Comparing Students by a Matching Analysis - On Early School Leaving in Dutch Cities

Kristof De Witte & Chris van Klaveren

TIER WORKING PAPER SERIES
TIER WP 14/10
Comparing Students by a Matching Analysis – On Early School Leaving in Dutch Cities

Kristof De Witte
TIER
Maastricht University
P.O. Box 616
6200 MD Maastricht, The Netherlands
and
KU Leuven
Naamsestraat 69
3000 Leuven, Belgium
kristof.dewitte@econ.kuleuven.be

Chris Van Klaveren
TIER
Maastricht University
P.O. Box 616
6200 MD Maastricht, The Netherlands
cp.vanklaveren@maastrichtuniversity.nl

Abstract

In case of regional discretionary on the implementation of policy measures, central governments often consider differences in outcomes as an indication that one policy was more effective than another policy. If uniform incentives are provided to motivate regional policy makers, these incentives can be discouraging when the underlying populations differ. Empirically, this study compares early school leaving between the four largest Dutch cities. It shows that considering regional differences as performance measures can be dangerous if differences in population characteristics are not properly taken into account. Methodologically, this study contrasts the use of a traditional probit model with a more advanced iterative matching procedure.

JEL Codes: C14, C61, C23, I21
Keywords: Uniform incentives; Early school leaving; Comparative; Matching analysis

1 Introduction

At the Lisbon 2000 summit, the European council decided to aim for a lower dropout rate, among other benchmarks. The average rate of early school leavers should be no more than 10% by 2012.\footnote{Dropout is also a major issue in other continents. For instance, U.S. president Obama mentioned in his inauguration speech: “Every American will need to get more than a high school diploma. And dropping out of high school is no longer an option”.

\footnote{We are grateful to two anonymous referees, Sofie Cabus, Wim Groot, Henriette Maassen van den Brink, and seminar participants at NICIS Institute and the Dutch Ministry of Education (OCW) for their constructive remarks. The authors acknowledge financial support of NICIS. The usual caveat applies.}}

The European Commission (2006) defines a ‘dropout’ (or early school leaver) as a young person...
(between 12 and 23 years old) who leaves secondary education without a diploma. To meet the Lisbon targets, central governments in all European countries developed various policy measures. A common factor in these policies is the subsidiarity principle: policy implementations are made at the decentralized level (usually the school level) and subsidies are given at the regional level.

If dropout policy is made at the regional level, differences across regions may appear in both policy and outcomes. Since central governments are interested in whether their money is well spend, it is natural that they focus on the regional differences in student dropout rate. It can be even rational for them to point to the dissimilarities in outcomes as this provides a ‘verbal’ incentive (in particular, naming and shaming) for the regions to perform better. It becomes tricky, however, when they consider these regional differences as performance measures. It is totally nonsense if incentives are based on the regional differences.

This paper aims to contribute to the literature from both a methodological perspective and from a policy point of view. First of all, and related to the methodological perspective, this paper compares the outcomes of two popular assessment procedures. Frequently, a probit or logit model is used to examine how regional differences in policy affect the outcome measure (in casu, dropout rates among students). In doing so, the researcher relates the outcome measure to a regional dummy, while ‘controlling’ for the heterogeneous underlying population. Although this analysis presumably measures the impact of policy differences (see, e.g., Allensworth, 2005), we point out in the analysis below, that this type of analysis could use an inappropriate reference or control group and, as a consequence, the analysis fails to account for differences in population characteristics properly. Therefore, to assess the influence of a policy measure, this study addresses the question: What if students living in one region would live in another region, how does this affect the outcome measure?. Angrist and Krueger (1999) mention that the most challenging empirical questions in economics involve similar “what if” statements about outcomes that are not observed. The “what if” statement is however necessary, in our case, because we are not interested in describing differences in outcome measures between regions, but are interested in how the outcome measure in one region compares to the outcome measure in another region. In the paper, the methodological contribution is explored using an iterative matching analysis. Matching techniques are acknowledged by Blundell and Costa Dias (2007, p.4) to be "a valuable part of the evaluation toolbox".

As a second contribution, the paper has a clear policy focus as it argues that one cannot consider straightforward differences in outcomes as performance measure, even for a uni-dimensional, simple and uniform indicator of performance (i.e., early school leaving of a student). Although this may seem common knowledge, in practice, many policy incentives are based on raw comparisons of outcomes. The effectiveness of regional policy does not only depend on the regional policy makers themselves, but also depends on the underlying population. If cities or regions are held
accountable for their performance on particular dimensions, central government should account for
regional heterogeneity in an appropriate manner.

The literature on regulation (e.g., Laffont and Tirole, 1993) shows that uniform incentives can
be demotivating. The principal (i.e., the central government) sets out an incentive for the agent
(i.e., the regional government). To keep the incentive as transparent as possible, the principal
determines a uniform incentive for all agents. If an \textit{a priori} specified target is obtained, the agent
receives the incentive. This uniform incentive is common practice in the health sector (e.g., for
quality targets), transport (e.g., keeping time tables), service centers (e.g., time to take a call), etc.

