How uniform incentives can provide a negative motivation: an application to early school leaving in two large cities

de Witte, K.; van Klaveren, C.

Publication date
2010

Document Version
Submitted manuscript

How Uniform Incentives can Provide a Negative Motivation – An Application to Early School Leaving in Two Large 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 may consider the differences in outcomes as an indication on the effectiveness of policy. In turn, to motivate regional policy makers uniform incentives might be developed. However, if the underlying population differs, uniform incentives may have an discouraging effect. This paper considers the differences in outcomes in early school leaving between the two largest Dutch cities. Using an exceptionally rich data set of all students in Amsterdam and Rotterdam, we test by probit and matching models whether uniform incentives are appropriate.

JEL Codes: C14, C61, C23, I21

Keywords: Uniform incentives; Early school leaving; Comparative; Matching analysis

1 Introduction

Since the Lisbon Agenda (2000), early school leaving is highly ranked on the European political agenda. The Lisbon Agenda stipulated that the percentage of drop out students should halve between 2002 and 2010. The European Commission (2006) defines a 'drop out' (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 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 drop out 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 drop out 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. However, it becomes tricky, when they consider these regional differences as performance measures. It is totally nonsense if incentives are based on the regional differences.

This paper shows that one cannot consider straightaway 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 contributions of this paper are threefold. First of all, this paper indicates why uniform incentives can be demotivating. The principal (i.e., the central government) sets out an incentive for the agent (i.e., the regional government).[1] To keep the incentive as transparent as possible, the principal determines a uniform incentive for all agents. In particular, if an *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 regions, obtaining the same target may be more requiring in a region with less advantageous characteristics than in a region with more advantageous characteristics. As such, both 'naming and shaming' and monetary rewards will provide a negative motivation for the region with a disadvantageous population.

As a second contribution, this paper compares the outcomes of two popular assessment procedures. Often a probit or logit model is used to examine how regional differences in policy affects the outcome measure (in casu, drop out 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 often uses 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 other words, if fully allowing for population heterogeneity, is an outcome measure comparable

[1] 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.

across regions. In the paper, the research question 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*".

A third contribution of the paper arises from the application at hand. In the Netherlands, the Ministry of Education, Culture and Sciences (OCW) tries to meet the targets in the Lisbon Agenda 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. While the Ministry of Education determines the general policy framework, the regions have a large discretion in filling this policy (and, thus, in spending the remaining 88.6% of the budget). While the larger part of the subsidy is not uniform, 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 particularly 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. In turn, this is explained by a more motivated alderman in Amsterdam (in comparison to the alderman in Rotterdam). If this is true, a uniform incentive is the most appropriate way to incentives policy makers (Laffont and Tirole, 1993). On the other hand, schools and regional policy makers point to the heterogeneity in population. If population significantly differs, 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 for the two largest cities in the Netherlands, i.e. Rotterdam and Amsterdam. At first sight, both cities have similar characteristics such that the uniform incentive seems 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 drop out status of *all* Dutch students in secondary education living in Amsterdam and Rotterdam (the two largest Dutch cities). 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 drop out probability for students living in Rotterdam would have been different if they would have lived in Amsterdam.

The remainder of this paper is organized as follows. Section 2 explores the data at hand. In particular, we briefly outline why policy makers are arguing that the alderman of Amsterdam is more motivated than the alderman of Rotterdam, and that this reduces the drop out rate faster. 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 present and discuss the empirical results of the iterative matching analysis. Finally, in Section 5 we conclude.

2 Blaming the alderman?

The data

Although student drop out 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.[2] 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, 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 drop out of secondary education between Amsterdam and Rotterdam, the two largest cities in the Netherlands). Student drop out in Amsterdam appears to be 0.76 percent lower than student drop out in Rotterdam (see Table 1). As has been discussed before, central policy makers consider this as a signal of differences in policy effectiveness. Indeed, the Dutch Ministry of Education is subsidizing regions of municipalities (note that both Amsterdam and Rotterdam constitute a separate region) to reduce the drop out rate and the subsidies (equal to 313 million euro in 2008, and increasing to 400 million euro in 2011) should be used for particular policy measures. It is beyond the scope of this paper to discuss extensively the policy measures, however, by large, Amsterdam and Rotterdam implemented similar policy measures to reduce drop out (see De Bruijn et al., 2010). As the policy measures are similar, civil servants argue that the differences in drop out are arising from differences in motivation between the two alderman.

However, the descriptive statistics in Table 1 show that Amsterdam and Rotterdam differ in various other characteristics, besides the percentage of student drop out. Since these differences may partially explain why student drop out differs between Amsterdam and Rotterdam, the policy maker should (at least) take them into account if (s)he wants to be able to conclude anything on

[2] The number of drop outs is more easily observed at post-secondary level. Therefore, a significant part of older literature is focussing on the drop out of university students as data are readily available by university professors.

