment Effects: Transition Rates

| | High Track | Mid Track | Low Track |
|---|---|---|---|
| Population density, in km$^2$ | 0.001 | -0.002 | -0.000 |
| | (0.250) | (0.417) | (0.583) |
| Household income, in 1000€ | 0.190* | -0.010 | -0.205* |
| | (0.083) | (0.750) | (0.083) |
| GDP p.c., in 1000€ | 0.007 | 0.001 | -0.027 |
| | (0.417) | (0.750) | (0.250) |
| Observations | 42 | 44 | 44 |

*Note:* This table shows the results of regressing the district-level treatment effects resulting from the SCM analyses on important economic covariates. *P*-values based on in-time permutations of treatment assignment shown in parentheses are computed analogously to the main results (see Section 4.2). * indicates significance at the 10% level in a one-sided test of the exact null hypothesis of zero effects.
*Source:* Own calculations based on various data sources. See Appendix A.

## 5.3 Matching Quality and Weights

As discussed in Section 4.1, similarity of the treated and comparison units in terms of their observed characteristics and pre-treatment outcomes is important in order to avoid biased estimation results. To check whether there is sufficient overlap of the distributions of the predictor variables, the last column of Table 1 compares the support of each variable between Baden-Wuerttemberg and Bavaria. The indicator is equal to one if $\min(x_{BW}) \geq \min(x_{BY}) - 0.1 \cdot \text{sd}(x_{BY})$ and $\max(x_{BW}) \leq \max(x_{BY}) + 0.1 \cdot \text{sd}(x_{BY})$, where $x$ denotes the respective predictor variable. With only four exceptions, the supports of the predictor variables in Bavaria are not markedly smaller than those in Baden-Wuerttemberg. In the vast majority of cases, the minimum and maximum values of the predictor variables in Bavaria are clearly more extreme than those in Baden-Wuerttemberg, and the standard deviations are always clearly larger in Bavaria.

Table 8 in Appendix B shows the average, minimum, median, and maximum number of Bavarian control districts with a nonzero weight for each of the 11 outcome variables. Depending on the outcome variable, the average number of control districts used to form a synthetic control ranges between 3.7 and 6.8 districts, and the maximum does not exceed 12 in all but one cases. In 0.8 percent of the cases (i.e., 4 out of $484 = 44$ treated districts $\times$ 11 outcome variables), the SCM algorithm was not able to construct a synthetic control from the Bavarian districts (see $N$ in Tables 3 to 5).[32] Thus, this evidence suggests that in nearly all cases, the SCM algorithm was successful in finding a sparse and unique combination of control districts. Another interesting feature to look at is the distribution of control unit weights, as depicted in Figure 11. For each outcome, the Bavarian control districts were grouped according to how often they were used in the estimation. Group "0" means that the district was never assigned a positive weight, while "1-2" means that the district was chosen up to two times, etc. As shown by the green bars, around 50 percent of all control districts in Bavaria were never used as a synthetic component (i.e., received a weight of zero in all 44 synthetic controls). Although there are some districts in Bavaria that were assigned a positive weight several times, it is not the case that one specific district was used excessively.

Table 9 in Appendix B displays the weighted average effects, computed according to Eq. (7), in the pre-treatment and the first post-treatment years. To facilitate interpretation, the effects are expressed relative to the value of the respective outcome variable in the last pre-treatment year (2011 or 2012). For outcome variables for which the effects are clearly significant in the post-treatment period (see Tables 3–5), the pre-reform placebo effects are typically much smaller than the actual treatment effects after the reform. Their magnitude exceeds that of the pre-reform placebo effects by a factor of 4 to 390 for the transition rates

[32] These are Kreisfreie Stadt Heilbronn and Landkreis Biberach for the transition rate to the academic track, Neckar-Odenwald-Kreis for the repetition rate in grade five at the academic track, and Landkreis Ravensburg for the repetition rate in grade five at the low track.

and 4 to 173 for the repetition rates in grade five. The pattern is somewhat less clear for the repetition rates in grade six though.

All in all, the descriptive statistics (Table 1), the evidence on the matching of Bavarian control districts to treated districts in Baden-Wuerttemberg (Table 8), and the evidence on pre-treatment fit (Table 9) suggest that the Bavarian districts constitute a well-suited donor pool for the districts in Baden-Wuerttemberg and that interpolation biases are unlikely to affect our estimated treatment effects in an important way.

Finally, Tables 10 to 12 in Appendix B show the weights of the predictor variables in the SCM estimations for the transition rates and grade repetition rates. The predictor weights indicate the relative importance of the variables used to predict the pre-treatment outcomes. Although the predictor weights may not be unique,[33] it is interesting to see that unlike in many previous studies applying the SCM, in our study, all predictor variables get a non-negligible predictor weight. Although the pre-treatment outcome has always the highest predictive power, it is not the case that its predictor weight is the only one that is important.