From the examples, it is clear that the uniform incentive is wide spread in policy. Nevertheless,
for agents (e.g., hospitals, transport companies, etc.) operating in different (geographical) regions,
obtaining the same target may be more requiring in a region with less advantageous characteristics
than in a region with more advantageous characteristics. Both ‘naming and shaming’ and mon-
etary rewards can therefore provide a negative motivation for the region with a disadvantageous
population.

The particular application focusses on the two largest cities in the Netherlands (although it is
extended in the appendix to a comparison between four large Dutch cities). In the Netherlands, the
Ministry of Education, Culture and Sciences (OCW) tries to meet the targets of the Lisbon Council
by a comprehensive dropout policy. The total budget spend on dropout prevention increased from
313 million euro in 2008 to 400 million euro in 2011. One of the policy measures consists of a
uniform monetary incentive of 2500 euro per early school leaver less in comparison to the base year
2005-2006. The Ministry of Education (OCW) allocated 5.4 % of this budget (i.e., 17.04 million
 euro) to this uniform incentive in 2008. This allocation increases to 11.4 % (i.e., 45.44 million
 euro) in 2011. It can be expected that regions with more disadvantageous population (e.g., more
difficult to reach) receive less subsidy (i.e., the 2500 euro per student less) than regions with a more
advantageous population.

The latter is the case in the cities of Amsterdam and Rotterdam. The city of Amsterdam
succeeds in reducing the dropout rate faster than the city of Rotterdam. This difference creates
two issues. On the one hand, it is a common believe among central government policy makers
that the difference in outcome (i.e., dropout rate) is thanks to a higher effectiveness of the policy
measures. For example, the differences in outcome are explained by the higher effectiveness of
the hands-on approach in Amsterdam (versus a hands-off approach in Rotterdam; De Bruijn et

\textsuperscript{2}As the examples below indicate, the implications of this paper are not restricted to the educational sector, but
can straightforwardly be extended to other sectors. To facilitate the use of the iterative matching procedure, the
Stata code is available upon request.

\textsuperscript{3}Following the official Dutch and European definition, we define an early school leaver (or dropout) as a person
younger than 23 who leaves education without a higher secondary education degree.
al., 2009, 2010; Section 2). If this is true, a uniform incentive is the most appropriate way to stimulate policy makers (Laffont and Tirole, 1993). On the other hand, schools and regional policy makers point to the heterogeneity in population. If populations significantly differ, it might be more difficult to obtain the targets with a disadvantageous population. If this is true, a uniform incentive provides a negative motivation.

This paper examines this scenario in detail for Amsterdam and Rotterdam. At first sight, both cities have many similar characteristics such that the uniform incentive seem appropriate. However, our matching results indicate that population characteristics are different, such that a uniform incentive can be demotivating. This paper benefitted from using an exceptionally rich registered data source for the year 2007, provided to us by the Dutch Ministry of Education. The data contain information on dropout status of all students in secondary education. Moreover, it includes information on several background characteristics on the school and student level. On the basis of these data we estimate a probit model and simulate how the dropout probability for students living in Rotterdam would have been different if they would have lived in Amsterdam.

After a detailed analysis of Amsterdam and Rotterdam, the application is (in Appendix) extended to two other large Dutch cities for which we obtained similar data: The Hague and Utrecht. We show that the differences between the cities are large, and that our results also yield for this extended sample.

The remainder of this paper is organized as follows. Section 2 explores the data at hand. We briefly outline the difference between the hands-on and hands-off approach. In Section 3 we present the results of a traditional probit model, which is often used to study regional differences in dropout rates. In Section 4, we focus on the methodological contribution as as we present and discuss the empirical results of the iterative matching analysis. In Section 5 we conclude. Finally, in the Appendix the data set is enlarged to two other large Dutch cities: Utrecht and The Hague.

2 Hands-on versus hands-off approach?

The data

Although student dropout is highly ranked on the political agenda in the Netherlands, as in other European countries, there did not exist an accurate estimate of the number of students dropping out of secondary education.\(^4\) Therefore, the Ministry of Education developed a tracking system for students. In this system, Dutch students receive a personal identification number which allows the central government to track them along their educational careers. This data set of all Dutch students...

\(^4\)The number of dropouts is more easily observed at post-secondary level. Therefore, a significant part of older literature is focussing on the dropout of university students as data are readily available by university professors.
students, called the Bron data [Basis Register Onderwijsnummer], is used to calculate how many students are dropping out of secondary education.

Besides pupil specific information (e.g., ethnicity, family structure, school track), the sample contains information on the neighborhood (by means of the zip code). This makes it possible to exploit this exceptionally rich data set and compare students in Rotterdam with those in Amsterdam for the year 2007. Because the data consider all students in secondary education in Rotterdam (i.e., 48,900 students) and Amsterdam (i.e., 49,671 students), we do not encounter the problem of having selective student samples in our analysis.

A central issue in the Dutch policy debate is due to the direct difference in the percentage of students who dropout of secondary education between Amsterdam and Rotterdam. Student dropout in Amsterdam appears to be 0.76 percent lower than student dropout in Rotterdam (see Table 1). Central policy makers consider this as a signal of differences in policy effectiveness. There are at least two issues worth exploring: the differences in populations and the similarity in policy.

**Difference in population**

Below, we shortly characterize some similarities and differences between Amsterdam and Rotterdam, which may explain the difference in student dropout rate. This characterization happens on the basis of a report commissioned by the Scientific Council for Government Policy (SCGP, 2005).

