ERASMUS UNIVERSITY

MASTER'S THESIS

DUTCH AREA-BASED SUBSIDY FOR DISADVANTAGED PRIMARY SCHOOL PUPILS REVISITED

Author:
Joris J. VAN EIJK

Student number:
380493

Supervisor:
dr. Sacha KAPOOR

Second assessor:
Matthijs OOSTERVEEN MSc

ABSTRACT

In this paper, I evaluate a Dutch area-based subsidy scheme that targets disadvantaged pupils in primary education. In line with the purpose of this subsidy scheme, I hypothesise that the extra funding has a positive impact on pupil performance. The strict eligibility rules of the subsidy scheme allow for a regression discontinuity design, which analyses a select group of observations. I use school-level government data from three school years covering the period 2013 to 2016. Using ordinary least squares models I find no evidence to support the hypothesis.

This thesis is submitted in fulfillment of the requirements for the degree of Master of Science in Economics and Business at the Erasmus School of Economics.

30 AUGUST 2017

Acknowledgements

I thank my supervisor, Sacha Kapoor, for accepting my request to supervise this project and guiding me along the way. I admire his forward-looking focus and his straightforwardness. I thank Matthijs Oosterveen for performing a second assessment of this thesis. I warmly thank my fellow student and dear friend Bart de Koning for his inexhaustible stream of helpful feedback and for our many cups of coffee and plenty fruitful conversations, chat sessions and phone calls, resulting in valuable insights and motivation. Special thanks go to Marijn Verspaandonk for providing data that he used in his 2016 master's thesis; to Daniel Bischof for making his Stata graph colour scheme plotplainblind *freely available; to the people that contributed to the open source LaTeX template that provided a basis on which I could build this final layout; and to those who pre-read this thesis. Finally, I thank my family for their unconditional support.*

Joris van Eijk

30 August 2017

"I want to spend this money effectively: the money has to end up there where it is needed the most. At the same time, I want to have a clear view on the benefits."

S.A.M. Dijksma

State Secretary of Education, Culture and Science

The Hague, 18 January 2008

1 Introduction

The Dutch government wants to offer all children the opportunity to utilise their talents (2008). This translates to creating equal chances for all children, so that they can perform to the best of their abilities in education. Most children do not need additional help to do so. For others, their upbringing or environment provide a poor start. For these children, the Dutch government is willing to provide additional resources to compensate for the unfavourable environmental factors.

A policy that put this into practice is called the impulse subsidy scheme. It was introduced in 2009 and provides additional funding for disadvantaged pupils in primary education. In the dataset of this paper, the amounts of impulse subsidy money per eligible pupil per year are 1,690 euro, 1,712 euro and 1,775 euro in the school years 2013-14, 2014-15 and 2015-16 respectively. These amounts are fourty percent of the average annual funding per pupil. The corresponding annual spendings on the impulse subsidy scheme range from 128 to 146 million euro. These amounts make up about 2.2 to 2.5 percent of the annual education budgets, which are close to six billion euro in this period. The additional resources that are provided through the impulse subsidy scheme come on top of existing funding schemes.

In this paper, I answer the following research question:

What is the impact of the impulse subsidy scheme in
primary education on pupil performance?

Pupil performance is measured through pupils' scores on national primary school leavers attainment tests ('national attainment tests'). In line with its purpose, I hypothesise that the impulse subsidy scheme increased pupil performance.

A school gets impulse subsidy funding for each disadvantaged pupil only if the school is located in a disadvantaged neighbourhood, called an impulse area. The determination of these impulse areas is based on two neighbourhood characteristics (percentage low income households and percentage welfare dependent households). The eligibility thresholds are very strict. Schools that fall just below and just above the thresholds can be expected to be similar, while the subsidy scheme causes significant differences in the amount of funding that both groups of schools receive. This creates a natural experiment with a treatment group (i.e. schools above, but close to the thresholds) and a control group (i.e. schools below, but close to the thresholds). The econometric specification that can exploit such a natural experiment is called a regression discontinuity design.

Lee and Lemieux (2010) provide an overview of economic literature that use the regression discontinuity design. One of the studies that Lee and Lemieux mention is the paper of Leuven, Lindahl, Oosterbeek, and Webbink (2007). Leuven et al. evaluate two Dutch subsidy schemes using a regression discontinuity framework and data from 1998 to 2003. One subsidy scheme provides extra funding for personnel for schools with at least seventy percent disadvantaged minority pupils. The other subsidy scheme provides extra funding for computers and software for schools with at least seventy percent pupils from any disadvantaged group. For both subsidy schemes, Leuven et al. find negative effects on pupil performance, some of which are significant. In follow-up research, Allaoui (2013) found that the personnel subsidy scheme did not have a positive impact on average teacher experience, teacher remuneration and the total number of full-time equivalent teacher jobs that schools employ.

More recently, at the request of the Ministry of Education, Culture and Science, the Netherlands Bureau for Economic Policy Analysis (CPB)[1] (2017) examined the impulse subsidy scheme that is also studied in this present paper. By visualising data from school and neighbourhood characteristics over the school years 2009-2010 to 2014-2015, it is shown that schools around the welfare dependency threshold are similar and thus that a regression discontinuity design is legitimised. The CPB found that eligible schools get about seven percent more funding, which is used to hire a similar percentage of full-time equivalent teachers (0.8 to 1.0 full-time equivalent teachers per 225 pupils). However, no statistically significant effects are found on any of the available indicators for education success. A suggested explanation for this result is that, contrary to schools that get large amounts of additional funding, for schools around the analysed threshold, the amount of additional funding is too small to substantially reduce class sizes. The CPB emphasises that these results thus may not be representative for the enitre policy. Notably, the CPB uses only one of the two eligibility thresholds in the regression discontinuity design, arguing that there are only few postal code areas that fall below the welfare dependency treshold but above the income treshold.

[1]The original Dutch name of this agency is: *"Centraal Planbureau"*.

Verspaandonk (2016) did address both eligibility thresholds by creating a new binary variable that indicates treatment. In this way, the regular regression discontinuity design (with a single threshold) can be used. Using data from the single school year 2014-2015, Verspaandonk found that the subsidy scheme increased the funding per pupil by 200 to 500 euro for all schools. This finding is similar to the findings of the Netherlands Bureau for Economic Policy Analysis.

This present paper provides an alternative methodology to address both eligibility thresholds. It adds to the literature by using a more select group of observations in the analysis and allowing for differences in the effect of treatment for three different clusters of treated schools. It compensates for the use of less observations per year by using three school years of data covering the period 2013 to 2016.

Using ordinary least squares models, I found that the impulse subsidy scheme significantly increased the funding per pupil of one of the three clusters of treated schools by about 266 euro, all else equal. However, the funding per pupil of two other clusters of treated schools did not significantly increase. This last finding is not in line with the researches of Verspaandonk (2016) and the Netherlands Bureau for Economic Policy Analysis (2017). In addition, I found that some, but not all clusters of treated schools were able to reduce the average class sizes by about 1.25 pupils, all else equal. This is higher than the reduction of 0.4 to 0.9 pupils that was found by Verspaandonk. These differences in conclusions may be explained by the fact that this present research uses two more years of data and uses a methodology that analyses a more select group of observations per year.

