Agence Française de Développement

working paper

September 2011 | 117

Student Loans: Liquidity Constraint and Higher Education in South Africa

Marc Gurgand, Paris School of Economics, Crest and J-Pal
Adrien Lorenceau, Paris School of Economics
Thomas Mélonio, Research Department, AFD
(meloniot@afd.fr)

Research Department

Agence Française de Développement
Direction de la Stratégie
Département de la Recherche

5 rue Roland Barthes
75012 Paris - France
www.afd.fr

AGENCE FRANÇAISE DE DÉVELOPPEMENT

PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

Disclaimer

The analyses and conclusions formulated in this document are those of the authors. They do not necessarily reflect the position of Agence Française de Développement or its partner institutions.

Publications Director: Dov ZERAH

Editorial Director: Robert PECCOUD

ISSN: 1958-539X

Copyright: 3rd quarter, 2011

Layout: Marcelle LARNICOL

Acknowledgements

We are very grateful to Herman Steyn and all Eduloan staff for providing invaluable data, many details and much advice regarding the South African higher education system and their work in the field, and to Jean Skene and Dr. Qhobela from the Department of Education (DoE) for providing us with the HEMIS data. We would like to thank Jocelyn Vass, Fabian Arends, Mariette Visser from HSRC for very valuable discussions and advice, and Geneviève Michaud for assistance with the data. We also thank the AFD team in South Africa. We are grateful to Clément de Chaisemartin, Eric Maurin and seminar participants at Pompeu Fabra, Paris School of Economics, European Development Network Workshop, EALE/SOLE (2010) and Econometric Society World Congress (2010) for useful comments. The authors gratefully acknowledge the financial support of Agence Française de Développement (AFD).

Contents

Abstract

Empirical evidence that access to higher education is constrained by credit availability is limited and usually indirect. This paper provides direct evidence by comparing university enrollment rates of potential South African students, depending on whether or not they get a loan to cover their university fees, in a context where such fees are high. We use matched individual data from a credit institution (Eduloan) and from the Department of Education. Using a regression-discontinuity design based on the fact that loans are granted according to a credit score threshold, we can estimate the causal impact of loan obtainment. We find that the credit constraint is substantial, as it reduces the enrollment rate into higher education by more than 20 percentage points in a population of student loan applicants.

Introduction

Although primary education is almost universal in South Africa, and secondary schooling has very wide outreach, higher education has become a severe problem in this emerging country. Enrollment stands at about 15%, a low figure for a country at this level of development. The black and colored population and, generally, the poor are those that have the most limited access to education. Wage returns to university degrees are, at the same time, high. This raises both efficiency and equity considerations that stand high on the political agenda.

Credit constraint seems a natural explanation of this situation. Higher education is costly, both in terms of direct and opportunity cost, and poorer people may be unable to borrow against future income if credit markets are imperfect. Although such imperfection is likely to exist, its magnitude remains debatable in what is a relatively highly financialized country. Moreover, the observed stylized facts can also be explained by other types of deprivation, for example, the poor may lack the necessary academic qualifications or a taste for university studies.

If credit constraint is a major problem, then a relevant policy would be to encourage the provision of student loans. This paper assesses the impact of a private company supported by international donors, Eduloan, which provides short-term loans to pay for university fees. In South Africa, average university fees are equivalent to the average monthly wage, and in many institutions it can be 2 to 5 times that much. Our estimation is based on a population wishing to enroll at a university and applying to Eduloan for a loan. We compare the enrollment rate of individuals who obtain a loan with that of individuals who are refused a loan. Identification of a causal effect is based on the observation that Eduloan uses a credit score threshold to decide whether or not to grant a loan: following the regression discontinuity approach, we can compare otherwise similar individuals with and without a loan, on either side of the discontinuity.

We were able to match application and customer data from Eduloan with individual data on university students from the South African Ministry of Education (Higher Education Management Information System [HEMIS] data). This allows us to observe loan requests, loan allocations and subsequent enrollment and graduation for a large sample of individuals. With this data, we can show that loan access substantially increases the probability of enrolling, by 20 to 25 points, representing a 50% increase. As expected, the effect tends to be even stronger for poorer families, indicating that they are more strongly constrained.