---

5 The body of the paper includes Amsterdam and Rotterdam as main focus. These cities are the largest two cities of the Netherlands, which arguably received most attention from policy makers. In Appendix, we extend the data to two other large Dutch cities: Utrecht and The Hague.
An important difference between Amsterdam and Rotterdam is the education level of the inhabitants. Amsterdam attracts both high and low educated people and these people want to work and live in the city. People with a middle or high education level work in Rotterdam but do not live there and as a consequence the low educated persons stay behind in the city (see SCGP, 2005).

Given the empirical evidence that parental schooling is causally related to the schooling of the the child (see Holmlund, Lindahl and Plug, 2008), we would expect that students in Rotterdam follow more often a lower educational track. Table 1 confirms this partially. Students in Rotterdam participate more often in pre-vocational education and less frequently in pre-university education.
However, when we focus on the differences between the other education levels/tracks, these differences are not so explicit. We note that the descriptives reported in Table 1 are representative because they are based on registered data on all students in Amsterdam and Rotterdam. In Table 1 ‘brug class’ denotes the first year of secondary education. Students in this track are therefore the youngest students in the sample.

A second difference between Rotterdam and Amsterdam is the characterization of the labor market. Rotterdam can be characterized as a city of industrial labor. Even though many people lost their jobs in this particular sector, these jobs were not replaced by other jobs in other sectors. In Amsterdam the dominant sectors in the labor market are the financial, knowledge and service sector, together with the cultural and tourist sector. Even though many people who worked in the industrial sector lost their jobs, these jobs were frequently replaced by (better paid) jobs in one of the dominant sectors (see SCGP, 2005). This, together with the observation that parents are, on average, higher educated in Amsterdam, suggests that it is more likely that children in Amsterdam are, on average, higher educated, which on it turns has a positive influence on student dropout.

With respect to gender, Rotterdam is very similar to Amsterdam. Hence, gender may affect student dropout, but it is unlikely that it explains the difference in student dropout between the two cities. Rotterdam and Amsterdam are different with respect to the ethnical background of the students. Amsterdam has less native Dutch students and a relatively high proportion is Surinamese or Moroccan. Rotterdam, on the other hand, has a relatively high proportion of students from Aruba and the Netherlands Antilles, and a relatively high proportion of the students are Turkish or Non-Western Immigrants. Because these differences may be related with student dropout, but not with the performance of the region, we should control for these differences in the analysis. It is expected that around 2030 the first and second generation of immigrants will be the population majority in both cities. This is caused by the inflow of ethnic minorities into the city, as well as by the outflow of non-immigrant families out of the city. With respect to the gender distribution, it holds that there are approximately as many women as men in the city, and, as we would expect, this gender distribution also applies to the student population of both cities.

### Similarity in policy

Following the Lisbon Council, the Dutch Ministry of Education developed ‘a dropout policy’. One of the main features of the policy consisted out of a general agreement between regional authorities (e.g., region of Amsterdam and Rotterdam) and the Ministry. It is a written agreement with various policy options and large discretion on their implementation for the regions. From the suggested list with potential policy measures, the regions could (given their needs) select one or more options. For example, regions could take measures for improved reporting of truancy; regions
could extend the number of apprenticeship places; regions could obtain money for better study curriculum management; regions could smooth the transition from pre-vocational to vocational education; etc. It is beyond the scope of this paper to discuss into detail each of the policy measures (see De Bruijn et al. (2009, 2010) for a qualitative discussion; and De Witte and Cabus (2010) for a quantitative evaluation). At this point, it is important to note that the city of Amsterdam and Rotterdam undertook almost the same policy measures. The main difference between the two cities arises from the practical implementation, which is obviously driven by the underlying population. The city of Amsterdam opted for a hands-on implementation, indicating that they closely monitor the pupils at risk. Civil servants are closely following the students along the lines of the policy agreements with the Ministry. On the contrary, the city of Rotterdam implemented a hands-off approach. This leaves the main responsibility with the student. The city implements the policy agreement with the Ministry, but monitors less closely the students. Given the discussed differences in outcomes, it is worthwhile to examine whether the outcome difference is due to the difference in policy implementation or due to the underlying population.

3 Probit analysis

Often a multivariate Probit model is used to estimate the probability that a student drops out of secondary education while controlling for a wide range of observable and exogenous characteristics (including region) that are assumed to influence the student dropout rate (see, e.g., Adams and Becker, 1990; Allensworth, 2005 and reference therein). In this study we perform such a multivariate analysis and regress a dummy variable indicating the dropout status of students on gender, age, ethnicity, the educational track, the percentage of non-native students at the school and a set of dummies that, subsequently, indicate if a student is coming from a disadvantageous area (as defined by the Netherlands Statistics on a wide range of indicators) or needs additional learning support (i.e., student with low intellectual capacities). Finally, we add a dummy variable indicating whether the student goes to school in Rotterdam or Amsterdam.

The estimation results are presented in Table 2 and the signs of the coefficients are roughly consistent with the existing literature (see Hebert and Reis, 1999 and references therein). Girls are less likely to drop out and so are native students. Turkish and Moroccan students, however, are just as likely to drop out than native students. Students living in disadvantageous areas seem to have a higher probability of dropping out, but this effect is not significant. Social segregation, measured by the percentage of migrants at a school, has an unfavorable impact on the dropout decision of the students.