### 5.4 Potential Border Effects

As Baden-Wuerttemberg and Bavaria have a common border, a concern with our setup might be that, in Bavarian districts close to the border to Baden-Wuerttemberg, parents could respond to the reform in the neighboring state. In fact, parents living in a border district in Bavaria could take advantage of the reform in Baden-Wuerttemberg by sending their child to a higher-than-recommended secondary school track in nearby Baden-Wuerttemberg. To examine whether this kind of sorting could affect our results, we conduct three different robustness checks. First, we restrict the analysis to all non-bordering districts of Baden-Wuerttemberg (34), using all non-bordering districts of Bavaria as the donor pool (83). The results based on this restricted sample are very similar to those obtained based on the full sample. Second, we include in our analysis all 44 districts of Baden-Wuerttemberg in the

[33] Different predictor weights may produce the same control unit weights and, consequently, the same treatment effects.

treatment group but only keep the non-bordering districts of Bavaria in the comparison group. Here, again, our results are stable. Third, we conduct a falsification test taking the bordering districts of Bavaria as treated and the non-bordering Bavarian districts as comparison group. Clearly, if there is no endogenous sorting taking place, we expect to observe treatment effect estimates that are zero on average for all outcomes. Indeed, we find that the shares of children in the bordering districts transferring to the three secondary school tracks are the same than in the non-bordering districts of Bavaria.

# 6 Conclusion

This study investigates the role of parents in secondary school track choice, exploiting the natural experiment of a reform in the German federal state of Baden-Wuerttemberg. The reform granted parents the right to freely choose among three secondary tracks. While, before the reform, the teacher's recommendation on the highest secondary school track suitable for the child was binding for the parents, after the reform the recommendation became purely informational. We study the effects of this reform on transition rates to secondary school tracks and grade repetition rates.

There are two reasons why the reform might have impacted on these outcomes. On the one hand, parents could know their child's ability better than the teacher and choose a higher-than-recommended, better-matching track for their child. If this is the case, the reform should impact positively on transition rates to the high and intermediate tracks and at the same time lead to no increases in grade repetition rates at the high track, and likely neither at the intermediate track, because the students who upgrade are not systematically worse in their academic ability than those who attend the higher tracks as recommended by the teacher. On the other hand, parents may be primarily led by their social aspirations or might misperceive their child's ability. If this is the case, the reform should have heterogeneous effects on transition rates depending on the socioeconomic status, with high SES families having a stronger incentive to deviate from a lower track recommendation than low SES families. Track choice based on social aspirations or upward biased beliefs would deteriorate average student ability at all tracks, which in turn would translate into higher grade repetition rates.

Our findings suggest that, after the reform, parents of children with a recommendation for the two lower tracks increasingly overturned the teacher's advice, sending their child to the next higher track instead. Similarly, parents of children with a recommendation for the high track downgraded less often than before the reform. The behavioral responses of parents stand in contrast with the behavior of teachers who continued to recommend the same proportions of students to the three tracks after the reform as before. According to the disaggregated SCM estimates, the abolition of mandatory teacher recommendations has a significantly positive effect on the transition rates towards higher school tracks. When looking at heterogeneity of responses across districts, we see that especially the district's average household income drives the effect magnitude: on average, higher household income is correlated with a stronger behavioral response towards (away from) the high (low) secondary school track.

The downside of this movement towards more and higher education can be seen when quantifying the effect of the reform on the repetition rate in the first two grades after transition, for which we find that it almost doubled. Taking the results of this study at face value, it is possible to calculate the additional, one-period cost induced by the reform of secondary school tracking. In 2012, annual public expenditures per student in the high school track amounted to €6,700 (Statistisches Bundesamt, 2015). This figure includes staff expenditures (teachers, administrative staff, social security contributions of civil servants), running material expenses, and capital expenditures. With $104,321$ students in grade five in 2012 and an estimated effect of about 0.43 percentage points on the repetition rate, the reform of secondary school tracking cost the German state of Baden-Wuerttemberg approximately €2.8 million. However, this number can be interpreted only as a lower bound of the additional public schooling costs because it is based solely on the increased repetition rate in grade five. As our empirical results indicate, grade repetition rates also increase in grade six. For a complete cost-benefit analysis, such costs would have to be compared to the benefits of the reform, which are, among others, mirrored in the increased transition rates towards higher educational school tracks. However, as the reform took place only in 2011, it is not possible to conclusively assess how many of the additional students on the higher school tracks will eventually complete their respective tracks.

What our findings do strongly suggest, though, is that, although the reform surely eased the pressure on students, parents, and teachers in primary school, this pressure has just been shifted to the next educational level. Moreover, our findings confirm the hypothesis that social aspirations and biased beliefs of parents are an important determinant of their choice of secondary school track, even at the risk that their child has to repeat a grade in secondary school. Our results hence illustrate how early tracking may contribute to fostering social inequalities, i.e., when track choice is determined by social aspirations and biased beliefs rather than academic ability.

# References


Abadie, A. (2019). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, forthcoming.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2011). Synth: An R Package for Synthetic Control Methods in Comparative Case Studies. *Journal of Statistical Software*, 42(13):1–17.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.

Abadie, A. and L'Hour, J. (2019). A Penalized Synthetic Control Estimator for Disaggregated Data. Working paper, Department of Economics, Massachusetts Institute of Technology.

Abdulkadiroglu, A., Angrist, J., and Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1):137–196.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The Value of Connections in Turbulent Times: Evidence from the United States. *Journal of Financial Economics*, 121(2):368–391.

Bach, M. and Fischer, M. (2020). Understanding the Response to High Stakes Incentives in Primary Education. Working paper, Leibniz Centre for European Economic Research (ZEW), Mannheim.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279.

Betts, J. R. (2011). The Economics of Tracking in Education. In Hanushek, E., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, chapter 7, pages 341–381. Elsevier Science.

Canaan, S. (2020). The Long-Run Effects of Reducing Early School Tracking. *Journal of Public Economics*, 187:104206.