None of the three clusters of treated schools show a significant increase in pupils' standardised test scores at the school level. The hypothesis that the impulse subsidy scheme increased pupil performance is thus rejected. This conclusion is in line with Leuven et al. (2007), who found no evidence of a positive effect on pupil performance of a subsidy scheme in a similar context. It is also in line with the conclusion of Verspaandonk (2016), who researched the same subsidy scheme.

The next section provides some context of the legislation surrounding the impulse subsidy scheme. Section 3 describes the data. Section 4 explicates the empirical strategy. Section 5 presents the results. Section 6 concludes and discusses.

2 Context

2.1 Legislation

Impulse areas are neigbourhoods that are defined by the government as disadvantaged neigbourhoods based on two characteristics.[2] A neighbourhood gets the impulse area status if either or both

1. the percentage of households with low incomes is equal to or above 11.5 percent, or
2. the percentage of households that depend heavily on financial government support ('welfare dependency') is equal to or above 11.3 percent.

Neighbourhood data from 2005 is used to determine the impulse area status. In determining impulse areas, there is a trade-off between up-to-date data and legal certainty for schools. The government chose the latter over the former and decided that the impulse areas are determined for periods of four years. The list of impulse areas has not changed since its introduction in 2009.

The two eligibility thresholds are both close to the 80th percentile of the respective characteristics. In other words, the impulse area status is assigned to the approximately twenty percent areas with the lowest incomes and the approximately twenty percent areas with the highest welfare dependency. In total, 1,013 of all 3,605 neighbourhoods ($\approx$ 28.1 percent) are impulse areas.

2.2 Dutch primary education

In general, Dutch children aged 4 to 12 are legally obligated to go to primary schools. All types of pupils, educationally strong and weak, are in the same class. The education system is heavily subsidised by the government. This comes with heavy monitoring, which also results in public access to some school data. Primary schools are funded predominantly based on the number of pupils.

[2]These neighbourhoods are four digit postal code areas. Dutch postal codes consist of four digits and two letters. Only the four digits are used to determine the impulse area status.

2.3 National attainment tests

An accurate description of national attainments tests is provided by the Dutch government on its website (2017a):

> *"The attainment test shows whether pupils have attained benchmark levels for language and number skills. It also indicates what type of secondary education would be most suitable for the individual pupil. So the test is also a tool for ensuring a smooth transition between primary and secondary education. The primary school leavers attainment test is not an exam; pupils cannot pass or fail it."*

The period for taking national attainment tests is 15 April to 15 May. As of the school year 2014-2015, schools are obligated to let their pupils take national attainment tests (Dutch Government, 2014), which schools do not have to pay for (Dutch Government, 2017b). Before that year, many schools voluntarily let their pupils take national attainment tests. As of one year after the obligation, schools are bound to the attainment tests that are prescribed by the government. As national attainment tests can be used to compare schools in terms of pupil performance on a national scale, they are used as the measure of pupil performance in this paper. Only one national attainment test, the central end test[3], is used sufficiently wide and consistent to be useful in this analysis in terms of number of observations.[4]

2.3.1 Central end test

The central end test is provided by the government itself: it is made by the Board for Tests and Exams[5] which is commissioned by the Ministry of Education, Culture and Science. The central end test is issued in two levels. Pupils that are expected to perform in a lower segment are handed the lower level version of the central end test. This prevents the test score from becoming inaccurate, as these pupils are unlikely to correctly answer the hard questions of the regular version of the central end test. The difference in test level is accounted for in the determination of the central end test score, which is measured in a scale from 501 to 550.[6]

[3]The central end test is also known as its predecessor: the *"CITO test"*.

[4]The total number of observations of the other three large national attainment tests over the course of the analysed three years, is just under 13.4 percent of the number of observations of the central end test in the same period.

[5]The original Dutch name of this board is: *"College voor Toetsen en Examens"*.

[6]This scale is chosen to avoid that the scores are interpreted as IQ scores or regular Dutch grades, which range from 1 to 10 (Squla, 2017).

The central end test consists of two compulsory components and one optional compontent (Cito, 2017). The compulsory components are language (135 exercises) and arithmetics (85 exercises). Pupils work two hours per day on these exercises for three days. The optional component is world orientation (90 questions), for which pupils get an extra fourty minutes per day.[7]

Only in extreme cases can pupils be exempted from taking national attainment tests. Examples of such cases are pupils that have severe learning difficulties or multiple disabilities (Dutch Government, 2017c). The Central End Test takes into account dyslexia, colour blindness and different backgrounds of pupils (Cito, 2017). For blind pupils, the test is available in braille. Pupils are not allowed to resit a national attainment test (Dutch Government, 2016), but each national attainment test does have a second test period to accommodate pupils that were absent during the first test period.

3 Data

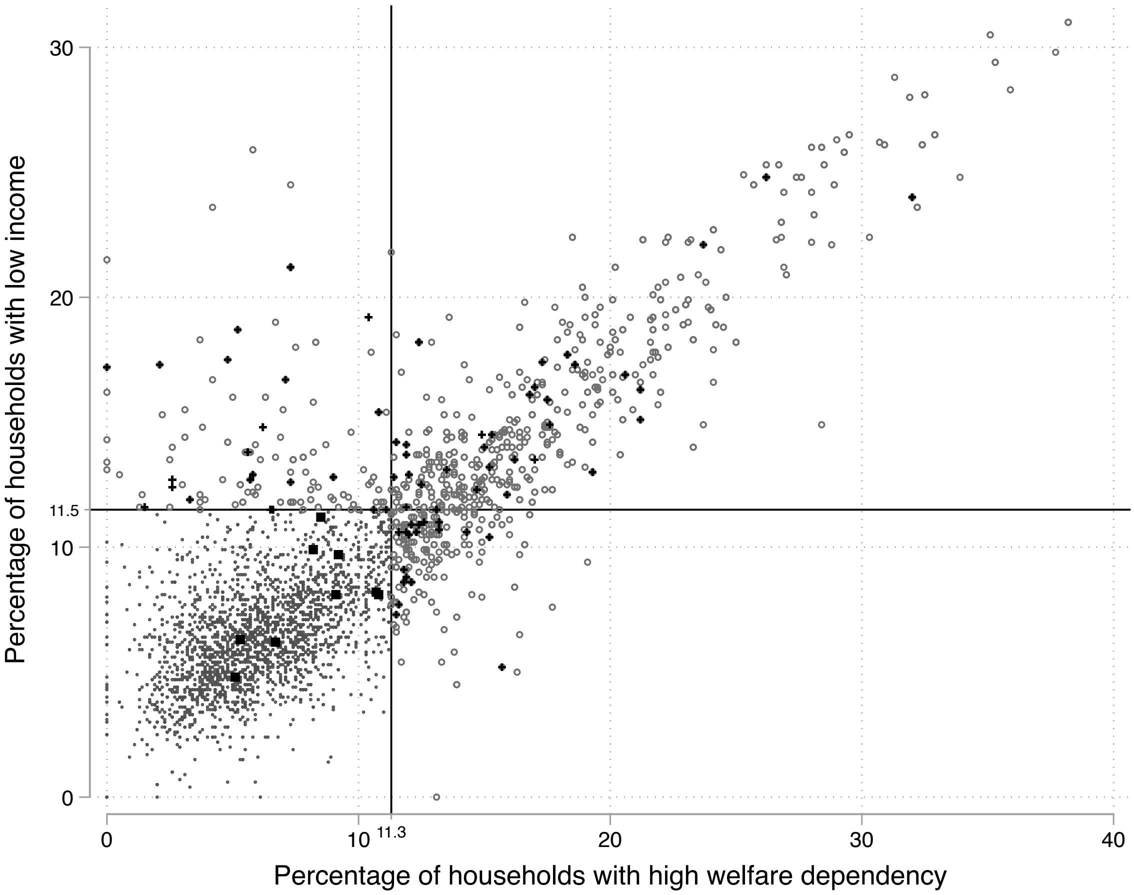
The dataset that is used in this paper is constructed using publicly available data from the website of the Dutch Ministry of Education, Culture and Science (2017) and nonpublic data from Statistics Netherlands.[8] The nonpublic data contains all four digit postal code areas that were in use in 2005, along with impulse area status. The public data contains all other data that is used in this paper.