\footnote{Robustness tests pointed out that replacing the more detailed ethnicity variable by a migrant dummy variable (i.e., native or migrant) does not significantly change the results.}
Table 2: Probit Analysis

<table>
<thead>
<tr>
<th></th>
<th>Coeff.</th>
<th>Std.Err.</th>
<th>Z-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.92</td>
<td>0.15</td>
<td>-5.99</td>
</tr>
<tr>
<td>Gender (female = 1)</td>
<td>-0.15</td>
<td>0.02</td>
<td>-10.24</td>
</tr>
<tr>
<td>Age</td>
<td>0.03</td>
<td>0.00</td>
<td>6.20</td>
</tr>
<tr>
<td>Ethnicity (native = ref)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surinamese</td>
<td>0.06</td>
<td>0.02</td>
<td>2.38</td>
</tr>
<tr>
<td>Aruba / Netherlands Antilles</td>
<td>0.20</td>
<td>0.03</td>
<td>5.97</td>
</tr>
<tr>
<td>Turkey</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.20</td>
</tr>
<tr>
<td>Morocco</td>
<td>-0.03</td>
<td>0.02</td>
<td>-1.28</td>
</tr>
<tr>
<td>Non-Western migrant</td>
<td>0.09</td>
<td>0.02</td>
<td>3.55</td>
</tr>
<tr>
<td>Western migrant</td>
<td>0.15</td>
<td>0.03</td>
<td>5.03</td>
</tr>
<tr>
<td>Disadvantageous area (=1)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.88</td>
</tr>
<tr>
<td>Segregation (% immigrants at school)</td>
<td>-0.37</td>
<td>0.05</td>
<td>-7.77</td>
</tr>
<tr>
<td>Educational track:†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'Brug' class</td>
<td>-0.90</td>
<td>0.13</td>
<td>-6.70</td>
</tr>
<tr>
<td>Pre-vocational training (vmbo)</td>
<td>-1.05</td>
<td>0.13</td>
<td>-7.81</td>
</tr>
<tr>
<td>Pre-vocational education with additional support</td>
<td>-1.37</td>
<td>0.14</td>
<td>-9.90</td>
</tr>
<tr>
<td>Vocational - Economical topics</td>
<td>-0.42</td>
<td>0.13</td>
<td>-3.16</td>
</tr>
<tr>
<td>Vocational - Technical topics</td>
<td>-0.29</td>
<td>0.13</td>
<td>-2.22</td>
</tr>
<tr>
<td>Vocational - Social care topics</td>
<td>-0.55</td>
<td>0.13</td>
<td>-4.16</td>
</tr>
<tr>
<td>Vocational - Agricultural topics</td>
<td>-0.40</td>
<td>0.16</td>
<td>-2.58</td>
</tr>
<tr>
<td>General training (havo)</td>
<td>-0.95</td>
<td>0.13</td>
<td>-7.14</td>
</tr>
<tr>
<td>Pre-university training (vwo)</td>
<td>-1.30</td>
<td>0.14</td>
<td>-9.63</td>
</tr>
<tr>
<td>Rotterdam (=1)</td>
<td>0.03</td>
<td>0.01</td>
<td>2.11</td>
</tr>
</tbody>
</table>

$R^2$ 0.099
LR $\chi^2$ 4169.52

†Reference group is vocational with combined subjects.

students in the school. Compared to the reference category of students who are following combined subjects in vocational training, all other categories do significantly drop out less. Obviously, the stronger the educational track (compare for example vocational training to pre-university training), the stronger the favorable impact on dropout. Finally, the estimation results show that students in Rotterdam are more likely to drop out, compared to their student colleagues in Amsterdam.\(^7\)

Based on the estimated coefficients, we can predict the dropout probability for each student (i.e., fitted value of observation $i$) and compute the difference in probability between students in Rotterdam and Amsterdam. For Amsterdam, we find an average probability of 6.11%, while in

\(^7\)We extensively analyzed alternative probit specifications as robustness tests (e.g., including birth year of the student, detailed information on ethnicity or excluding variables). However, the difference between Amsterdam and Rotterdam proved not to be an artefact of the model specification but consistently remained significant under all specifications.
Rotterdam we find an average probability of 6.82%. Hence, a benevolent central government could conclude that the city of Amsterdam reduces the dropout rate more effectively than Rotterdam, even when controlling for compositional differences between the two cities.

However, based on the probit analysis we examine the difference in probability of dropping out in Rotterdam and Amsterdam, but the more relevant question is whether dropout rates among students in Rotterdam would have been different if these students would have lived in Amsterdam. This is fundamentally different than examining how student dropout rates differ between Rotterdam and Amsterdam. In the latter case we evaluate how living in Rotterdam and not Amsterdam influences dropout rates, while in the former case, we describe the dropout rate differs between the two populations. Based on the Probit analysis, the central government can therefore not conclude that Amsterdam reduces dropout rates more effectively than Rotterdam, because students in Rotterdam should be compared with an comparable group of students in Amsterdam (or vice versa) to make such a statement.