Table 1: Descriptive statistics

	Amsterdam	Rotterdam
Drop out (% of population in city)	5.89	6.65
Gender (male % of population in city)	50.30	50.37
Migrant (% of population in city)	69.56	64.01
Disadvantageous area (% of population in city)	77.97	75.97
Segregation (% migrants at school)	0.41	0.47
'Brug' class	12.92	11.74
Pre-vocational education (vmbo)	15.96	19.26
Pre-vocational education with additional support	11.66	9.63
Vocational - Economical topics	17.54	16.12
Vocational - Technical topics	7.60	10.66
Vocational - Social care topics	9.69	11.96
Vocational - Agricultural topics	0.39	0.69
Vocational - Combined topics	0.04	0.22
General training (havo)	8.74	9.20
Pre-university education (vwo)	15.46	10.52
Total number of students	49,671	48,900

how the implemented policy is linked to the student drop out rates observed in Amsterdam and Rotterdam.

Difference in population

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

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 often 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 often 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 drop out.

With respect to gender and ethnical background, Rotterdam is very similar to Amsterdam. Hence, gender and ethnicity may affect student drop out, but it is not likely that these characteristics explain the difference in student drop out between the two cities. With respect to ethnical background, 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.

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 drop out 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 drop out status of students on gender, 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. 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 drop out decision of the 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 drop out. Finally, the estimation results show that students in Rotterdam are more likely to drop out, compared to their student

Table 2: Probit Analysis

	Coeff.	Std.Err.	Z-value
Constant	-0.60	0.13	-4.72
Gender (female = 1)	-0.15	0.01	-10.80
Immigrant Students‡ (=1)	0.07	0.02	3.97
Disadvantageous area (=1)	-0.01	0.02	-0.52
Segregation (% immigrants at school)	-0.39	0.04	-9.39
Educational track:†			
'Brug' class	-1.43	0.13	-11.09
Pre-vocational training (vmbo)	-0.10	0.13	-7.99
Pre-vocational education with additional support	-0.96	0.13	-7.60
Vocational - Economical topics	-0.38	0.12	-3.04
Vocational - Technical topics	-0.26	0.12	-2.10
Vocational - Social care topics	-0.51	0.13	-4.11
Vocational - Agricultural topics	-0.38	0.15	-2.56
General training (havo)	-1.09	0.13	-8.63
Pre-university training (vwo)	-1.35	0.13	-10.57
Rotterdam (=1)	0.03	0.01	2.51

†Reference group is vocational with combined subjects. ‡Immigrants are defined according to the definition of the Netherlands Statistics, i.e. both parents are not born in the Netherlands.

colleagues in Amsterdam.[3]

Based on the estimated coefficients, we can predict the drop out 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 a probability of 6.11%, while in Rotterdam we find a probability of 6.82%. Hence, a benevolent central government could conclude that the city of Amsterdam reduces the drop out 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 drop out 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 drop out rates differ between Rotterdam and Amsterdam. In the latter case we *evaluate* how living in Rotterdam *and not* Amsterdam influences drop out rates, while in the former case, we *describe* the drop out rate differs between the two populations. Based on the Probit analysis, the central government can therefore not conclude that Amsterdam reduces drop out 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.

[3]We extensively analyzed alternative probit specifications as robustness tests (e.g., including birth year of the student 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.

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 drop out status when students live in Rotterdam and y_{0i} represents the drop out status when students live in Amsterdam (see (Splawa)-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 drop out 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 drop out rates. Even if we control for compositional differences by conditioning on a vector of observables, $\mathbf{x_i}$, i.e. we determine $E(y_{1i} - y_{0i}|\mathbf{x_i})$, 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 $\mathbf{x_i}$ 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 drop out rate (which is, as argued before, not the case).

Let I represent a discrete treatment variable that takes the value one if a students live 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):

$$\begin{aligned} &E(y_{1i}|I=1) - E(y_{0i}|I=0) \\ &\qquad = E(y_{1i} - y_{0i}|I=1) + \{E(y_{0i}|I=1) - E(y_{0i}|I=0)\}. \end{aligned} \tag{1}$$

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 condition on those characteristics, $\mathbf{x_i}$, that significantly explain the variation in drop out rates and that are known to affect the status of living in Rotterdam. We then have that $y_{0i} \perp I|\mathbf{x_i}$ such that $E(y_{0i}|\mathbf{x_i}, I = 1) = E(y_{0i}|\mathbf{x_i}, I = 0)$. To $y_{0i} \perp I|\mathbf{x_i}$ 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 treatement effect conditional on $\mathbf{x_i}$ is given by:

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

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). Differences in parent's schooling levels between the two populations may be related to the students drop out status. Although, we partly control for the parent's education level by including education type and ethnicity as conditioning variables in $\mathbf{x_i}$, we can not be certain that these variables capture the entire schooling effect of the parents on drop out 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 drop out 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 $\mathbf{x_i}$ as the set $A_j(\mathbf{x}) = \{j|\mathbf{x_j} \in c(\mathbf{x_i})\}$, where $c(\mathbf{x_i})$ is the characteristics neighborhood of $\mathbf{x_i}$. 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 treatement effect on the treated is:

$$\Delta = \frac{1}{N_R} \sum_{i \in \{I=1\}} [y_{1,i} - \sum_j w(i,j) \cdot y_{0,j}], \tag{3}$$

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.

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 neightbor 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, $\mathbf{x}$. 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:

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

where Σ^{-1} represents the within sample covariance matrix and where $w(i,j) = 1$ if a match is possible.

We emphasize that kernel estimators or matching estimators based on a propensity score, i.e. the conditional probability of being a student in Rotterdam, 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.[4]

[4]We matched students from Rotterdam to one, five and ten students from Amsterdam using the propensity score,

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 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 simulation 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

In the matching analysis we use Mahalanobis distances to match each student in Rotterdam to the best look-alike student in Amsterdam on the basis of a vector of observed characteristics ($\mathbf{x}$). To enhance comparability, the covariates used in the matching analysis are the same variables that are used in the Probit analysis.

Figure 1 shows the distribution of the average treatment effect on the treated (ATET) based on

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, $p(\mathbf{x})$, that a student lives in Rotterdam given $\mathbf{x}$. The matching set is then $A_i(p(\mathbf{x})) = \{p_j \min_j \|p_i - p_j\|\}$. Caliper matching is essentially a propensity score matching estimator where we impose that $p_i - p_j < \varepsilon$. For ε we take the values 0.05 and 0.01. When we performed Kernel matching we used an Epanechnikov kernel funtion with 0.6 as bandwidth and the weight that defines the Kernel matching estimator is then $w_{i,j} = \frac{K(\mathbf{x_j}-\mathbf{x_i})}{\sum_{j=1}^{N_{C,i}} K(\mathbf{x_j}-\mathbf{x_i})}$. The outcomes of the alternative matching models are available upon request.

the 500 simulations we perform. The figure represents the difference in student drop out 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) drop out probability. Intuitively, these differences indicate wether the drop out rate of students in Rotterdam would change, if they would have lived in Amsterdam.

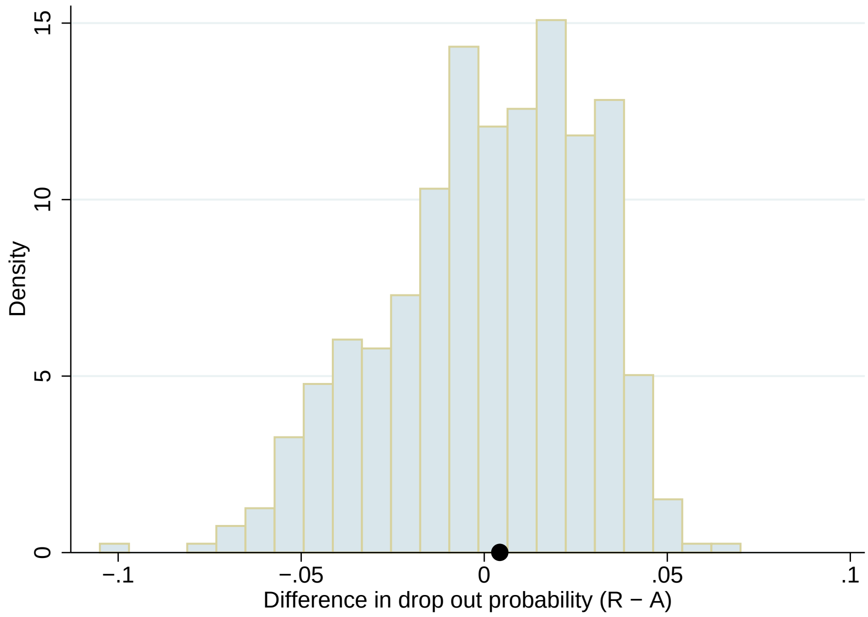

Figure 1: Simulated distribution of the treatment effect on the treated

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 drop out 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> χ^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).