Figure 3.1 visualises all primary schools in the Netherlands based on the two neighbourhood characteristics that determine impulse subsidy eligibility (percentage of households with low income and percentage of households with high welfare dependency). Some schools in impulse areas receive regular funding because the number of disadvantaged pupils is zero. Due to the eligibility rules, this perfectly fits within the impulse subsidy scheme. It is the case for 77 unique schools in total, 22 of which (totalling 36 observations) are used in the analysis in this paper. However, it is considered to be a mistake that some schools in regular areas receive impulse funding. This is the case for nine unique schools in total, two of which are used in the analysis of this paper (both observations occur in the school year 2013-2014).

[7] World orientation includes excersises on geography, history and nature.

[8] The nonpublic data is provided to me by Erasmus University alumnus Marijn Verspaandonk.

FIGURE 3.1: REGULAR AND IMPULSE AREAS



Note: Schools that are located in regular areas (i.e. the control group) are represented by a grey dot (·) if they received regular funding and represented by a solid black square (■) if they received impulse funding; schools that are located in impulse areas (i.e. the treatment group) are represented by a grey circle (○) if they received impulse funding and a black plus (+) if they received regular funding. The continuous horizontal and vertical lines at 11.5% and 11.3%, respectively, represent the thresholds that determine the impulse area status. Data from three school years (2013 to 2016) at the 4 digit postal code level.

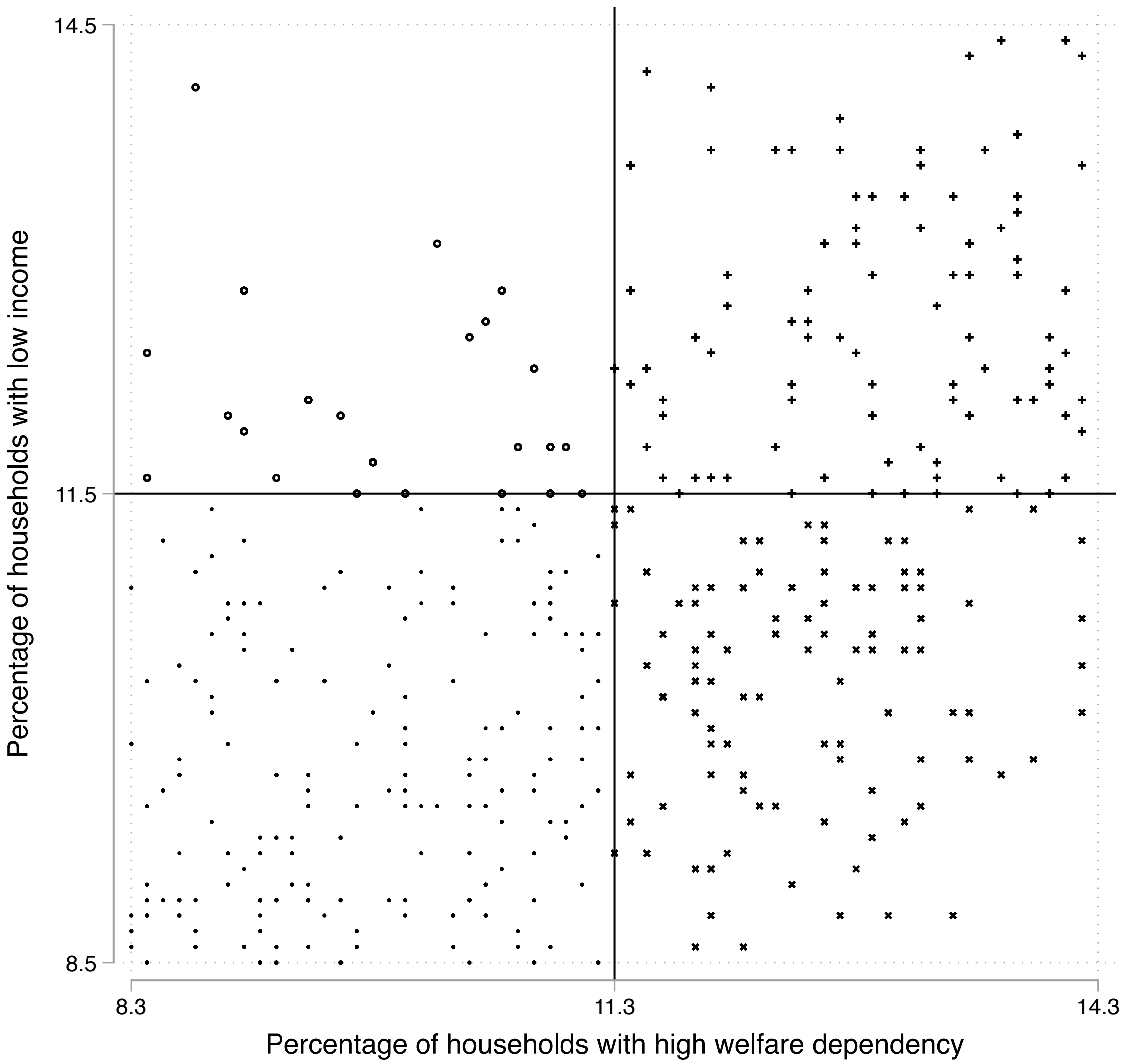
Not all observations will be used in the analysis of this paper, as will be discussed in section 4. The observations that are included in the analysis lie close (i.e. within three percentage points distance) to the point where both impulse area eligibility thresholds intersect. This point is visualised in figure 3.1 by the intersection of the two continuous lines (which represent the thresholds). The thresholds lie at 11.5 percent (households with low income) and 11.3 percent (households with high welfare dependency). The used observations thus lie in postal code areas with more than 8.5% (= 11.5% - 3%) and less than 14.5% (= 11.5% + 3%) low incomes, and more than 8.3% (= 11.3% - 3%) and less than 14.3% (= 11.3% + 3%) welfare dependency. These observations are visualised in figure 3.2, which shows four quadrants. Within the southwest quadrant lie observations that fall below both thresholds and thus are regular areas. All other observations meet either one or both of the impulse area eligibility criteria and thus are impulse areas.

Table 3.1 shows the number of observations in each quadrant per year. There is a large gap between the number of observations of the northwest and the southwest quadrants. There is no clear explanation for this finding. This is simply how the four digit postal code areas are distributed based on the two eligibility characteristics.