4 Matching Analysis

4.1 Matching Theory

According to the potential outcome model there are two potential outcomes for each student. The first outcome, \( y_{1i} \) represents the dropout status when students live in Rotterdam and \( y_{0i} \) represents the dropout status when students live in Amsterdam (see (Sless)-Neyman, J., 1923, 1990; Roy, 1951; Rubin, 1974; Rubin, 1976 and Holland, 1986). Obviously, we never observe both outcomes at the same time for any student and the outcome that we do not observe is generally referred to as the counterfactual outcome.

We could assume that the student population in Amsterdam represents the counterfactual outcomes, \( y_{0i} \), and determine the effect of living in Rotterdam and not in Amsterdam by the average treatment effect, \( E(y_{1i} - y_{0i}) \). However, differences in dropout probability between the two populations cannot be attributed to ‘living in Rotterdam’ if the student population of Amsterdam differs from the student population of Rotterdam in characteristics that are related to dropout rates. Even if we control for compositional differences by conditioning on a vector of observables, \( x_i \), i.e. we determine \( E(y_{1i} - y_{0i}|x_i) \), the student population of Amsterdam may include students who are non-comparable to any student in the student population of Rotterdam. In this case, the student population of Amsterdam does not accurately represent the counterfactual outcome \( y_{0i} \). As a consequence, a probit analysis where the probability of dropping out is regressed on a vector of observables \( x_i \) and a variable that indicates whether a student lives in Amsterdam or Rotterdam is not sufficient to determine the effect of ‘living in Rotterdam’. This reasoning inspired researchers
to try to estimate the average treatment effect on the treated rather than the average treatment
effect (Angrist and Krueger, 1999). We note, however, that estimating a simple probability model
is sufficient if the student population in Rotterdam resembles the population in Amsterdam in
those characteristics that determine the variation in dropout rate (which is, as argued before, not
the case).

Let $I$ represent a discrete treatment variable that takes the value one if a student lives in
Rotterdam and zero if students live in Amsterdam. Given the outcomes $y_{1i}$ and $y_{0i}$ for students
in Rotterdam and Amsterdam, respectively, the average treatment effect can be written as (see
Cameron and Trivedi, 2005):

$$E(y_{1i}|I = 1) - E(y_{0i}|I = 0) = E(y_{1i} - y_{0i}|I = 1) + \{E(y_{0i}|I = 1) - E(y_{0i}|I = 0)\}. \tag{1}$$

The first term on the second line is the average treatment effect on the treated and the second term
in braces represents a ‘bias’. Since, we are interested in the average treatment effect on the treated
this requires that $E(y_{0i}|I = 1) = E(y_{0i}|I = 0)$. However, this condition may not be met due to
composition differences, selection on observables and selection on unobservables.

To control for differences in student composition and selection on observables, we should condi-
tion on those characteristics, $x_i$, that significantly explain the variation in dropout rates and that
are known to affect the status of living in Rotterdam. We then have that $y_{0i} \perp I|x_i$ such that
$E(y_{0i}|x_i, I = 1) = E(y_{0i}|x_i, I = 0)$. To $y_{0i} \perp I|x_i$ is generally referred to as unconfoundedness
(Imbens, 2005), or ignoribility (Rubin, 1978; Wooldridge, 2001). Under the assumption that the
ignorability assumption is satisfied, Angrist and Krueger (1999) show that the average treatment
effect conditional on $x_i$ is given by:

$$E(y_{1i} - y_{0i}|I = 1) = E(\Delta x_i|I = 1) = E(y_{1i}|x_i, I = 1) - E(y_{0i}|x_i, I = 0). \tag{2}$$

The ignoribility assumption, however, does not ensure that we control for unobserved factors that
partly determine $I$ and $y$; the so-called selection on unobservables. In the registered data we use
there is, for example, no information on parental schooling, while there is evidence that parental
schooling is causally related to children’s schooling (see Holmlund, Lindahl and Plug, 2008). Dif-
fences in parent’s schooling levels between the two populations may be related to the students
dropout status. Although, we partly control for the parent’s education level by including education
type and ethnicity as conditioning variables in $x_i$, we can not be certain that these variables capture
the entire schooling effect of the parents on dropout status.

For this study it is important to keep in mind that the a priori expectation of policy makers,
based on Table 2, is that students in Amsterdam perform better, i.e. \( E(y_{1i} - y_{0i} | I = 1) < 0 \). Under the assumption that the expectation of the policy maker is correct and on the basis of the empirical results, we can determine whether the selection on unobservables is positive or negative and can reason if this selection on unobservables is plausible.

Suppose we find that the student dropout rate in Rotterdam is similar to that of Amsterdam. If the expectation of the policy maker is correct (\( E(y_{1i} - y_{0i} | I = 1) < 0 \)), then it must be that \( E(y_{0i} | I = 1) - E(y_{0i} | I = 0) > 0 \). This means that the expectation of the policy maker can only be true if students who are less likely to drop out are more likely to live in Rotterdam. But the latter is not very plausible, because higher educated parents are more likely to work in the commercial/service sector and less likely to work in the industrial labor market, and so higher educated parents are more likely to self-select in Amsterdam (SCGP, 2005).