Card, D. and Giuliano, L. (2016). Can Tracking Raise the Test Scores of High-Ability Minority Students? *American Economic Review*, 106(10):2783–2816.

Dickert-Conlin, S., Elder, T., and Teltser, K. (2019). Allocating Scarce Organs: How a Change in Supply Affects Transplant Waiting Lists and Transplant Recipients. *American Economic Journal: Applied Economics*, 11:210–239.

Dube, A. and Zipperer, B. (2015). Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies. IZA Discussion Paper No. 8944, Institute for the Study of Labor (IZA).

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2):209–230.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). The Long-Term Effects of Early Track Choice. *The Economic Journal*, 172(603):1348–1380.

Fisher, R. A. (1971). *The Design of Experiments*. Hafner Publishing, New York, 9th edition.

Gobillon, L. and Magnac, T. (2016). Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls. *Review of Economics and Statistics*, 98(3):535–551.

Hall, C. (2012). The Effects of Reducing Tracking in Upper Secondary School: Evidence from a Large-Scale Pilot Scheme. *Journal of Human Resources*, 47(1):237–269.

Hanushek, E. A. and Wössmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116(510):C63–C76.

Imbens, G. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2015). Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictors. Working Paper, University of Hohenheim.

Kerr, S. P., Pekkarinen, T., and Uusitalo, R. (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3):577–602.

Klößner, S. and Pfeifer, G. (2015). Synthesizing Cash for Clunkers: Stabilizing the Car Market, Hurting the Environment. Working Paper, University of Hohenheim.

Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di Una Legge di Distribuzione. *Giornale dell' Istituto Italiano degli Attuari*, 4:83–91.

Leschinsky, A. and Mayer, K. U. (1990). *The Comprehensive School Experiment Revisited: Evidence from Western Europe*. Peter Lang, Frankfurt am Main.

Malamud, O. and Pop-Eleches, C. (2011). School Tracking and Access to Higher Education among Disadvantaged Groups. *Journal of Public Economics*, 95(11–12):1538–1549.

Matthewes, S. H. (2020). Better Together? Heterogeneous Effects of Tracking on Student Achievement. *The Economic Journal*, forthcoming.

Meghir, C. and Palme, M. (2005). Educational Reform, Ability, and Family Background. *The American Economic Review*, 95(1):414–424.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2):135–146.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.

Smirnov, N. V. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, 19(2):279–281.

Splawa-Neyman, J., Dabrowska, D., and Speed, T. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472.

Statistisches Bundesamt (2015). Bildungsausgaben - Ausgaben je Schülerinnen und Schüler 2012. Technical report, Wiesbaden.

Statistisches Bundesamt (2016). Bevölkerungsfortschreibung auf Grundlage des Zensus 2011. Technical report, Fachserie 1, Reihe 1.3, Wiesbaden.

Tamm, M. (2008). Does Money buy Higher Schooling?: Evidence from Secondary School Track Choice in Germany. *Economics of Education Review*, 27(5):536–545.

Waldinger, F. (2007). Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores? Unpublished manuscript, London School of Economics, London.

# Appendix

## A Data Appendix

This data appendix describes the data sources used, the processing of the original data into the final data set used for our main analyses, and the construction of our dependent and independent variables. All data used in this study were collected from several sources, which are, for most variables, freely available to the public (see Table A.2). The district-level identifier is the official German district code, which corresponds to the European NUTS-3 definition.

## A.1 Variable Description

Our analysis uses two sets of outcome variables and nine further control variables. All variables are measured at the district level. With regard to the availability of our outcome measures, the analysis employs data for the years 2005 to 2017. Note that we refer to the school years (which start in September and end in July) in the year they start. The school year 2009/2010, for instance, will be denoted as the year 2009.

### A.1.1 Dependent Variables

*Transition rates*: Our first set of dependent variables consists of the three transition rates to the respective secondary school type, that is, (i) the transition rate to the high, that is, academic school track (called "Gymnasium" in German), (ii) the transition rate to the intermediate school track (called "Realschule"), and (iii) the transition rate to the low school track (called "Hauptschule").

*Repetition rates*: Our second set of dependent variables is the respective repetition rates in grades five and six, as well as an overall repetition rate that sums the repeaters across all school tracks. The rationale for using the aggregate repetition rate across all tracks lies in the data structure: repeaters who change school tracks when they repeat the grade are counted in their new schools, which would lead to a misclassification (which would in any case underestimate the effect in the highest school track). When using the official statistic of the number of repeaters, one must also note that the data does not count the students in the year they fail class, but in the year they attend the respective grade for the second time. Thus, we use the lead of this variable to get the number of repeaters in that grade in the same year. However, for this reason, the treatment effects can be calculated only with data up to the school year 2015/16. This number is then related to the number of students in the grade (see next below) they failed, which is the most accurate measure given the data structure that we face.

*Number of students in respective grade*: The exact number of students per grade is used as the denominator of the repetition rates, our second set of dependent variables. If the number of students is missing, which concerns less than 1 percent of the sample, we use the following three methods to impute the missing value (in ascending order):

1. First, we calculate the number of students by using the information on the share of repeaters. From cross-checks, this is the most accurate method. However, if there are no repeaters, the share of repeaters is zero and this method fails.

2. If the first method fails, we use the number of transitions from grade four to grade five

and the number of repeaters in each grade to update the number of students year by year.