Table 3.2 provides descriptive statistics that are derived from the observations that are used in the analysis, discriminated by quadrant. The total number of pupils at the school appear similar. The funding per pupil has a higher mean for each of the treatment quadrants relative to the control quadrant. This is as expected. The total amount of impulse subsidy money that is actually transferred to each quadrant is: southwest 116.610 euro, northwest 6.350.212 euro, northeast 40.585.512 euro, southeast 43.098.504 euro. There are only two schools that are located in the control quadrant that received impulse subsidy funding.[9] The northwest quadrant has a lower mean of the relative share of impulse subsidy funding than the other two treatment quadrants. This is not as expected. If the difference in means is significant, this might suggest that schools in the northwest quadrant are less able to increase performance. Schools in the treatment quadrants have lower means of class size and teacher age. Surprisingly, only the northwest quadrant has a higher mean of performance than the control quadrant: the northeast and southeast quadrants have a lower mean of performance than the control quadrant. Note that these are just averages. The controlled impact of the impulse subsidy scheme on

[9] The share of this impulse subsidy funding in the total funding of one school is visualised as maximum in the southwest quadrant (5.77 percent). The relative share of the other school is 3.68 percent.

FIGURE 3.2: REGRESSION DATAPOINTS SPLIT INTO FOUR QUADRANTS



Note: Schools that are located in regular areas (i.e. the control group) are represented by a dot (·) in the southwest quadrant; schools that are located in impulse areas are represented by respectively a circle (○) in the northwest quadrant, a plus (+) in the northeast quadrant and a cross (×) in the southeast quadrant. The continuous horizontal and vertical lines at 11.5% and 11.3%, respectively, represent the thresholds that determine the impulse area status. The presented datapoints are used in the regressions in this paper and are located in the area ±3 percentage points around both thresholds. Data from three school years (2013 to 2016) at the 4 digit postal code level.

performance is discussed in section 5.

TABLE 3.1: NUMBER OF OBSERVATIONS PER QUADRANT PER YEAR

	2013-14	2014-15	2015-16
Southwest (regular areas)	354	344	305
Northwest (impulse areas)	48	40	36
Northeast (impulse areas)	168	171	141
Southeast (impulse areas)	207	217	193
Total	777	772	675

The national attainment test scores are standardised over all observations (not just the observations that are used in the analysis) to have a mean of zero and a standard deviation of one in every year. Besides, due to privacy considerations, observations that are made up of less than five test-taking pupils are not available in this public dataset and thus cannot be used in the analysis.

The use of national attainment test scores as the measure of pupil performance comes with the assumption that pupils who change from a treatment school to a nontreatment school, or vice versa, do not significantly influence the results. Another assumption on the data is one on continuity of postal code–province combinations. Gaps in postal code and province data in observations from 2013-2014 were filled using 2014-2015 data (or, if that was not sufficient, 2015-2016 data). In filling the gaps, I implicitly assumed that these characteristics did not change over the years.

4 Empirical Strategy

In general, a sharp regression discontinuity design is used when treatment status is a deterministic and discontinuous function of a single covariate (Angrist & Pischke, 2008). However, in the impulse subsidy scheme there are not one, but two covariates that determine the treatment status of a school. This makes it difficult to apply the regular methodology of a sharp regression discontinuity design. The two covariates are the percentage of low income households,

TABLE 3.2: DESCRIPTIVE STATISTICS

	Number of Pupils	Funding per Pupil	% Impulse Funding	Class Size	Teacher Age	Performance
Southwest (control quadrant)						
Mean	230.64	4044.66	.0094	16.42	41.28	-.0143
Standard Deviation	125.33	680.05	.22	3.20	3.61	1.0144
Minimum	19	3105.18	0	1.20	30	-7.5333
Median	206	3859.32	0	16.60	41.56	.0740
Maximum	815	12642.21	5.77	30.48	50	2.7014
Northwest (treatment quadrant)						
Mean	251.27	4127.12	4.20	15.43	39.64	.0536
Standard Deviation	140.51	662.37	2.97	2.86	4.02	1.1750
Minimum	34	3409.95	0	7.12	30.82	-6.9663
Median	235	3968.83	3.70	15.77	40.01	.1568
Maximum	723	7521.65	12.63	22.26	48.81	1.9779
Northeast (treatment quadrant)						
Mean	229.30	4522.66	6.57	14.64	40.59	-.2660
Standard Deviation	133.40	849.56	4.18	3.67	3.85	1.0566
Minimum	26	3296.00	0	2.78	30.28	-4.0299
Median	208.5	4292.17	6.01	14.78	41.19	-.1616
Maximum	846	9656.74	17.14	39.42	49.93	3.1114
Southeast (treatment quadrant)						
Mean	222.05	4429.05	6.06	15.35	40.76	-.2327
Standard Deviation	121.25	846.43	3.99	3.42	3.65	.9791
Minimum	30	3093.57	0	4.59	30	-4.5565
Median	203	4183.60	5.68	15.75	41.13	-.1663
Maximum	752	10247.59	17.47	27.67	48.95	2.2712

Note: The presented statistics are are derived from the observations that are used in the regressions of this paper. These observations cover three school years (2013-2016) and are visualised in figure 3.2. Percentage impulse funding is defined as the total amount of impulse subsidy money divided by the total funding of the school, and is scaled from 0 to 100. Class size is defined as the number of pupils divided by the number of full-time teacher equivalents at the school. Teacher age is a weighted average that takes into account the amount of working hours and has a minimum of 30 and a maximum of 60. Performance is standardised over the entire dataset to have a mean of zero and a standard deviation of one.

I, and the percentage of welfare dependent households, W, in the postal code area in which a school is located. Using the threshold values of I and W, the dummy variable that indicates treatment status can thus be defined as:

$$\text{IMPULSE AREA} = \begin{cases} 1, & \text{if } I \geq 11.5\% \text{ or } W \geq 11.3\% \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

The regression discontinuity design comes with the assumption that there are no confounding discontinuities at the thresholds. In other words, the exact values of the thresholds are assumed to be arbitrary. If this assumption holds, schools in areas close to the thresholds are comparable in every way except treatment (i.e. receiving impulse subsidy money). This creates a natural treatment group and control group, so that observed differences between the two groups can be interpreted as the causal effect of treatment.

Areas with similar values of I and W are expected to be similar in other characteristics as well. This can be tested using observable characteristics. If the observable characteristics of areas around the thresholds are similar, it is considered to be likely that unobservable characteristics are similar as well. This is the basis of the regression discontinuity design. To ensure that the treatment group and the control group are sufficiently similar, the observations that are used in the analysis should lie sufficiently close to the thresholds. However, moving closer to the thresholds comes at the cost of losing observations and thus the estimation of the effect of treatment becomes less precise. Observations that lie further away from the thresholds can be added in order to increase precision. However, this is likely to come with bias, which decreases the accuracy of the estimation of the effect of treatment. The choice of what observations to use in the analysis is thus a trade-off between the accuracy and the precision of the effect of treatment. Based on tests of Calonico, Cattaneo, and Titiunik (2014) and in line with previous reserach (Verspaandonk, 2016), I use observations that lie within a three percentage points range around both thresholds (see section A.1 on page 26).