If our conditional variables do not correct (enough) for the parental education effect, then students in Rotterdam are performing better, if higher educated parents with children who are less likely to drop out of school select themselves in Amsterdam. On the other hand, if our conditional variables properly correct for selection on unobservables then our conclusion should be that students in Amsterdam and Rotterdam are evenly likely to drop out of school.

The problem with selection on unobservables is that we cannot check whether this selection occurs and, consequently, we cannot determine its impact. Nevertheless, and as we try to point out in the example above, it is possible to reason how selection on unobservables can influence the empirical result under the assumption that the expectation of the policy maker is correct.

### 4.2 Matching Procedure

The description of the matching procedure relies on Cameron and Trivedi (2005). Denote the comparison group for student \( i \) in Rotterdam with characteristics \( x_i \) as the set \( A_j(x) = \{ j | x_j \in c(x_i) \} \), where \( c(x_i) \) is the characteristics neighborhood of \( x_i \). Furthermore, \( N_A \) and \( N_R \) denote the number of students in, respectively, Amsterdam and Rotterdam and the weight given to the \( j^{th} \) case, that could serve as a potential match for the \( i^{th} \) treated case, is denoted as \( w(i,j) \) with \( \sum_j w(i,j) = 1 \). The matching estimator of the average treatment effect on the treated is:

\[
\Delta = \frac{1}{N_R} \sum_{i \in \{I=1\}} [y_{1i} - \sum_j w(i,j) \cdot y_{0j}],
\]

where \( 0 < w(i,j) \leq 1 \), and \( \{I=1\} \) is the set of students who are living in Rotterdam and \( j \) is an element of the set of matched students in Amsterdam. An intuitive interpretation of the matching algorithm is that it links each student in Rotterdam to the best look-alike student in Amsterdam.
given a set of characteristics that presumably influence the drop out probability.

From equation (3) it follows that different matching estimators are generated by choosing different weights. We can choose between an exact matching estimator, a kernel estimator or an estimator that is based on some distance measure. We do not choose for an exact matching estimator because the probability of finding an exact match depends on the number of matching variables. In our case this induces a bias, because it is less likely that a match will occur for households with characteristics that are less likely and, consequently, the estimate will show a regression towards the mean.

The results presented in this study are based on nearest neighbor matching using the mahalanobis distances and this means that we match each student in Rotterdam to the best look-alike student in Amsterdam on the basis of a vector of observables, \( x \). The advantage of using the mahalanobis distance is that it is intuitive and fully non-parametric so that the outcome of the match does not rely on any functional form or distribution. Mahalanobis matching minimizes the distance between students according to the following rule:

\[
w(i, j) = 1 \text{ if } j = \arg \min_{j=1,\ldots,N^A} (x_i - x_j)' \Sigma^{-1} (x_i - x_j), \tag{4}
\]

where \( \Sigma^{-1} \) represents the within sample covariance matrix and where \( w(i, j) = 1 \) if a match is possible. This rule simply states that each student from Rotterdam is matched to a student from Amsterdam who is closest to him/her.\(^8\)

There is one important issue that we should consider when performing the analysis. The student population of Rotterdam counts 48,900 students and the student population of Amsterdam counts 49,671 students. On the one hand, we have that the quality of the match becomes worse if we do not allow that students in Amsterdam are matched to students in Rotterdam more than once. This is because, as the matching procedure continues, there will be less students from Amsterdam to choose from and, evidently, students with characteristics that are most likely will be matched first. On the other hand, if we allow that students in Amsterdam are matched to students in

---

\(^8\)Kernel estimators or matching estimators based on a propensity score, are not necessarily inferior to Mahalanobis matching. Each matching method has its own advantages and disadvantages and for an elaborate description of the available matching methods we refer to Cameron and Trivedi (2005). As a robustness check we matched students based on a conditional probability of living in Rotterdam and based on a kernel function and we found that the results and conclusions were similar to results with Mahalanobis matching. We matched students from Rotterdam to one, five and ten students from Amsterdam using the propensity score, and we matched students on the basis of caliper and kernel matching. When we match on the propensity score we match on the conditional probability, \( p(x) \), that a student lives in Rotterdam given \( x \). The matching set is then \( A_i(p(x)) = \{ p_j \min_j \| p_i - p_j \| \} \). Caliper matching is essentially a propensity score matching estimator where we impose that \( p_i - p_j < \varepsilon \). For \( \varepsilon \) we take the values 0.05 and 0.01. When we performed Kernel matching we used an Epanechnikov kernel function with 0.6 as bandwidth and the weight that defines the Kernel matching estimator is then \( w_{i,j} = \frac{K(x_j - x_i)}{\sum_{j=1}^N K(x_j - x_i)} \). The outcomes of the alternative matching models are available upon request.
Rotterdam more than once then it may be that the estimate we will find is driven by a small group of Amsterdam students. Additionally, and worse, the way students are ordered in the data determine which student in Amsterdam is matched to a student in Rotterdam. For example, it is likely that many students in Rotterdam with common characteristics can be matched to many students in Amsterdam, but the ordering of the data ensures that the same student from Amsterdam is picked as a match and, consequently, this one student is overrepresented in the analysis.