This result can be interpreted as a positive evaluation of Eduloan's impact. But it also brings new and straightforward evidence that liquidity constraint is a significant obstacle to higher education. Given the high level of fees, even a short-term loan can affect enrollment, as many people (in particular among the poorest potential students) obviously have no alternative when they are refused a loan from Eduloan. Our results imply both that the rest of the market exposes individuals to liquidity constraint and that this constraint has a large (and quantified) impact on enrollment decisions.

Beyond the Eduloan case, this paper thus contributes to the literature on borrowing constraint and the demand for education. Proving the existence of such constraint and measuring its magnitude has proved a difficult task, and the literature has followed indirect routes. To emphasize the source of the difficulty, think of demand for education *S* in the standard Beckerian framework as a function of potential wages and interest rate: *S(w(.),r)*. Credit market imperfection implies that individuals face interest rates that are higher than the market rate and decrease with assets or parental income, or that they face a limit to their debt that is also a function of their current resources. Demand for education would then have the form *S(w(.),r(I),d(I))*, where *I* is a measure of family income and *d* is maximum debt. Comparing these two demand functions is hard because *r(I)* is rarely observed, the market interest rate *r* is empirically difficult to determine, and debt, if observed, could well be an optimal, not a constrained, amount of debt.

A first strand of the literature has estimated the causal effect of parental income *I* on education level *S*. Some authors, for instance Acemoglu and Pishke (2001) or Maurin (2002), claim that there is a positive effect. But Cameron and Heckman (2001) and Carneiro and Heckman (2002) argue that such a link reflects rather the impact of cultural traits or very early investment during childhood. At any rate, the reduced-form demand function with credit constraint is indistinguishable from *S(w(.), r, I)*, a demand function with perfect credit markets but a consumption motive. This approach thus exhibits a credit constraint only if education is believed to be a pure investment good. Another approach is based on the discount rate bias, thus labeled after Lang and Ruud's (1986) and Lang's (1993) estimation of idiosyncratic discount rates. Card (2001) basically takes the marginal return to schooling to be an estimate of the value of *r(I)*. He argues that for some instruments for schooling in a wage equation, marginal rates of returns are estimated over a population potentially constrained by liquidity. Because, with such instruments, estimated returns are much higher than OLS returns, this could be evidence that *r(I)* is indeed higher for individuals of modest origins. Cameron and Taber (2004) develop the argument further using a model where only the credit market for human capital is imperfect. In this model, the credit constraint only applies to the direct cost of education. They reconsider Card's argument in this context and estimate a structural model of the form *S(w(.), r, C* × *r(I),I)*, where *C* is the amount of direct cost: this interaction allows to differentiate the effect of *r(I)* from *I*. They find no evidence of a credit constraint.

Attanasio and Kaufmann (2009) have recently taken a different route. They claim that subjective expectations of earnings at different education levels ($\tilde{w}(.)$) are the relevant argument in the demand function $S(\tilde{w}(.),r)$. As such, $\tilde{w}(.)$ can be observed simply by asking people. Education should increase with expected returns. If the demand for education is constrained by some binding level of debt *d(I)*, however, then this relationship no longer holds. They do find that expected returns are correlated with education levels for the richer part of their sample, but not for the poorer, which seems to indicate that the poor are credit-constrained.

A few other papers, such as Keane and Wolpin (2001), Brown *et al.* (2009) and Lochner and Monge-Naranjo (2011), rely on structural or calibrated models. Generally, there is little agreement over the existence and importance of credit constraints. The literature is inconclusive and strongly focused on the developed world. Moreover, the empirical methods used are extremely *indirect*, in order to circumvent a basic observability problem. In contrast, this paper takes a very *direct* approach, using a quasi-experiment on loan provision. This is probably the most straightforward way to document this issue. If the loan reduces (at least part of) a credit constraint, then it should increase higher education enrollment. Conversely, if credit constraint is not binding, the loan may well increase individual welfare, but not enrollment.