Based on recent data at the four digit postal code level, the Netherlands Bureau for Economic Policy Analysis (2017) showed that schools that are located in areas around the welfare threshold are not significantly different. This legitimises the use of the regression discontinuity design for the analysed data, which covers the period 2009-2014. Without having access to recent data at the four digit postal code level, this present research can legitimise the use of the regression discontinuity design for the period 2013-2016 by testing whether background characterisics at the school level have significantly changed since the school year 2013-2014. In addition, it can be tested whether the treatment group and the control group show significant differences in these background characteristics. To legitimise the use of the regression discontinuity design, there should be no significant differences between the treatment group and the control group (apart from the treatment). The background characteristics are the school denomination and the total number of pupils at the school. These background characteristics should not be significantly influenced by the treatment in order to be informative for these tests. This is why, for example, class size data is not useful for these tests. For the data on total number of pupils, Mann-Whitney U (rank sum) tests are performed. For the data on school denomination, Fisher-Exact tests are performed. Both tests compare the means of two groups. None of the tests find significant differences, even at the ten-percent level. This means that the number of pupils and the distribution over denominations of each quadrant did not significantly change in 2015 and 2016 compared to 2014. In addition, there are no significant differences between the four quadrants in terms of number of pupils and distribution of denominations. The regression discontinuity design is thus considered to be legitimised.

In a regular regression discontinuity design, a distribution test is performed to test if the treatment group is much larger than the control group, as this would be an indication that schools self-selected into treatment. It is expected that no such effect is found in the present analysis, because subsidy eligibility is area-based. Only new schools that determine where to be located, or schools that move across a four digit postal code area border, could self-select into treatment. Still, it seems highly unlikely that the impulse status of a four digit postal code area is decisive in the determination of school location, particularly since the list of impulse areas can be changed every four years. Table 3.1 shows that all treatment quadrants have less observations than the control quadrant. This is what would be observed if self-selection into treatment does not pose a problem.

I identify the treatment effect using ordinary least squares regressions. Firstly, I identify the combined treatment effect using a single dummy variable (IMPULSE AREA). Secondly, I split up the effect of interest by using dummy variables for each treatment quadrant in figure 3.2 (i.e. NORTHWEST, NORTHEAST and SOUTHEAST). These dummy variables equal one if the observation lies within the respective quadrant and zero otherwise. The base group is SOUTHWEST, which is the control group. This second step allows for differences in the effect of treatment for different clusters of impulse areas that are relatively alike in terms of the eligibility characteristics (low incomes and welfare dependency).

These two specifications of the effect of interest are used in regressions for four dependent variables. The first three dependent variables are funding per pupil, average class size and weighted teacher age. These regressions are robustness checks. The final dependent variable is the standardised pupil test score on national attainment tests at the school level ('performance'). These regressions are the core of this paper.

The regression that identifies a single treatment effect is specified as

$$\mathrm{Y}_{it} = \alpha + \beta\textsc{impulse area}_i + \mathbf{X}_{it} + \gamma_t + \varepsilon_{it} \tag{4.2}$$

where Y_{it} is FUNDING PER PUPIL, average CLASS SIZE, weighted TEACHER AGE or pupil PERFORMANCE of school i at year t, α is a constant, β is the coefficient of interest, IMPULSE AREA$_i$ is a binary variable that equals one if school i is located in an impulse area and zero otherwise,[10] $\mathbf{X}_{it}$ is a vector of control variables containing linear and quadratic terms of the running variables (percentage low income and percentage high welfare),[11] province, school size and denomination, γ_t captures time fixed effects, and ε_{it} is an error term. Robust standard errors are used in all regressions in this paper.

Equation 4.3 is similar to equation 4.2, but differs in the identification of the effect of interest, which is split into three.

$$\mathrm{Y}_{it} = \alpha + \beta_1\textsc{northwest}_i + \beta_2\textsc{northeast}_i + \beta_3\textsc{southeast}_i + \mathbf{X}_{it} + \gamma_t + \varepsilon_{it} \tag{4.3}$$

where β_1, β_2 and β_3 are the coefficients of interest and NORTHWEST$_i$, NORTHEAST$_i$ and SOUTHEAST$_i$ are dummy variables that equal one if the area lies within the respective quadrant and zero otherwise. The base group of this regression are observations that lie in the southwest quadrant. These are regular areas (i.e. the control group).

[10] The variable IMPULSE AREA only has subscript i, because impulse area status is constant over time in this dataset.

[11] Following the recommendation of Gelman and Imbens (2017) and in line with previous research (Verspaandonk, 2016), only linear and squared polynomials are used in the main regressions of this paper. Given that the running variables are controlled for, the effect of interest is robust to different degrees of polynomials (see section A.2 on page 29).

5 Results

TABLE 5.1: FUNDING IN EURO

	(1)	(2)	(3)	(4)
Impulse area	412.8***	195.5***		
	(54.52)	(71.50)		
Northwest (impulse areas)			82.35	40.51
			(106.2)	(112.4)
Northeast (impulse areas)			529.1***	156.2
			(74.59)	(137.9)
Southeast (impulse areas)			391.2***	266.2***
			(67.07)	(83.16)
Additional controls	No	Yes	No	Yes
Observations	2224	2208	2224	2208

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. The dependent variable of all models in this table is funding per pupil in euro at the school level. Even numbered models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

The results are presented in four tables. All tables present the effect of treatment in two forms. The first form is the combined treatment effect. The second form is the treatment group split up into three clusters (northwest, northeast and southeast), where the base is still the control group (southwest). Each regression is presented with and without control variables. All models are individually numbered, where odd numbers present raw regression models and even numbers present controlled regression models.

5.1 Funding

As a robustness check, table 5.1 presents the impact of the impulse subsidy scheme on the funding per pupil at the school level. Model 1, which includes no controls, shows that there is a strong combined treatment effect that is positive and significant. Model 2, which does include controls, shows that the analysed schools in impulse areas jointly have about 196 euro

TABLE 5.2: CLASS SIZE

	(5)	(6)	(7)	(8)
Impulse area	−1.438***	−0.969***		
	(0.214)	(0.306)		
Northwest (impulse areas)			−0.929**	−0.520
			(0.400)	(0.485)
Northeast (impulse areas)			−1.925***	−1.250**
			(0.298)	(0.560)
Southeast (impulse areas)			−1.159***	−1.242***
			(0.255)	(0.363)
Additional controls	No	Yes	No	Yes
Observations	2150	2140	2150	2140

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. The dependent variable of all models in this table is average class size at the school level. Class size is defined as the total number of pupils divided by the total number of full-time teacher equivalents Even numbered models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

more funding per pupil on average than regular areas, all else equal. This effect is significant at the one-percent level.

Models 3 and 4 in table 5.1 specify the impact of treatment on the funding per pupil over the three treatment quadrants (see figure 3.2 for a visualisation of all four quadrants). The controlled model, model 4, shows that the impulse subsidy scheme significantly increased the funding per pupil in the southeast quadrant by about 266 euro, all else equal, at the one-percent level. Schools in the southeast quadrant are located in areas with relatively high percentages of welfare dependent households and relatively low percentages of low income households, compared to the other schools in the analysis. The other two treatment quadrants do have positive point estimates that are not significantly different from zero. The combined treatment effect of model 2 thus originates mainly from the southeast quadrant.