3. If both above-mentioned methods do not work, we impute the lagged value.

### A.1.2 Control Variables

*GPD per capita*: GDP per capita relates the GDP in the respective district (in EUR, nominal) to the number of its residents. The number of residents enters as the average number of persons living in this district, which is calculated based on the system of current population estimation in Germany.

*Number of schools per capita*: Number of schools per capita sets the number of schools (private and public) in relation to the population of the district. It is constructed for each of the main school types separately.

*Population density*: Population density is the number of individuals per district in relation to its area measured in square kilometers. Because there was missing area data in the years 2005 to 2007, we used the mean area for the year 2004 in the respective district for the denominator (which is the last, previous value).

*Student-teacher ratio*: The students-to-teacher ratio is calculated by dividing the number of students by the number of full-time equivalent teachers. Thus, the number of part-time teachers enters with a factor of 0.5. Again, the measure is school-type specific. For the years 2005 to 2008, there is no information on the number of teachers. We therefore impute the number of teachers in these years by using the share of teachers within the population in 2009 and projecting it to population figures in years with missing teacher information.

*Primary household income per resident*: The primary income of private households includes income from work and wealth received by domestic private households. This income includes, in particular, compensation of employees, self-employment income of sole proprietorships, and self-employed persons including remuneration for family workers, operating surplus from the production of owner-occupied housing services and net property income received.

*Share of foreign students*: Share of foreign students measures the share of the non-German population aged 11 to 15 within the district, that is, students at the beginning of secondary school.

*Unemployment rate*: Unemployment rate measures the number of unemployed individuals in terms of the entire civilian working population.

## A.2 List of Data Sources (alphabetical order)

Table 7: Data Sources

| Variable | Years Available | Data Source | Source Code |
|---|---|---|---|
| Number of students in BW | 2005-2017 | Official school statistic BW<br>www.statistik-bw.de<br>(last accessed 05.03.2018) | n/a |
| Number of students in BY | 2004-2017 | Bavarian statistics<br>www.statistik.bayern.de<br>(last accessed 05.03.2018) | n/a |
| Transition rates | 2005-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 03.01.2015) | D12.1i |
| Number of repeating students | 2005-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | D13.1i |
| Number of down-/upgaders | 2007-2016 | Statistical Offices of the States<br>www.statistischebibliothek.de<br>(last accessed 18.11.2020) | Fachserie / 11 / 1:<br>Table 3.7 |
| Population density | 2000-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | A01.4i |
| Unemployment rate | 2001-2017 | Federal Employment Agency<br>www.regionalstatistik.de/genesis<br>(last accessed 18.03.2015) | 13211-02-05-4 |
| GDP per capita | 2000-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | A02.1i |
| Number of teachers | 2009-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | D09.1i |
| Number of students | 2005-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | D07.1i |
| Share of Foreign Students | 2004-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | A01.3i |
| Number of schools | 2004-2017 | Statistical Offices of the States<br>www.bildungsmonitoring.de<br>(last accessed 22.02.2015) | D06.1i |
| Primary household income | 2004-2016 | National Account Systems<br>www.vgrdl.de<br>(last accessed 22.02.2015) | 1.3 |

*Note:* This table shows the data sources used in our analyses.

## A.3 Data and Sample Adjustments

### A.3.1 Bamberg and Schweinfurt – merging city and outer region

Bamberg city (with district identifier 9461) and Bamberg rural district (with district identifier 9471) share common schools, especially for the highest school track. Because the data do not allow one to disentangle the number of students commuting from the outer district to the school located in the city, we decided to merge these two Bavarian districts together. The same rationale holds for Schweinfurt city (with district identifier 9622) and Schweinfurt rural district (with district identifier 9671). To construct the respective dependent and independent variables, we either just build the sum of the districts (e.g., for the number of teachers, population, area) or scale the variables using the population size as weighting factor.

### A.3.2 Baden-Baden and Heidelberg – number of repeaters

The information on the number of repeaters in the intermediate school track for Baden-Baden (district identifier 8211) and Heidelberg (district identifier 8221) is missing for data privacy reasons. Because there are fewer than three intermediate track schools in Baden-Baden, the entries for both Baden-Baden and Heidelberg have been set to missing. For these two districts, we imputed the number of repeaters based on the difference in the number of repeaters in Baden-Wuerttemberg as a whole and the 42 districts with non-missing information. This difference is then assigned to Baden-Baden and Heidelberg according to their population size.