To our knowledge, the only two papers to take a similar route are Canton and Blom (2004) and Stinebrickner and Stinebrickner (2008). The latter ask American students the hypothetical question of whether they would like to take out a loan at a fair interest rate. Although this is close to our test as a thought experiment, constrained students are identified on the basis of a subjective question, which may be very different from actually obtaining or not obtaining a loan. Furthermore, they do not estimate the effects on attendance. Canton and Blom use data on actual loan provision in Mexico. However, they cannot measure the impact on enrollment because all of their population is already enrolled. They estimate impacts on academic performance instead, but they are exposed to strong selectivity bias if loan provision is also a determinant of enrollment.

In this paper, we have an actual quasi-experiment on the provision of loans and we observe a sample of potential students, some of whom will end up not going to university. This is a unique setup for obtaining evidence on credit constraints. Because the loans that we observe are short term, we must make the distinction between “liquidity constraints”, for those who would have sufficient income to enroll, in present value, but may lack the savings to pay for the tuition fees up front, and “solvency constraints”, for those who would need to increase their income after their studies to pay back their loan. Although the latter would arguably affect a larger share of the population, we show that even a pure liquidity constraint has important consequences for enrollment.

In the following section, we present the Eduloan scheme in the general South African context and show that although other loans are available to some segments of the population (the poorest and the richest), most of the individuals wishing to enroll at university may have limited access to credit. We stress the fact that the high level of fees in this country makes a loan an important option. We also outline a model of Eduloan customer behavior, in order to clarify the interpretation of the estimated parameter. Section 2 describes the data. Section 3 presents the results and Section 4 shows several robustness checks using different sources of identification and samples. Finally, Section 5 discusses the results, notably their interpretation and external validity. We then conclude.

1. The Eduloan Scheme in the South African Context

Since the end of the apartheid regime in 1994, the higher education system has experienced profound changes. The government faced a challenging trade-off: to improve access for the historically disadvantaged people while ensuring the development of the educational system in keeping with international standards. In pursuit of the second of these objectives, the government has reorganized public institutions into three types: Universities, Universities of Technology and Comprehensive Universities (providing both general and vocational qualifications). Distance learning represents more than one third of total enrollment.

However, while primary education is universal and secondary enrollment stands at over 90%, enrollment in higher education is only around 15%; Black Africans account for 60% of this figure, although they represent 80% of the population. Moreover, the graduation rate is extremely low, between 15% and 20% depending on the qualification level and population group (Department of Education, 2009). In this context, the issue of access to higher education, especially for the benefit of the historically disadvantaged, continues to be high on the South African political agenda.

In contrast, wage returns to higher education seem to be very high: Branson *et al.* (2009) and Keswell and Poswell (2004) argue that marginal returns to education increase with education level and are as high as 50% per year at the tertiary level. Altogether, this set of facts - low attendance and high return - is compatible with some forms of constraint in access to higher education.

An obvious source of constraint could be the "shared cost" principle adopted by the South African higher education system: since private returns to tertiary education are high, "users" are asked to finance it in part. As a result, tuition fees represent about 25% of the higher education budget. In 2004 (the beginning of our sample period) they amounted to ZAR 5,251 million (Stumpf *et al.*, 2008), for 744,000 students. The yearly average fee is thus about ZAR 7,000,[1] with in fact substantial variation between institutions: it is not unusual for fees to be between ZAR 15,000 and ZAR 35,000, especially in contact education (as opposed to distance education).[2] These fees are to be compared to the average monthly wage, which is around ZAR 7,500 in this period (Statistics South Africa, 2006) or to the annual GDP per capita at ZAR 36,000 in 2006.[3] In the presence of liquidity constraints, such fees could well explain low enrollment and low graduation in spite of high returns.

In order to empower the historically disadvantaged people and increase participation in higher education for the poorest, the government has implemented a contingent loan program (National Student Financial Aid Scheme [NSFAS]). The loans are granted on the basis of a means test. They are only to be paid back when the student is employed, and the installments depend on his/her salary; moreover, 40% of the loan can be converted into a grant depending on the student's academic results. In 2004, the amounts lent ranged between ZAR 2,000 and ZAR 25,000, and the program benefited 15% of the students in public institutions (Stumpf *et al.*, 2008), 98% of whom were historically disadvantaged.