Some, not all, of the analysed schools that are located in impulse areas received significantly more money per pupil than schools that are located in regular areas. This difference within treated schools was not expected. It stresses the relevance of the distinction between different clusters of treated schools that is made in this paper. Given that only the southeast quadrant received significantly more money than the base group, it may be expected that schools in the southeast quadrant are better able to increase pupil performance than the other two treatment

TABLE 5.3: TEACHER AGE

	(9)	(10)	(11)	(12)
Impulse area	−0.650***	−1.019***		
	(0.238)	(0.383)		
Northwest (impulse areas)			−1.672***	−1.025*
			(0.539)	(0.621)
Northeast (impulse areas)			−0.597*	−0.740
			(0.319)	(0.678)
Southeast (impulse areas)			−0.478*	−0.970**
			(0.283)	(0.457)
Additional controls	No	Yes	No	Yes
Observations	2224	2208	2224	2208

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses. The dependent variable of all models in this table is weighted teacher age at the school level. The dependent variable takes into account the amount of working hours and has a minimum of 30 and a maximum of 60. Even numbered models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

quadrants (northwest and northeast).

5.2 Class size

Table 5.2 presents the impact of the impulse subsidy scheme on the average class size at the school level. Model 5, which includes no controls, shows that there is a strong combined treatment effect that is negative and significant. Model 6, which does include controls, shows that the analysed schools in impulse areas jointly have on average about one less pupil per full-time teacher equivalent than the analysed schools in regular areas, all else equal. This effect is significant at the one-percent level.

Models 7 and 8 specify the impact of treatment on the average class size over the three treatment quadrants. The controlled model, model 8, shows that the impulse subsidy scheme significantly decreased the class size of schools in the northeast and southeast quadrants by about 1.25 pupils, all else equal. These effects are significant at respectively the five-percent and one-percent level. Schools in the northeast and southeast quadrants are located in areas with relatively high percentages of welfare dependent households, compared to the other schools in the analysis. The northwest quadrant does have a negative point estimate, which indicates

TABLE 5.4: STANDARDISED PUPIL PERFORMANCE

	(13)	(14)	(15)	(16)
Impulse area	−0.217***	−0.0176		
	(0.0539)	(0.0862)		
Northwest (impulse areas)			0.116	0.110
			(0.0990)	(0.128)
Northeast (impulse areas)			−0.292***	−0.00437
			(0.0746)	(0.151)
Southeast (impulse areas)			−0.226***	−0.0785
			(0.0640)	(0.102)
Additional controls	No	Yes	No	Yes
Observations	2224	2208	2224	2208

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors in parentheses. The dependent variable of all models in this table is standardised pupil performance at the school level. Even numbered models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

that the average class size for this quadrant is lower than the control group, but this estimate is not significantly different from zero. The combined treatment effect of model 5 thus originates mainly from the northeast and southeast quadrants.

The analysed schools that received impulse subsidy money and are located in relatively highly welfare dependent areas, were able to successfully spend this money on reducing class sizes (i.e. hiring additional teachers or letting present teachers work more hours).

5.3 Teacher age

Table 5.3 presents the impact of the impulse subsidy scheme on the average weighted teacher age at the school level. This variable takes into account the amount of working hours and has a minimum of 30 and a maximum of 60. Model 10, which includes controls, shows that analysed schools in impulse areas jointly have teachers that are on average about one year younger than analysed schools in regular areas, all else equal. This effect is significant at the one-percent level.

Models 11 and 12 specify the impact of treatment on the weighted teacher age over the three treatment quadrants. The controlled model, model 12, shows that the impulse subsidy scheme decreased the weighted teacher age of schools in the southeast quadrant by about one year.

This effect is significant at the five-percent level. A similarly sized effect is found for schools in the northwest quadrant. However, this effect is only significant at the ten-percent level. The subsidy scheme did not significantly reduce the teacher age for schools in the northeast quadrant. It is notable that the point estimate is negative and larger than the corresponding standard error.

There is some evidence that indicates that the impulse subsidy funding was used to attract new teachers that are younger than the weighted average teacher age at the schools, or letting the present young teachers work more hours.

5.4 Performance

Table 5.4 presents the impact of the impulse subsidy scheme on standardised pupil performance at the school level. Model 13 and 14 present the combined treatment effect of the subsidy scheme on the standardised test scores. Model 13, which includes no controls, shows that the treated schools jointly score on average more than 0.2 standard deviation worse than non-treated schools. This effect is significant at the one-percent level. However, when controls are included, the coefficient size drops and the coefficient is no longer significantly different from zero. Thus, jointly, the eligible schools do not perform better than the noneligible schools.

Models 15 and 16 specify the impact of treatment on standardised pupil performance over the three treatment quadrants. The uncontrolled model, model 16, shows that the negative change in performance originates from the northeast and southeast quadrants. When controls are introduced, in model 16, the coefficient size of the northeast quadrant becomes very small with a large standard error. The coefficient of the southeast quadrant is now also insignificant, although the size of the negative coefficient is closer to the corresponding standard error. It is compelling that the southeast quadrant, which showed a large and significant increase in funding per pupil, a large and significant reduction in average class size, shows a negative coefficient. Contrary, the northwest quadrant, which showed no significant increase in funding and no significant reduction in average class size, shows a positive coefficient of performance. The coefficient size is about 0.1 standard deviation, which is close to its standard error. It is not significantly different from zero.

There is no clear positive effect of treatment. The hypothesis that the impulse subsidy scheme increases performance is thus rejected. In other words, the analysed schools that received impulse subsidy funding did not show an observable and significant increase in performance.

5.5 Heterogeneity analysis

It might be that the impact of treatment differs depending on some observable characteristic of the school. For example, it might be that the impulse subsidy scheme is most effective for schools that have the largest average class sizes. However, it is not possible to derive such a causal effect from the data as treatment may directly influence class size. Table 5.2 shows that this is the case. Using pretreatment data can be a solution. However, as the dataset that is used in this paper does not include such data, average class size and average teacher age cannot be used for a heterogeneity analysis.

Fortunately, data on school denomination is available. Due to its nature, school denomination is not influenced by the treatment. As it is likely that unobservable characteristics differ between pupils that go to schools with different denominations, it might be that schools with different denominations react differently to the treatment. There are four types of school denomination: public, Protestant, Catholic and residual category other. Public schools are the base category. Table 5.5 presents the impact of the impulse subsidy scheme on standardised pupil performance at the school level. Model 17 is an exact copy of model 16 in table 5.4. Model 28 includes interaction terms of each treatment quadrant and school denomination.

Model 18 in table 5.5 shows that the impact of the impulse subsidy scheme on Protestant schools in the northeast quadrant is almost half a standard deviation higher compared to public schools in that quadrant. This effect is significant at the one-percent level. However, across quadrants, Protestant schools do not consistently react stronger to treatment relative to public schools: the interaction coefficients of Protestant schools in the northwest and souteast quadrants are not significantly different from zero and are positive and negative respectively. It may seem that Protestant schools react relatively well to the treatment in terms of performance, but this finding is insufficiently consistent to draw a general conclusion.