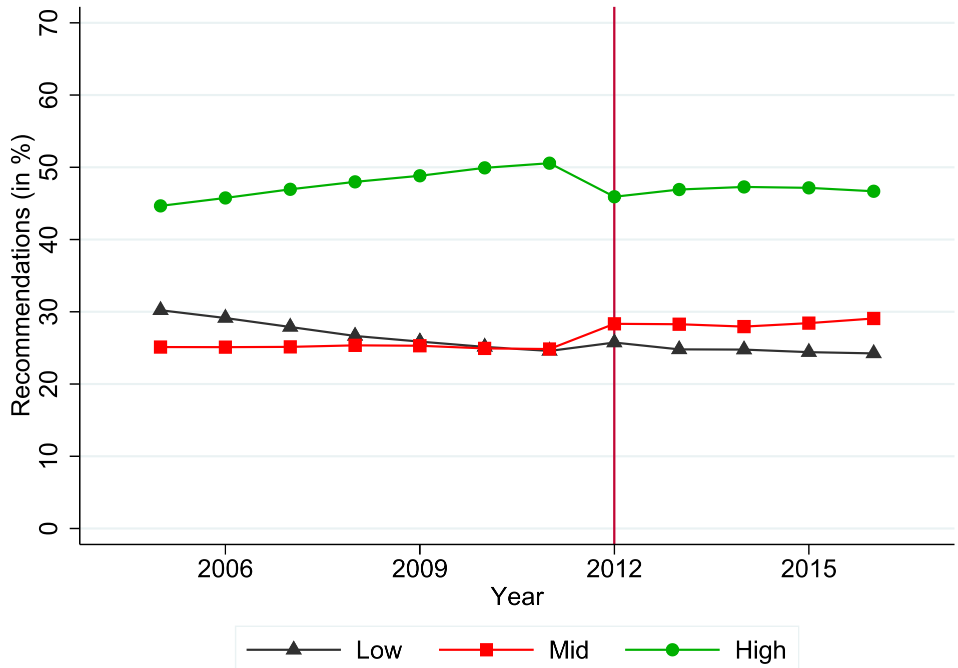
### A.3.3 Bavaria and the year 2014

The administrative data collection process in Bavaria for the school year beginning in fall 2014 was slightly different from that in other years. We therefore interpolate the values for 2014 with those of 2013 and 2015. The imputations affect the transition rates in 2014, the numerator of the repetition rates in 2013, and the denominator of the repetition rates in 2014.

# B Additional Descriptive Evidence and Estimation Results

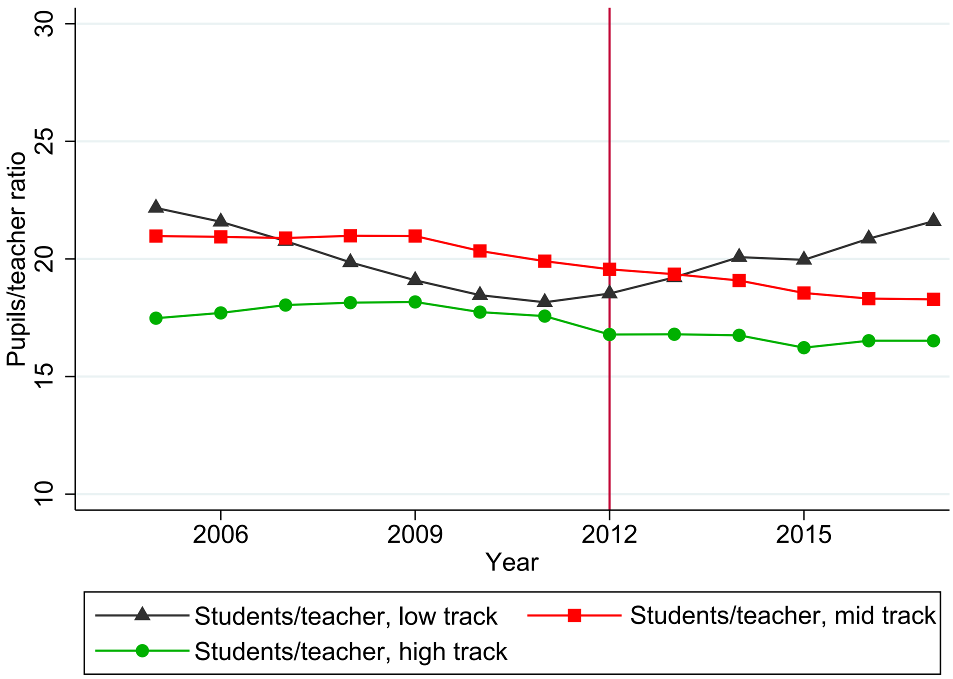
## B.1 Additional Descriptive Evidence

Figure 9: Evolution of Track Recommendations




*Note:* This figure plots the development of track recommendations by school type in percent of total recommendations over the sample period 2005-2016. The green-dotted line represents the share of recommendations to the highest school track, while the red-squared and the black-triangulated line show the respective time series for the mid and low track.
*Source:* Own calculations based on school statistics of Baden-Wuerttemberg.

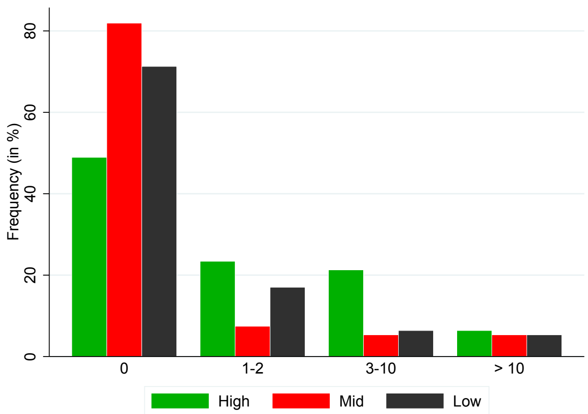
Figure 10: Evolution of Student-Teacher Ratios by Track Type



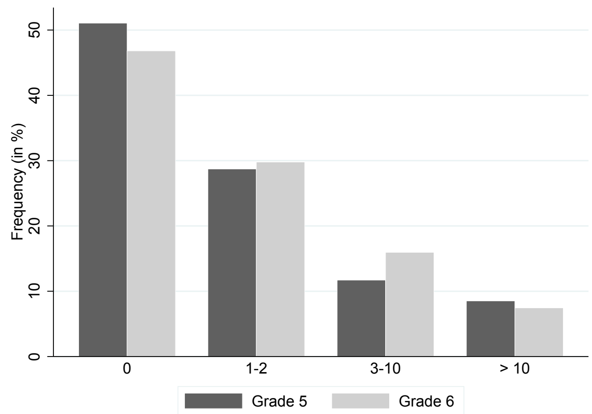

*Note:* The graph shows the evolution of student-teacher ratios in the academic track (light green connected dots), the intermediate track (red connected squares), and in the low track combined with primary schools (black connected triangles) in the sample period 2005-2016.
*Source:* Own calculations based on school statistics of Baden-Wuerttemberg.