To make our results less dependent on the ordering of students we simulate the distribution of the matching estimator and, first of all, control for the non-random ordering of students and, second, control for the fact that students with characteristics that are relatively unlikely receive a lower weight in the analysis. We perform 500 simulations and in each simulation we select one thousand students from Rotterdam at random. These students are then randomly ordered by generating and ordering a variable that assigns a uniform pseudorandom number on the interval \([0,1)\) to each of the 1000 students from Rotterdam and to each of the 49,017 students from Amsterdam. The treatment effect of the treated is obtained using equations (3) and (4).

By performing 500 simulations we essentially simulate the distribution of the treatment effect on the treated. The mean of this distribution is the estimated treatment effect on the treated and the standard deviation shows how reliable this estimate is. We note that the distribution of the matching estimator is not necessarily normal. If we evaluate whether students from Rotterdam differ significantly from students in Amsterdam we should consider this distribution.

5 Matching students in Amsterdam to students in Rotterdam

Figure 1 shows the distribution of the average treatment effect on the treated (ATET) based on the 500 simulations we perform. The figure represents the difference in student dropout rate between students in Rotterdam and look-alike students in Amsterdam. A negative (positive) value means that we find that students from Rotterdam have a lower (higher) dropout probability. Intuitively, these differences indicate whether the dropout rate of students in Rotterdam would change, if they would have lived in Amsterdam.\(^9\)

\(^9\)Remark that in the matching analysis, we experimented with various robustness tests. For example, we included control variables such as age, detailed information on ethnicity and family status (divorced or married parents). The results did not significantly change.
The standard deviation of this distribution equals 0.028 and the black circle shows the distribution mean of 0.013. Although the mean suggests that, on average, student dropout in Amsterdam is lower, we should test whether the distributional mean is significantly different from zero. In order to do so, we first perform a test of normality by eye-ball empirics. This is graphically illustrated in Figure 2, where we plot the simulated values of the ATET against the normal distribution. The red diagonal line represents the normal distribution, while the blue dots represent the simulated values. The distribution is more normal as the blue dots tend to cleave more to the red line. The simulated distribution seems approximately normally distributed, except that there is one outlier at the lower end of the distribution and that the blue dots seem to cleave less to the red line at the upper end of the distribution. Assuming normality, we would find that the mean of the distribution does not differ significantly from zero, with a t-value of 0.46.

A more formal test than the graphical presentation consists of a Kolmogorov-Smirnov Normality test. This test rejects normality of the simulated distribution (prob$>\chi^2 = 0.0002$). Therefore we also determine the non-parametric Wilcoxon signed-rank statistic, but again find that the distribution mean does not differ significantly from zero ($z=2.61$).

Figure 1: Simulated distribution of the treatment effect on the treated
We note that several matching analysis were performed, where first we conditioned only on educational track and then extended the number of conditioning variables. For example, we started with conditioning only on the educational track such that students in Rotterdam are comparable to students in Amsterdam in the educational track they follow, but not necessarily in other characteristics that may also affect dropout in secondary education. We found that the difference in dropout rate between Amsterdam and Rotterdam is already insignificant when we condition only on educational track, and this insignificance remains when we include the other covariates, mentioned in Table 1, in the set of matching variables.

The general conclusion is that dropout rates are not significantly different when we compare students in Rotterdam to look-alike students in Amsterdam. This conclusion clearly differs from the probit conclusion that suggested that dropout rates were lower in Amsterdam. The difference in result arises because the probit analysis compares the students in Rotterdam with all students in Amsterdam, instead of comparing them with only look-alike students of Amsterdam. The matching analysis therefore uses a better control group and is more suitable to draw conclusions upon. Hence, central government policy makers should (at least) take into account whether the control group that is used is appropriate before drawing any conclusions.
6 Conclusion

Central governments are interested in whether their money is well spend. From an evaluation perspective, it is rational to consider the differences in outcome between regions. It becomes tricky, however, when central governments consider these regional differences as performance measures and conclude that regions with inferior outcomes did not implement best practice policy. If uniform (monetary) subsidies are allocated to the outcome measure, the implication of differences in outcome are even stronger.

From an empirical perspective, this paper focussed on dropout prevention in the Netherlands. The Lisbon European Council (2000) stipulated that the percentage of dropout students should halve between 2002 and 2010. As in many other policy frameworks, the dropout prevention uses a subsidiarity principle: policy implementations are made at the decentralized level (usually the school level) and subsidies are given at the regional level. In the Netherlands, one of the incentive schemes consists of a uniform incentive of 2500 euro per students that drops out less than base year 2005-2006. A similar uniform incentive might be appropriate if the underlying population is completely similar. However, this paper indicated that this is not the case. In particular, we address the question: What if students living in one city (Rotterdam) would live in another city (Amsterdam), how does this affect dropout rates?

From a methodological perspective, we clarify and show the difference between applying a traditional probit model and applying an iterative matching model. The probit analysis measures the difference in student dropout between the two cities and this is fundamentally different from the research question addressed above. In the former case we evaluate how living in Rotterdam and not Amsterdam influences dropout rates, while in the latter case, we describe the difference in dropout rate between the two populations. Based on the Probit analysis, the central government can therefore not conclude that Amsterdam reduces dropout rates more effectively than Rotterdam, because students in Rotterdam should (at least) be compared with a comparable group of students in Amsterdam (or vice versa). If policy makers hold cities or regions accountable for their performance on particular dimensions, and if uniform incentives are based on these outcome measures, central government should account for regional heterogeneity in an appropriate manner. Otherwise, the incentives that are given may be demotivating for the region with the most disadvantaged population.