[1] Accounting for inflation, this is about 1,200 current US dollars.

[2] Social Surveys, 2009.

[3] Relative to GDP per capita, the ZAR 7,000 fee would be comparable to an average fee of US$ 9,500 in the US.

In the South African financial context, the NSFAS is the main opportunity for poor students to finance their education. Commercial banks constitute an alternative source of financing as they also offer student loans (Social Surveys, 2009). However, the requirements for loan approval are such that probably only the wealthiest families will use this option. Informal money lenders also exist, but they charge very high interest rates (40 to 50% on an annual basis). In the light of this financial environment, Eduloan holds a very specific market position.

1.1 Eduloan

Eduloan is a private financial company created in the mid-1990s that receives support from international donors, essentially in the form of guarantees for loans taken out by Eduloan from national banks. It provides loans to cover tuition fees for individuals planning to enroll at a university (public or private) in South Africa. The position of Eduloan in the student loan market seems to be between the NSFAS and the commercial banks. It targets middle to upper-middle income households, most of whom would not be eligible for the NSFAS but may not be wealthy enough to get funding from commercial banks. This is consequently a population likely to face borrowing constraints.

Eduloan provides short- to medium-term loans (typically 12 to 24 months) at a moderate rate (around 1% above the prime rate, which is the reference rate for households). In order to be eligible, borrowers must be employed and have a minimum level of income. In addition, the installment must not exceed 25% of the monthly salary. Customers can borrow to finance their own studies or to sponsor the studies of a relative.

Whether the loan is granted or not also depends on a credit score, called the *Empirica score*. It is calculated by a credit bureau based on a nationwide banking history. Although the algorithm is not made public, we know that it includes information such as current debts, the number of credit cards possessed, delinquency and numerous other variables. The final decision to grant a loan to an applicant is largely dependent on the applicant's Empirica score being above a certain threshold, which is not publicly available and cannot be revealed here for confidentiality reasons (the threshold will be normalized to zero). The Empirica will thus be our forcing variable for the regression discontinuity identification strategy. Individuals are unaware of their score and it is very unlikely that they could manipulate its value in the neighborhood of the threshold (which they do not even know).

Loan applications work as follows: Eduloan has an office on most public university campuses. Students must first decide on the university they wish to attend and the course they wish to study. Once the university has accepted their application and given them the corresponding fee quotation, they can apply directly to Eduloan to cover part or all of the fees. If the loan request is accepted, Eduloan pays the tuition fees directly to the university. If necessary, the student can ask for additional loans during the year. The important feature for us is that choice of a university is a prerequisite for loan application and loans are necessarily provided for that university because of the direct payment system. This will allow us to restrict most of our analysis to students who requested a loan to attend a public university: they cannot use the loan they receive to pay for a different university or for consumption.

1.2 The parameter of interest

In such a context, it is important to clearly describe the parameter that we can estimate. The fact that loans are short-term and must be repaid during the studies, and that potential students are required to have an explicit education project before having access to a loan, are important and specific features. The following simple model clarifies the extent to which the impact of this specific program can be taken to reveal liquidity constraints. Let us describe the intertemporal utility of an agent who borrows from Eduloan. We assume the agent has access to resources I every period while studying. This income is a requirement. It can be the income of his/her parents, relatives or spouse or his/her own income if he/she continues to work while studying. If he/she wants to attend university, he/she will have to pay a fee f. As mentioned above, for many people, f may not be negligible with respect to I. When we consider 3 periods, in period 1, the agent decides whether or not to enroll. If he/she does enroll, he/she has I income and has to pay f. To pay the fees f, he/she can borrow d from Eduloan or use any proportion of I. In the second period, he/she has a new income I and must repay the loan, if any. In the third period, he/she receives his/her new wage as a more highly-educated worker w^H. To make very clear that the loan is short-term, we assume that repayment must occur before the agent actually receives his/her wage w^H. The agent has to solve the following program:

$$\max \quad u(c_1) + \delta u(c_2) + \delta^2 u(c_3)$$

$$\begin{cases} c_1 = I + d - f \\ c_2 = I - rd \\ c_3 = w^H \\ d \leq \bar{d} \\ d \geq 0 \end{cases}$$

where δ is a subjective discount factor, r the interest rate on the loan, d the maximum amount that can be borrowed from Eduloan, and c_i the consumption in period i = 1, 2, 3. In general, the reduced form utility from this program will be some function $V^H(w^H, I, r, f, \max(d))$, with $\max(d) = f$, because Eduloan offers a loan that can cover no more than the university fees f. If the agent decided not to enroll in higher education, he/she would earn w^L every period and his/her intertemporal utility would be of the form $V^L(w^L)$. Eduloan customers, whom we observe, thus have several characteristics: they wish to enroll if they obtain a loan from Eduloan: $V^H(w^H, I, r, f, \max(d') = f) > V^L(w^L)$; also, their utilities and incomes are such that they wish to take out a loan $d > 0$. Whatever parameter we estimate is only valid for that specific population. Evaluating the impact of Eduloan as a scheme amounts to comparing enrollment outcomes when access to the scheme is available and when it is not. If, in the absence of Eduloan, the same person had access to a commercial bank instead and could borrow a maximum amount max(d'), then his/her intertemporal utility, if enrolled, would be $V^H(w^H, I, r, f, \max(d'))$.[4] The proportion of people who asked for a loan from Eduloan and, everything else being equal, enroll when they are granted the loan and do not enroll otherwise is thus:

$$P[V^H(w^H, I, r, f, \bar{d} = f) > V^L(w^L)] \\ - P[V^H(w^H, I, r, f, \bar{d}') < V^L(w^L) < V^H(w^H, I, r, f, \bar{d} = f)]$$

and this is the parameter we can estimate when we observe the higher education enrollment of similar people who, for arguably exogenous reasons, are or are not granted the Eduloan loan they applied for.

If positive, this parameter contains two pieces of information. The first is that, for a set of individuals, max(d') < f: this implies that, but for Eduloan, there is a borrowing constraint in the South African financial market such that these individuals cannot borrow at least the full amount of fees.[5] In our setup, liquidity constraint is obviously evaluated with respect to Eduloan. Eduloan is by no means financial market perfection: it offers a low interest rate, but only over the short term and for limited amounts. But this estimation can reveal that for some individuals, borrowing capacity is even more limited.

4 The argument also runs if we define liquidity constraint as having access to a higher interest rate $r' > r$.

5 It is easy to show that VH is non-decreasing with max(d).

The second piece of information is that the constraint max(d') < f is ***binding*** for that proportion of people. It is possible for liquidity constraints to exist, but without preventing people from entering higher education, because they are prepared to sharply reduce their current consumption, for instance. When the objective is to increase enrollment in higher education rather than the welfare of the students, unbinding liquidity constraint is of limited importance. We only identify the extent of binding liquidity constraint, but this is the most important one from a policy-making perspective.

This paper uses discontinuity in the Empirica score, which serves to decide whether or not to give access to an Eduloan loan, as a means to compare similar people with and without a loan. If the market were highly competitive and many banks were willing to lend to the same people under the same conditions, Eduloan would bring no value added at the margin. On the contrary, if there is an impact, this implies that Eduloan reduces the level of liquidity constraint, and liquidity constraint is indeed a reason why some people may not attend university. From the evaluation of this scheme, we learn something more general for which, as already mentioned, there is no such direct evidence in the literature.

However, we quantify the consequences of such constraints on a specific population: people who want to enter higher education provided they get a loan. To this extent, we probably underestimate the impact of liquidity constraint: some people would need a long-term loan or a loan covering more than just tuition fees in order to enter higher education. We do not have that population in the data. On the other hand, we do not expect credit constraint to be a relevant issue for people who, because of taste or ability, do not consider attending university. Sorting out this latter population from individuals who are liquidity constrained is a problem in the literature. We directly exclude this population here.