## B.2 Additional Estimation Results

Figure 11: Control Unit Weights




(a) Transition rates




(b) Overall repetition rates

*Note:* Both figures show the distribution of control unit usage in our estimations. For each outcome, the Bavarian control districts were grouped according to how often they were used in the estimation. Group "0" means that the district was never assigned a positive weight, while "1-2" means that the district was chosen up to two times etc. Subfigure (a) plots the distribution for the three transition rates by school type and Subfigure (b) shows the distribution for the overall repetition rates in grades 5 and 6.
*Source:* Own calculations based on various data sources. See Appendix A.

Table 8: Number of Control Districts Forming a Synthetic Control

| | Mean | SD | Min | Median | Max |
|---|---|---|---|---|---|
| Transition Rate High Track | 5.976 | 2.018 | 3 | 6 | 11 |
| Transition Rate Mid Track | 3.659 | 1.180 | 2 | 4 | 7 |
| Transition Rate Low Track | 4.182 | 1.559 | 2 | 4 | 9 |
| Repetition Rate Grade 5 Overall | 5.591 | 2.453 | 2 | 5 | 11 |
| Repetition Rate Grade 5 High Track | 5.349 | 1.876 | 2 | 5 | 11 |
| Repetition Rate Grade 5 Mid Track | 5.955 | 1.509 | 3 | 6 | 9 |
| Repetition Rate Grade 5 Low Track | 7.233 | 5.080 | 3 | 6 | 37 |
| Repetition Rate Grade 6 Overall | 5.841 | 2.068 | 2 | 6 | 11 |
| Repetition Rate Grade 6 High Track | 4.659 | 2.112 | 1 | 4.5 | 10 |
| Repetition Rate Grade 6 Mid Track | 6.045 | 1.928 | 3 | 6 | 11 |
| Repetition Rate Grade 6 Low Track | 6.841 | 2.458 | 2 | 6.5 | 12 |

*Note:* This table shows the distribution of the number of districts used for the construction of the synthetic control districts for every outcome variable.
*Source:* Own calculations based on various data sources. See Appendix A.

Table 9: Average Effects in Pre- and Post-Reform Periods in Percent

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|
| (a) Transition Rate (Post-Reform: 2012, 2013) | | | | | | | | | |
| High Track | 0.95 | -0.17 | 0.34 | -0.08 | -1.42 | -0.02 | 0.28 | 7.80 | 9.02 |
| Mid Track | 2.01 | 3.80 | 6.87 | 11.12 | 8.57 | -5.38 | -2.00 | 5.69 | 3.02 |
| Low Track | -6.09 | 1.38 | -5.62 | -4.88 | -3.47 | 8.08 | 3.50 | -28.82 | -46.33 |
| (b) Repetition Rate Grade Five (Post-Reform: 2012, 2013) | | | | | | | | | |
| Overall | -7.01 | 0.33 | -2.80 | 1.00 | -5.40 | -21.86 | -8.60 | 84.96 | 87.39 |
| High Track | -4.85 | 0.63 | -12.51 | -14.41 | -2.25 | -11.81 | -12.07 | 109.05 | -0.03 |
| Mid Track | -6.69 | 9.08 | 7.76 | -5.99 | -5.78 | -9.33 | -19.50 | 108.45 | 135.67 |
| Low Track | -6.85 | -7.83 | -10.05 | -7.10 | -1.59 | -15.66 | 5.57 | 7.03 | 5.26 |
| (c) Repetition Rate Grade Six (Post-Reform: 2013) | | | | | | | | | |
| Overall | -5.38 | 7.69 | -7.91 | -16.6 | -17.33 | -2.23 | -23.69 | -5.65 | 30.06 |
| High Track | -9.61 | -11.83 | -8.11 | -17.02 | -6.35 | -23.92 | -26.91 | -10.28 | 6.78 |
| Mid Track | 11.32 | 4.37 | -8.32 | -3.74 | -5.30 | -2.90 | -25.03 | -13.27 | 30.73 |
| Low Track | -13.63 | -0.40 | -14.38 | -9.98 | -13.77 | 4.02 | -6.98 | 3.31 | 9.79 |

*Note:* This table shows weighted average effects as defined in Eq. (7) for the pre-treatment and the first post-treatment years. The effects are expressed in percent of the value of the respective outcome variable in the last pre-reform year (2011 or 2012).
*Source:* Own calculations based on various data sources, see Appendix A.