Appendix

The main text presents the estimation results for Amsterdam and Rotterdam, the two most discussed Dutch cities (at least at the Ministry of Education). It is, however, interesting to also
perform the analysis for alternative cities. In this Appendix, we repeat the analysis for all students of the following city couples: The Hague-Amsterdam, The Hague-Rotterdam, Utrecht-Amsterdam, Utrecht-Rotterdam and Utrecht-The Hague.

Table 3 presents the descriptive statistics for The Hague and Utrecht. The descriptives statistics for Rotterdam and Amsterdam can be found in Table 1. The tables clearly indicate that the underlying population characteristics differ substantially among all four cities. Controlling for the observed differences is therefore crucial.

| Table 3: Descriptive statistics The Hague and Utrecht |
|---------------------------------|-----------|
| The Hague Utrecht               |
| dropout (% of population in city)| 5.87      |
| Gender (male % of population in city)| 50.56     |
| Disadvantageous area (% of population in city)| 44.98     |
| Segregation (% migrants at school)| 0.49      |
| Ethnicity:                      |
| Native Dutch                   | 51.06     |
| Surinamese                     | 4.21      |
| Aruba / Netherlands Antilles   | 1.2       |
| Turkey                         | 10.22     |
| Morocco                        | 20.82     |
| Non-Western migrant            | 6.09      |
| Western migrant                | 6.29      |
| Educational track              |
| 'Brug' class                   | 12.92     |
| Pre-vocational education (vmbo)| 11.41     |
| Pre-vocational education with additional support | 18.65 22.43 |
| Vocational - Economical topics | 14.49     |
| Vocational - Technical topics  | 8.79      |
| Vocational - Social care topics| 8.79      |
| Vocational - Agricultural topics| 0.82     |
| Vocational - Combined topics   | 0.24      |
| General training (havo)        | 8.98      |
| Pre-university education (vwo) | 17.49     |
| Total number of students       | 16,431    |
|                               | 32,687    |

Multivariate Probit models, controlling for a wide range of observable and exogenous characteristics (including a region dummy), are frequently estimated to predict the probability that students
drop out of secondary education. The estimated coefficient of the region dummy variable cannot be considered as a performance measure, i.e. as information that shows that one city outperforms another city. Table 4 presents the estimation coefficients for the region dummy in the multivariate probit framework. A positive coefficient denotes a higher student drop out in the column city than the row city (e.g., controlled for various observables, Rotterdam has a 3.1% higher dropout level than Amsterdam).

Table 4: City ‘effects’ from a probit analysis†

<table>
<thead>
<tr>
<th></th>
<th>Amsterdam</th>
<th>Rotterdam</th>
<th>The Hague</th>
<th>Utrecht</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amsterdam</td>
<td>0</td>
<td>0.031**</td>
<td>-0.011</td>
<td>0.042*</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.016)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>Rotterdam</td>
<td>0</td>
<td>-0.045***</td>
<td>0.039*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Hague</td>
<td>0</td>
<td>0.038*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.021)</td>
<td></td>
</tr>
</tbody>
</table>

† */ **/ *** means statistically significant at the 10/5/1 percent level. The control variables used in the analysis are similar to those mentioned in Table 2

Based on Table 4, policy makers could conclude that Amsterdam outperforms Rotterdam and Utrecht because drop out in Amsterdam is significantly lower. In a similar way, they could conclude that Utrecht and Rotterdam perform equally well, and that both cities perform worse than The Hague.

In Section 4.1 and 4.2 we argued that the multivariate probit examines the difference in dropout probability between to cities, and that this is fundamentally different than examining how dropout rates among students living in one city would have been different if these students would have lived in another city.

Table 5 shows the iterative matching results. The iterative matching procedure chooses a more appropriate control group than the multivariate probit. Table 5 reveals that the significant differences in dropout in Table 4 are now insignificant, in fact all dropout differences turn out to be insignificant. On the basis of the matching results, policy makers can no longer conclude that one city outperforms the other city in terms of dropout statistics. The extension of the analysis once again shows that it is dangerous to interpreted regional differences in a multivariate probit models as performance measures.
Table 5: City ‘effects’ from a probit analysis†

<table>
<thead>
<tr>
<th></th>
<th>Amsterdam</th>
<th>Rotterdam</th>
<th>The Hague</th>
<th>Utrecht</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amsterdam</td>
<td>0</td>
<td>0.013</td>
<td>-0.019</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.050)</td>
<td>(0.045)</td>
<td></td>
</tr>
<tr>
<td>Rotterdam</td>
<td>0</td>
<td>-0.005</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0512)</td>
<td>(0.046)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Hague</td>
<td>0</td>
<td>0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.046)</td>
</tr>
</tbody>
</table>

† *, **/ ** means statistically significant at the 10/5/1 percent level. The control variables used in the analysis are similar to those mentioned in Table 2

References


Cameron, A.C. and P.K. Trivedi (2005), Microeconometrics: methods and applications, Cambridge University Press, New Yor.