2. Data

The data used in this paper have two sources. The first is customer data from Eduloan, describing loan applications and acceptance or rejection decisions. This is necessary in order to compute a "treatment" variable over a population of interest. The second is provided by the Ministry of Education and identifies the students entering a public higher education institution, and thus the outcome variable. These two data sets are matched using national identification numbers.

2.1 Eduloan data

As a private credit company, Eduloan maintains customer files on both the whole set of applicants and on their actual customers. They have provided us with two data sets. The first one contains information on Eduloan applicants between 2004 and 2008. The key variables are the Empirica score, the national identification number of the student (who is not necessarily the applicant when parents borrow for their children) and the application date. In addition, the files include characteristics on the applicant such as the borrower's net salary, the institution he/she applies to, the loan amount requested, his/her age and so on. The second data set contains actual customers, *i.e.* the applicants whose loan application was accepted and who received a loan. Again, the key variables for our purpose are the national identification number and the agreement date.

In the first data set, we can observe several application dates per applicant and per year. These may be either duplicate administrative records for the same request or individuals who actually apply for more than one loan over the year. When a loan has been granted, we have no direct information about which application it corresponds to. Because our outcome (university enrollment) is a yearly event, it is enough for us to know whether, for a given year, certain applications were sent and certain loans were obtained.

In most of the empirical analysis, we use data from 2004 to 2007, because this is the period over which the threshold Empirica value set by Eduloan's internal procedures generates a discontinuity on loan grants. During this period, the threshold remained unchanged. In 2008, Eduloan's activities were strongly impacted by the credit crunch following the financial crisis, and the threshold had much less explanatory power. We use the 2008 data only for a robustness analysis.

2.2 HEMIS data

The second source of data is provided by the Ministry of Education, which bases its management of public subsidies to higher education institutions on enrollment figures. The HEMIS has therefore been created to collect accurate individual data on every student entering the public higher education system. The data contain information on all the courses and qualifications undertaken by a student throughout his/her studies in the public institutions. This includes the name of the institution, the type of courses or qualifications, educational credits completed among those taken, whether the student is in contact or distance mode, etc.

As this data contains the student's national identification number, it can be matched with the Eduloan applicant and customer data. Our database is unique, starting with a list of more than 15,000 applications for a loan at Eduloan, complemented with systematic information on whether applicants obtained a loan from Eduloan and whether they enrolled and completed their credits in a public higher education institution during the relevant year.

2.3 Data constraints

The major limitation of this data stems from the fact that HEMIS files only contain information on students entering *public* higher education institutions. Therefore, we do not know whether individuals who applied to private higher education institutions eventually enrolled. In South Africa, the private higher education sector is quite developed with around seventy notable institutions.

Fortunately, loans are granted in order to pay fees to a specific institution and they are paid directly to that institution by Eduloan. When a loan has been requested for a public institution (which we will refer to hereafter as the "HEMIS perimeter"), then we know whether granting the loan has indeed increased the likelihood that the applicant actually enrolled. Our data contains a variable for the type of institution for which the student has requested a loan. However, this variable is not properly informed for about 18% of observations. Where the information is available, a large majority of students (80%) applied to public institutions, compared to 20% for private ones.

Our baseline analysis will be restricted to applicants within the HEMIS perimeter, excluding loan requests for private or unknown institutions. We will check that this sample selection is independent from having an Empirica score on either side of the threshold. Because this is verified, the sample restriction has no implication for internal consistency; but it does affect external validity. In our robustness analysis, we will include the sample with unknown institutions and show that we can then estimate a lower bound on the effect on HEMIS perimeter individuals. But we will never make any claim about the population that wishes to enter private institutions.

The other technical difficulty is to match dates between applications, loans granted and enrollment. The academic year is the calendar year in South Africa. The norm is that students register in January and ask for a loan right away: 55% of our application dates are in January or February, and 62% in the first three months. But some administrative processing may take time and some students may ask for help to pay additional fees or a second fee installment later on, so that additional applications appear throughout the year. We keep only one observation per student and per year. We consider that loans requested year t have been granted whenever the same student has made one or more applications during year t and has received a loan during the same year.