Table 10: Predictor Weights for Transition Rates

| | High Track | | Mid Track | | Low Track | | Average | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Transition rate, high track | 0.269 | 0.332 | 0.028 | 0.063 | 0.008 | 0.016 | 0.102 | 0.137 |
| Transition rate, mid track | 0.035 | 0.115 | 0.332 | 0.385 | 0.041 | 0.143 | 0.136 | 0.214 |
| Transition rate, low track | 0.009 | 0.016 | 0.022 | 0.044 | 0.242 | 0.359 | 0.091 | 0.140 |
| Rep. rate in 5, high track | 0.041 | 0.105 | 0.034 | 0.110 | 0.009 | 0.018 | 0.028 | 0.078 |
| Rep. rate in 5, mid track | 0.020 | 0.035 | 0.015 | 0.022 | 0.013 | 0.030 | 0.016 | 0.029 |
| Rep. rate in 5, low track | 0.045 | 0.154 | 0.015 | 0.036 | 0.031 | 0.061 | 0.030 | 0.084 |
| Rep. rate in 6, high track | 0.022 | 0.083 | 0.053 | 0.099 | 0.008 | 0.014 | 0.028 | 0.065 |
| Rep. rate in 6, mid track | 0.042 | 0.155 | 0.056 | 0.155 | 0.039 | 0.064 | 0.046 | 0.125 |
| Rep. rate in 6, low track | 0.086 | 0.234 | 0.072 | 0.193 | 0.009 | 0.020 | 0.056 | 0.149 |
| Population (in 1000) | 0.077 | 0.215 | 0.037 | 0.104 | 0.077 | 0.205 | 0.064 | 0.175 |
| Population density, in $km^2$ | 0.069 | 0.174 | 0.045 | 0.124 | 0.016 | 0.035 | 0.043 | 0.111 |
| GDP p.c. (in 1000 €) | 0.022 | 0.057 | 0.025 | 0.075 | 0.025 | 0.064 | 0.024 | 0.066 |
| Unemployment rate | 0.020 | 0.039 | 0.080 | 0.207 | 0.011 | 0.030 | 0.037 | 0.092 |
| Household income (in 1000 €) | 0.054 | 0.163 | 0.008 | 0.019 | 0.174 | 0.311 | 0.079 | 0.164 |
| Stud./teacher, high track | 0.039 | 0.130 | 0.014 | 0.031 | 0.044 | 0.152 | 0.032 | 0.104 |
| Stud./teacher, mid track | 0.038 | 0.086 | 0.023 | 0.039 | 0.045 | 0.091 | 0.036 | 0.072 |
| Share foreign stud. aged 11-15 | 0.074 | 0.202 | 0.017 | 0.075 | 0.089 | 0.193 | 0.060 | 0.157 |
| High track schools per 100,000 | 0.023 | 0.037 | 0.023 | 0.052 | 0.057 | 0.162 | 0.034 | 0.084 |
| Mid track schools per 100,000 | 0.013 | 0.040 | 0.100 | 0.196 | 0.062 | 0.081 | 0.058 | 0.106 |

*Note:* This table shows the means and standard deviations of the weights given to each of the predictors in the SCM estimations of the treatment effects on the transition rates. The columns labeled 'Average' refer to the average across all outcome variables shown in the table. Note that the predictor weights are not unique. Different predictor weights can produce the same control unit weights and consequently the same treatment effects.
*Source:* Own calculations based on various data sources. See Appendix A.

Table 11: Predictor Weights for Repetition Rates in Grade Five

| | Overall | | High Track | | Mid Track | | Low Track | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Transition rate, high track | 0.054 | 0.157 | 0.051 | 0.162 | 0.046 | 0.156 | 0.059 | 0.145 | 0.053 | 0.155 |
| Transition rate, mid track | 0.009 | 0.021 | 0.040 | 0.106 | 0.007 | 0.014 | 0.011 | 0.029 | 0.017 | 0.042 |
| Transition rate, low track | 0.020 | 0.055 | 0.039 | 0.144 | 0.059 | 0.194 | 0.064 | 0.186 | 0.046 | 0.145 |
| Rep. rate in 5, high track | 0.051 | 0.147 | 0.157 | 0.248 | 0.034 | 0.076 | 0.064 | 0.161 | 0.077 | 0.158 |
| Rep. rate in 5, mid track | 0.085 | 0.141 | 0.045 | 0.147 | 0.131 | 0.205 | 0.020 | 0.033 | 0.070 | 0.131 |
| Rep. rate in 5, low track | 0.064 | 0.172 | 0.016 | 0.063 | 0.066 | 0.208 | 0.051 | 0.131 | 0.050 | 0.144 |
| Rep. rate in 6, high track | 0.043 | 0.116 | 0.121 | 0.266 | 0.029 | 0.069 | 0.083 | 0.224 | 0.069 | 0.169 |
| Rep. rate in 6, mid track | 0.101 | 0.200 | 0.040 | 0.147 | 0.023 | 0.044 | 0.038 | 0.060 | 0.050 | 0.113 |
| Rep. rate in 6, low track | 0.024 | 0.051 | 0.025 | 0.037 | 0.030 | 0.098 | 0.036 | 0.076 | 0.029 | 0.066 |
| Population (in 1000) | 0.030 | 0.119 | 0.046 | 0.148 | 0.065 | 0.213 | 0.047 | 0.142 | 0.047 | 0.156 |
| Population density, in km$^2$ | 0.029 | 0.069 | 0.076 | 0.172 | 0.114 | 0.205 | 0.055 | 0.116 | 0.068 | 0.141 |
| GDP p.c. (in 1000 €) | 0.121 | 0.284 | 0.069 | 0.170 | 0.093 | 0.259 | 0.111 | 0.256 | 0.098 | 0.242 |
| Unemployment rate | 0.099 | 0.213 | 0.075 | 0.208 | 0.078 | 0.218 | 0.042 | 0.103 | 0.073 | 0.185 |
| Household income (in 1000 €) | 0.055 | 0.210 | 0.047 | 0.140 | 0.017 | 0.035 | 0.043 | 0.150 | 0.040 | 0.134 |
| Stud./teacher, high track | 0.052 | 0.175 | 0.057 | 0.205 | 0.076 | 0.241 | 0.051 | 0.165 | 0.059 | 0.197 |
| Stud./teacher, mid track | 0.090 | 0.219 | 0.016 | 0.026 | 0.039 | 0.144 | 0.081 | 0.212 | 0.056 | 0.150 |
| Share foreign stud. aged 11-15 | 0.003 | 0.004 | 0.024 | 0.057 | 0.061 | 0.164 | 0.019 | 0.040 | 0.027 | 0.066 |
| High track schools per 100,000 | 0.067 | 0.210 | 0.018 | 0.055 | 0.018 | 0.033 | 0.105 | 0.248 | 0.052 | 0.137 |
| Mid track schools per 100,000 | 0.003 | 0.007 | 0.039 | 0.085 | 0.013 | 0.050 | 0.020 | 0.086 | 0.019 | 0.057 |