TABLE 5.5: STANDARDISED PUPIL PERFORMANCE: QUADRANT INTERACTED WITH SCHOOL DENOMINATION

	(17)	(18)
Northwest (treatment quadrant)	0.110	0.0488
	(0.128)	(0.200)
Northeast (treatment quadrant)	−0.00437	−0.185
	(0.151)	(0.178)
Southeast (treatment quadrant)	−0.0785	−0.0211
	(0.102)	(0.149)
Northwest (treatment quadrant) × Protestant school		0.250
		(0.279)
Northwest (treatment quadrant) × Catholic school		−0.0506
		(0.255)
Northwest (treatment quadrant) × Other denomination		0.143
		(0.253)
Northeast (treatment quadrant) × Protestant school		0.481***
		(0.170)
Northeast (treatment quadrant) × Catholic school		0.232
		(0.176)
Northeast (treatment quadrant) × Other denomination		0.243
		(0.245)
Southeast (treatment quadrant) × Protestant school		−0.160
		(0.169)
Southeast (treatment quadrant) × Catholic school		0.00310
		(0.151)
Southeast (treatment quadrant) × Other denomination		0.0289
		(0.203)
Additional controls	Yes	Yes
Observations	2208	2208

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. The dependent variable of all models in this table is standardised pupil performance at the school level. All models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

6 Conclusion and Discussion

In this paper, I evaluated the Dutch impulse area subsidy scheme that targets disadvantaged pupils in primary education. The research question was: *What is the impact of the impulse subsidy scheme in primary education on pupil performance?* In line with the purpose of this subsidy scheme, I hypothesised that the extra funding has a positive impact on pupil performance. The eligibility rules of the impulse subsidy scheme are very strict: only disadvantaged pupils that go to schools that are located in impulse areas are substantially subsidised. These strict eligibility rules allowed for a regression discontinuity design. I used school-level government data from three school years covering the period 2013 to 2016. This research is distinctive in the method of addressing the two eligibility thresholds: I use a select group of observations, which I divide into a control group and three clusters of treated schools that are relatively alike. This method allows for differences in the effect of treatment.

Using ordinary least squares models, I found that the impulse subsidy scheme significantly increased the funding per pupil of only one of the three clusters of treated schools by about 266 euro on average, all else equal. The two other clusters of treated schools show positive, but, unexpectedly, nonsignificant point estimates. Most of the analysed treated clusters of schools were able to successfully spend the impulse subsidy funding on reducing class sizes (i.e. hiring additional teachers or letting present teachers work more hours). There is some evidence that indicates that the impulse subsidy funding was used to attract new teachers that are relatively young, or letting relatively young teachers work more hours, resulting in a decrease of the weighted average age by about one year on average, all else equal. There is a cluster of analysed Protestant schools that reacts relatively well to the treatment in terms of pupil performance. However, this finding is insufficiently consistent to draw a general conclusion.

There is no clear effect of the impulse subsidy scheme. The analysed schools that received impulse subsidy funding did not show an observable and significant increase in pupil performance at national attainment tests. The hypothesis that the impulse subsidy scheme increased pupil performance is thus rejected. This finding is complementary to findings of existing literature.

An explanation for why the effects of the impulse subsidy scheme are not being as intended might be decreasing marginal returns to funding: the schools that are targeted by the impulse subsidy scheme already receive additional funding for disadvantaged pupils by another subsidy scheme. An alternative explanation might be that the money was spent on additional teachers that are relatively young and less experienced. Hence, the added value of the new teachers would be low. The results of this research provide some indication that suggests that schools used the impulse subsidy money to hire relatively young teachers.

It is important to note that the conclusions of this research are based on the analysed schools. Due to the nature of the regression discontinuity design, only a select group of observations is included in the analysis. Therefore, the conclusions of this research cannot be expanded to the entire impulse subsidy scheme. Particularly because schools further away from the eligibility thresholds generally have relatively more eligible disadvantaged pupils. Consequently, the impulse subsidy funding is a larger share of the total funding of these schools. For these schools, the effect of receiving this money might thus be different from the conclusions of this paper.

Given that the effects of this subsidy scheme in its current form on the analysed schools are not as intended, it may be possible to increase social benefit. For example, it might be beneficial to increase the minimum amount of additional funding. This could be the subject of future research. In addition, I recommend future research to focus on developments around primary school legislation in regard of extra funding for disadvantaged pupils. Changes in policy may provide for other natural experiments that allow for the identification of the causal effect of the impulse subsidy scheme on schools that fall outside the regression discontinuity design analysis.

References

Allaoui, F. (2013). *Where did the subsidy go? the effect of extra funding for schools with disadvantaged students on teacher retention and hiring rates.* Retrieved from http://hdl.handle.net/2105/15099 (Master's thesis, Erasmus University, Rotterdam, the Netherlands)

Angrist, J. D., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Cito. (2017). *De toets uitgelegd.* Retrieved 06 August 2017, from http://www.cito.nl/onderwijs/primair%20onderwijs/centrale_eindtoets/de-toets-uitgelegd

Dutch Government. (2008). Aanpak onderwijsachterstanden. *Tweede Kamer, 2007-2008 Kamerstuk 27 020, No. 57.* Retrieved 06 August 2017, from https://zoek.officielebekendmakingen.nl/kst-27020-57.html

Dutch Government. (2014). *Toegelaten eindtoetsen basisonderwijs bekend.* Retrieved 06 August 2017, from https://www.rijksoverheid.nl/actueel/nieuws/2014/11/12/toegelaten-eindtoetsen-basisonderwijs-bekend

Dutch Government. (2016). *Regeling leerresultaten po 2014.* Retrieved 10 August 2017, from http://wetten.overheid.nl/BWBR0035918/2016-01-01

Dutch Government. (2017a). *Compulsory primary school leavers attainment test.* Retrieved 06 August 2017, from https://www.government.nl/topics/primary-education/compulsory-primary-school-leavers-attainment-test

Dutch Government. (2017b). *Keuzemogelijkheden verplichte eindtoets basisschool.* Retrieved 06 August 2017, from https://www.rijksoverheid.nl/onderwerpen/toelating-middelbare-school/keuzemogelijkheden-verplichte-eindtoets-basisschool

Dutch Government. (2017c). *Welke leerlingen hoeven geen eindtoets te maken?* Retrieved 06 August 2017, from https://www.rijksoverheid.nl/onderwerpen/

toelating-middelbare-school/vraag-en-antwoord/welke-leerlingen-hoeven-geen-eindtoets-te-maken

Dutch Ministry of Education, Culture and Science. (2017). *Databestanden open onderwijsdata.* Retrieved May 2017, from https://duo.nl/open_onderwijsdata/databestanden/po/

Gelman, A., & Imbens, G. (2017). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*(just-accepted).

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, *89*(4), 721–736.

Netherlands Bureau for Economic Policy Analysis. (2017). *Een empirische evaluatie van het onderwijsachterstandenbeleid in het primair en voortgezet onderwijs.*

Squla. (2017). *Eindtoets groep 8 in 2017: een overzicht voor ouders.* Retrieved 06 August 2017, from https://www.squla.nl/eindtoets-groep-8-2017-een-overzicht

Verspaandonk, M. J. (2016). *The effect of six year subsidizing primary schools in poor neighborhoods.* (Unpublished master's thesis, Erasmus University, Rotterdam, the Netherlands)

A Robustness checks

A.1 Bandwidth

Table A.1 presents the impact of the impulse subsidy scheme on standardised pupil performance at the school level. Models 19 to 23 show the impact of changing the bandwidth (in terms of percentage points) of used observations around the two eligibility thresholds on the effect of interest. Model 21 includes observations that fall within a three-percentage points bandwidth around the two eligibility thresholds. This is the bandwidth that is used in the main regressions of this paper: model 21 is an exact copy of model 16 in table 5.4. A bandwidth of three percentage points is supported by a Calonico et al. (2014) test based on performance data from the entire dataset (not only the data close to the thresholds). However, this test uses a single threshold. In order to be able to use it for the present dataset, a new variable is created that takes on the value of the distance to the *closest* threshold in terms of percentage points. As a robustness check, two additional tests are presented using each running variable in its original form. These tests also support the three-percentage points bandwidth. The results of the three tests are visualised in figure A.1.