There is an ambiguity, however, when loan applications are made late in the year and a loan is granted at the beginning of the following year. We do not know whether it is intended to pay for late fees or if it is in provision for the coming year. We are thus unsure whether this request has been accepted and whether we should relate it to enrollment in the current year or the following one. As a result, our baseline estimation excludes individuals for which the only application of the year was posted in November or December (we then keep 86% of our sample). As a robustness check, we check that results are not sensitive to inclusion of these observations. Of course, it is still possible to allocate an application to the wrong academic year, despite excluding the late applications.

Finally, it is worth mentioning that we had to drop some observations for which the national identification number was missing or obviously incorrect. Also, some individuals with no credit history did not have an Empirica score: they are excluded from the whole analysis.

2.4 Descriptive Statistics

Table 1 presents our sample for the years 2004 to 2007, on which most of the analysis is based. Each observation corresponds to a loan request for a given student and a given year. As explained above, when the earliest application was made in November or December, the loan/student/year observation is not included in the baseline sample.

The table shows the characteristics of the student, of the loan request and of enrollment in a public university, if any. The figures are presented separately for individuals who requested a student loan for a public university (HEMIS perimeter), for a private institution, and for whom this information was missing in the data. We also split the sample between loan applications that were accepted and those that were turned down.

It is important to note that the average student age is high, typically around 27. This is mostly explained by the fact that a large share of the students are the borrowers themselves who, by Eduloan rules, have to be employed with a regular income and a pay slip. A substantial share of the sample population are employees who want to upgrade their qualifications in order to gain access to better paid jobs, and not just parents borrowing for their children's education. This is common practice in South Africa, where the largest university in the country (the University of South Africa [UNISA]) is dedicated to distance education.

Borrowers declare wages that are relatively high by South African standards: their average monthly wage is between ZAR 6,000 and ZAR 7,500. This is to be compared with the average wage of the population in formal employment, which was around ZAR 7,500 in this period (Statistics South Africa, 2006). Given that wages are usually log-normally distributed and taking into account the existence of informal employment, it is very likely that our population of borrowers is somewhat above the median wage. Therefore, our sample can be regarded as a collection of potential students from middle-class South African households, although probably not the most well-off. This is precisely the population that we expect to pursue higher education (having graduated from high school and been accepted academically by a university), but who may face a liquidity constraint in doing so. As a matter of fact, requested loan values represent on average one to two month's wages, an amount that households may find difficult to make available up front, but are capable of repaying over 12 to 24 months. This is also a reminder that our sample is obviously not representative of the South African population as a whole, but may correspond to those for whom liquidity is a binding constraint.

Overall, Eduloan accepts 42% of applications. Loans are granted more often to borrowers that declare higher wages (by about ZAR 1,000 in all samples). However, the gender distribution, the proportion of students who are themselves the borrowers and their ages do not differ significantly according to loan status.

When we consider loans requested for a public university, 75% of students who were granted a loan actually enrolled, according to the HEMIS database, compared to only 53% of those who were refused a loan from Eduloan. As a result, a naive estimation of loan impact would be an additional 22 points, or 41% increase in the enrollment rate. The fact that a quarter of the students who had their loan application accepted did not subsequently enroll has no single explanation. One obvious possibility is that they changed their minds, faced unexpected constraints, did not obtain complementary resources, etc. A very likely explanation is that they dropped out early in the year: HEMIS data do not include early dropouts, and as we have already mentioned, dropout rates are huge in South Africa. If students drop out in spite of the loan, this will logically reduce the estimated loan impact. Finally, we cannot exclude mistyped ID numbers or other sorts of mismatches, such that some enrolled persons are treated as non-enrolled or vice versa. However, given that enrollment is an explained variable and that we will use an instrument that must be independent from such measurement errors in the outcome, this should only come at the cost of statistical precision.

Among students actually enrolled in a public university, loan status is only associated with a small difference in the number of courses they register for and in the number of credits they obtain by the end of the year.