*Note:* This table shows the means and standard deviations of the weights given to each of the predictors in the SCM estimations for the treatment effects on the repetition rates in grade five. The columns labeled "Average" refer to the average across all outcome variables shown in the table. Note that the predictor weights are not unique. Different predictor weights can produce the same control unit weights and consequently the same treatment effects.
*Source:* Own calculations based on various data sources. See Appendix A.

Table 12: Predictor Weights for Repetition Rates in Grade Six

| | Overall | | High Track | | Mid Track | | Low Track | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Transition rate, high track | 0.055 | 0.154 | 0.048 | 0.137 | 0.041 | 0.091 | 0.054 | 0.144 | 0.049 | 0.132 |
| Transition rate, mid track | 0.023 | 0.101 | 0.061 | 0.155 | 0.030 | 0.072 | 0.034 | 0.151 | 0.037 | 0.120 |
| Transition rate, low track | 0.035 | 0.111 | 0.053 | 0.141 | 0.056 | 0.155 | 0.034 | 0.154 | 0.044 | 0.140 |
| Rep. rate in 5, high track | 0.054 | 0.168 | 0.007 | 0.015 | 0.036 | 0.090 | 0.027 | 0.110 | 0.031 | 0.096 |
| Rep. rate in 5, mid track | 0.002 | 0.006 | 0.011 | 0.015 | 0.012 | 0.039 | 0.001 | 0.002 | 0.006 | 0.015 |
| Rep. rate in 5, low track | 0.010 | 0.032 | 0.008 | 0.025 | 0.112 | 0.240 | 0.027 | 0.084 | 0.039 | 0.095 |
| Rep. rate in 6, high track | 0.027 | 0.077 | 0.050 | 0.152 | 0.072 | 0.160 | 0.137 | 0.261 | 0.071 | 0.162 |
| Rep. rate in 6, mid track | 0.040 | 0.146 | 0.022 | 0.047 | 0.018 | 0.045 | 0.013 | 0.032 | 0.023 | 0.068 |
| Rep. rate in 6, low track | 0.014 | 0.057 | 0.012 | 0.040 | 0.040 | 0.116 | 0.066 | 0.209 | 0.033 | 0.105 |
| Population (in 1000) | 0.083 | 0.195 | 0.098 | 0.207 | 0.048 | 0.138 | 0.061 | 0.174 | 0.073 | 0.178 |
| Population density, in km$^2$ | 0.080 | 0.169 | 0.050 | 0.102 | 0.020 | 0.026 | 0.061 | 0.158 | 0.053 | 0.114 |
| GDP p.c. (in 1000 €) | 0.087 | 0.234 | 0.080 | 0.190 | 0.036 | 0.099 | 0.036 | 0.092 | 0.060 | 0.154 |
| Unemployment rate | 0.071 | 0.154 | 0.124 | 0.187 | 0.069 | 0.166 | 0.021 | 0.094 | 0.071 | 0.150 |
| Household income (in 1000 €) | 0.094 | 0.198 | 0.035 | 0.065 | 0.146 | 0.208 | 0.035 | 0.128 | 0.077 | 0.150 |
| Stud./teacher, high track | 0.061 | 0.154 | 0.120 | 0.249 | 0.116 | 0.207 | 0.074 | 0.175 | 0.093 | 0.196 |
| Stud./teacher, mid track | 0.046 | 0.161 | 0.027 | 0.071 | 0.014 | 0.032 | 0.047 | 0.165 | 0.033 | 0.107 |
| Share foreign stud. aged 11-15 | 0.120 | 0.257 | 0.079 | 0.181 | 0.045 | 0.114 | 0.142 | 0.300 | 0.097 | 0.213 |
| High track schools per 100,000 | 0.058 | 0.156 | 0.069 | 0.200 | 0.060 | 0.154 | 0.037 | 0.108 | 0.056 | 0.155 |
| Mid track schools per 100,000 | 0.040 | 0.150 | 0.048 | 0.155 | 0.030 | 0.086 | 0.093 | 0.234 | 0.053 | 0.156 |

*Note:* This table shows the means and standard deviations of the weights given to each of the predictors in the SCM estimations for the treatment effects on the repetition rates in grade six. The columns labeled "Average" refer to the average across all outcome variables shown in the table. Note that the predictor weights are not unique. Different predictor weights can produce the same control unit weights and consequently the same treatment effects. *Source:* Own calculations based on various data sources. See Appendix A.