In table A.1, all standard deviation sizes decrease as the bandwidth increases. This may be the result of the increased number of observations that are analysed in each model. As was mentioned in section 4, the downside of increasing the number of analysed observations (which increases precision) is the bias that is likely to come with it (which reduces accuracy). Model 22 and model 23 show marginally significant coefficients for the northwest quadrant. These effects may be the result of bias. Mann-Whitney U (rank sum) tests and Fisher-Exact tests show that for bandwidths of four and five percentage points, there are significant differences in the total number of pupils and the distribution over school denominations.[12] For a bandwidth of three percentage points, there are no significant differences in terms of total number of pupils

[12]For the bandwidth of four percentage points, the mean number of pupils at the school level is higher in the northwest and southeast quadrant relative to the control group (southwest quadrant) at respectively the five-percent and the one-percent level. For the bandwidth of five percentage points, the distribution over the school denominations shows significant differences between both the northwest and northeast quadrants relative to the control group (southwest quadrant), both at the five-percent level.

TABLE A.1: STANDARDISED PUPIL PERFORMANCE: USING DATA FROM DIFFERENT BANDWIDTHS

	(19)	(20)	(21)	(22)	(23)
Northwest (treatment quadrant)	0.0138	−0.0637	0.110	0.187*	0.169*
	(0.281)	(0.170)	(0.128)	(0.106)	(0.0928)
Northeast (treatment quadrant)	0.558	−0.0126	−0.00437	−0.0693	−0.0142
	(0.403)	(0.211)	(0.151)	(0.124)	(0.104)
Southeast (treatment quadrant)	0.227	−0.109	−0.0785	−0.0437	−0.0191
	(0.329)	(0.136)	(0.102)	(0.0763)	(0.0660)
Additional controls	Yes	Yes	Yes	Yes	Yes
Bandwidth	±1	±2	±3	±4	±5
Observations	336	1153	2208	3794	5583

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. The dependent variable of all models in this table is standardised pupil performance at the school level. Bandwidth is scaled in percentage points. ±1 means: all observations that lie within a one-percentage point range around both eligibility thresholds are included in the model. All models include additional controls; these are linear and squared terms of both running variables (i.e. percentage low income and percentage welfare dependency), a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects.

FIGURE A.1: RAW STATA OUTPUT: TEST FOR BANDWIDTH

```
Bandwidth estimators for RD local polynomial regression


           Cutoff c = 0 | Left of c  Right of c            Number of obs  =      14853
----------------------+----------------------            NN matches     =          3
          Number of obs |     11270        3583            Kernel type    = Triangular
    Order loc. poly. (p) |         1           1
        Order bias (q) |         2           2
Range of DistanceToClosestThreshold |    11.200    26.900

---------------------------------------------
      Method |       h           b          rho
-------------+-------------------------------
         CCT | 3.005017     5.78161    .5197544
---------------------------------------------


Bandwidth estimators for RD local polynomial regression


        Cutoff c = 11.5 | Left of c  Right of c            Number of obs  =      14853
----------------------+----------------------            NN matches     =          3
          Number of obs |     12141        2712            Kernel type    = Triangular
    Order loc. poly. (p) |         1           1
        Order bias (q) |         2           2
Range of PercentageLowIncome |    11.400    19.500

---------------------------------------------
      Method |       h           b          rho
-------------+-------------------------------
         CCT |  2.97474    4.665484     .637606
---------------------------------------------


Bandwidth estimators for RD local polynomial regression


        Cutoff c = 11.3 | Left of c  Right of c            Number of obs  =      14853
----------------------+----------------------            NN matches     =          3
          Number of obs |     11652        3201            Kernel type    = Triangular
    Order loc. poly. (p) |         1           1
        Order bias (q) |         2           2
Range of PercentageWelfareDependency |    11.200    26.900

---------------------------------------------
      Method |       h           b          rho
-------------+-------------------------------
         CCT | 3.210919     5.66478    .5668214
---------------------------------------------
```

Note: All three tests use data on the four-digit postal code level and pupil performance data. The first test uses a variable that takes on the value of the distance to the *closest* threshold. The two tests below use the two running variables (respectively the percentage of low income households and the percentage of welfare dependent households in the four-digit postal code area). The variables DistanceToClosestThreshold, PercentageLowIncome and PercentageWelfareDependency, which represent percentages, are scaled from 0 to 100.

and distribution over school denominations, even at the ten-percent level. In other words, the necessary assumptions for validating the use of the regression discontinuity design do not hold for the bandwidths of four or five percentage points.

A.2 Polynomials running variables

Table A.2 presents the impact of the impulse subsidy scheme on standardised pupil performance at the school level. Models 24 to 28 show the impact of changing the degree of polynomials of the running variables (i.e. percentage low income and percentage welfare dependency) on the effect of interest. As was mentioned in section 4 of this paper, following the recommendation of Gelman and Imbens (2017) and in line with previous research (Verspaandonk, 2016), only linear and squared polynomials are used in the main regressions of this paper. Model 26, which presents the results with second order polynomials, is an exact copy of model 16 in table 5.4.

Models 24 and 25 in table A.2 show that including at least a linear polynomial of the running variables substantially changes the effect sizes of the effect of interest for the northeast and southeast quadrant. This stresses the importance of controlling for the running variables. Further increasing the degree of polynomials in models 26 to 28 does not substantially change pattern of results. Thus, given that the running variables are controlled for, the effect of interest is robust to different degrees of polynomials.

TABLE A.2: STANDARDISED PUPIL PERFORMANCE: CONTROLLING FOR DIFFERENT NUMBERS OF POLYNOMIALS OF THE RUNNING VARIABLES

	(24)	(25)	(26)	(27)	(28)
Northwest (treatment quadrant)	0.0921	0.114	0.110	0.122	0.129
	(0.0984)	(0.127)	(0.128)	(0.154)	(0.156)
Northeast (treatment quadrant)	−0.284***	−0.0612	−0.00437	0.0524	0.0647
	(0.0714)	(0.145)	(0.151)	(0.190)	(0.196)
Southeast (treatment quadrant)	−0.250***	−0.0757	−0.0785	−0.0370	−0.0375
	(0.0623)	(0.102)	(0.102)	(0.131)	(0.131)
Additional controls	Yes	Yes	Yes	Yes	Yes
Degree of polynomials running variables	0	1	2	3	4
Observations	2208	2208	2208	2208	2208

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. The dependent variable of all models in this table is standardised pupil performance at the school level. All models include additional controls; these are a linear term of the total number of pupils at the school, categorical variables of province and school denomination, and time fixed effects. Both running variables (i.e. percentage low income and percentage welfare dependency) are controlled for with different numbers of polynomials.