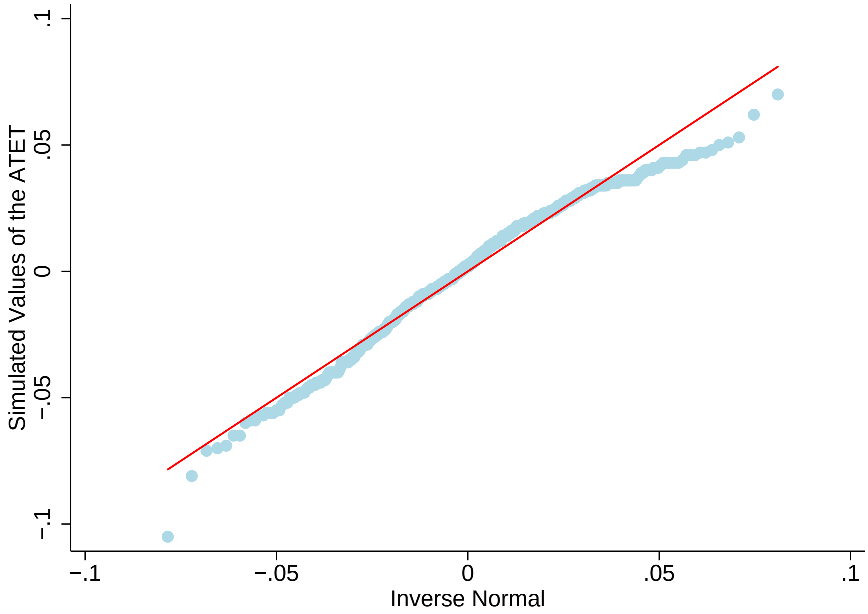

Figure 2: Graphical test of normality of the treatment effect on the treated

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 drop out in secondary education. We found that the difference in drop out 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 drop out 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 drop out rates where 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.

This paper focussed on dropout prevention in the Netherlands. The Lisbon Agenda (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 the year before. 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 drop out rates?*. In doing so, 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 drop out 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 drop out rates, while in the latter case, we *describe* the difference in drop out rate between the two populations. Based on the Probit analysis, the central government can therefore not conclude that Amsterdam reduces drop out rates more effectively than Rotterdam, because students in Rotterdam should (at least) be compared with an *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 disadvantageous population.

References

Adams, J.L. and W.E. Becker (1990), 'Course Withdrawals: A Probit Model and Policy Recommendations', *Research in Higher Education* **31**(6), 519–538.

Allensworth, E.M. (2005), 'Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion', *Educational Evaluation and Policy Analysis* **27**(4), 341.

Angrist, J.D. and A.B. Krueger (1999), *Empirical Strategies in Labor Economics*, Vol. 3 of *Handbook of Labor Economics*, Elsevier Science BV, chapter 23, pp. 1277–1364.

Blundell, R. and M. Costa Dias (2007), 'Alternative approaches to evaluation in empirical microeconomics.'. Institute for Fiscal Studies, Londen.

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

Commission, European (2006), Detailed analysis of progress towards the lisbon objectives in education and training: Analysis of benchmarks and indicators?, Technical report.

De Bruijn, H., M. Groenleer, H. van der Voort, M. Noordink, B. Dunning and B. Gooskens (2010), Inzicht in resultaat 2, Technical report, TU Delft.

Engelbersen, G., E. Snel and A. Weltevrede (2005), Sociale herovering in Amsterdam en Rotterdam – een verhaal over twee wijken. Wetenschappelijke Raad voor het Regeringsbeleid (Scientific Council for Government Policy).

Hebert, T.P. and S.M. Reis (1999), 'Culturally diverse high-achieving students in an urban high school', *Urban Education* **34**(4), 428.

Holland, P.W. (1986), 'Statistics and causal inference (with discussion)', *Journal of the American Statistical Association* **81**, 945–970.

Holmlund, H., M. Lindahl and E. Plug (2008), The causal effect of parent's schooling on children's schooling: A comparison of estimation methods, IZA Discussion Papers 3630, Institute for the Study of Labor (IZA).

Imbens, G. (2005), 'Semiparametric estimation of average treatment effects under exogeneity: A review', *Review of Economics and Statistics* .

Laffont, J.J. and J. Tirole (1993), *A theory of incentives in procurement and regulation*, the MIT Press.

Roy, A. (1951), 'Some thoughts on the distribution of earnings', *Oxford Economic Papers Psychology* **3**, 135–146.

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

Rubin, D.B. (1976), 'Inference and missing data', *Biometrika* **63**, 581–592.

Rubin, D.B. (1978), 'Bayesian inference for causal effects', *Annals of Statistics* **6**, 34–58.

(Splawa)-Neyman, J. (1923, 1990), 'On the principles of probability theory to agricultural experiments. essays on principles. section 9', *Statistical Sciences* **5**, 465–471.

Wooldridge, J.M. (2001), 'Asymptotic properties of weighted m-estimators for standars stratified samples', *Econometric Theory* **17**, 451–470.