When we consider loans requested to attend private higher education institutions, we find that a small fraction actually end up in public universities, according to the HEMIS database. This is the case for 18% with a loan and 11% without a loan. Here again, it is not unlikely that some people changed their plans, but this does not seem to be in response to a loan refusal: in fact this 7 point difference does not survive a causal estimation (see Figure 6, later). Also, looking at courses and credits, which are conditional on studying in a public university, these students do not appear different from the rest of the enrolled population.

Table 1: Descriptive statistics on loan demands, 2004-2007

	Loan requested for a public institution (Hemis perimeter)			
	No loan obtained		Loan obtained	
	mean	standard error	mean	standard error
Gender distribution (figures represent males)	0.46	0.50	0.48	0.50
Age	27.83	8.42	27.57	7.86
Monthly wage	6,420	5,018	7,525	7,372
Missing wage information	0.15	0.36	0.00	0.00
Requested loan/monthly wage	1.53	1.65	1.04	0.80
Missing requested loan value	0.03	0.18	0.00	0.06
Student is the borrower	0.49	0.50	0.49	0.50
Enrollment in public University	0.53	0.50	0.75	0.43
Credits completed (if enrolled)	0.46	0.37	0.44	0.37
Number of courses registered (if enrolled)	7.14	4.37	6.82	4.13
# observations	5,166		4,814	
	Loan requested for a private institution			
	No loan obtained		Loan obtained	
	mean	standard error	mean	standard error
Male	0.46	0.50	0.50	0.50
Age	26.46	9.30	26.18	8.55
Monthly wage	5,918	4,339	6,736	4,929
Missing wage information	0.15	0.35	0.00	0.00
Requested loan/wage	2.18	2.23	1.44	1.26
Missing requested loan value	0.07	0.26	0.06	0.24
Student is the borrower	0.35	0.48	0.36	0.48
Enrollment in public University	0.11	0.31	0.18	0.38
Credits completed (if enrolled)	0.39	0.35	0.44	0.35
Number of courses registered (if enrolled)	6.41	4.56	7.41	4.36
# observations	1,707		766	
	Institution unreported or ambiguous			
	No loan obtained		Loan obtained	
	mean	standard error	mean	standard error
Male	0.48	0.50	0.48	0.50
Age	27.25	8.42	27.27	8.44
Monthly wage	5,941	4,694	6,884	5,012
Missing wage information	0.11	0.32	0.00	0.00
Requested loan/wage	1.58	1.68	1.10	0.86
Missing requested loan value	0.14	0.35	0.33	0.47
Student is the borrower	0.44	0.50	0.43	0.49
Enrollment in public University	0.51	0.50	0.74	0.44
Credits completed (if enrolled)	0.47	0.37	0.47	0.36
Number of courses registered (if enrolled)	7.42	4.33	7.57	4.09
# observations	1,897		766	

Note: The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. When several requests have been sent for a given student the same year, we use the average requested loan.

Source: authors.

3. Results

3.1 Empirical strategy

The main objective of this paper is to estimate the causal impact on enrollment in higher education of obtaining a loan from Eduloan. With no loss of generality, we consider the following model, estimated over a sample of applicants:

$$Y = \alpha + \beta L + \varepsilon$$

where Y is a dummy for enrollment and L a dummy for the loan. α and β are parameters to be estimated and ε is a residual that contains unobserved determinants of enrollment other than the Eduloan loan. Because ε may be correlated with L, simple correlation between enrollment and the loan does not provide a parameter that has a causal interpretation.

In order to identify a causal impact, we use the regression-discontinuity design (see Imbens and Lemieux, 2007, and Lee and Lemieux, 2010, for presentations of this method). We take advantage of the presence of the Empirica score, a credit score E, that strongly influences Eduloan's decision to provide the loan. There is a threshold E_0 that determines eligibility: in principle, Eduloan agents are not to grant a loan if the borrower's value of E is below E_0, although there are exceptions.

Figure 1 shows the probability of obtaining a loan, as a function of the threshold (normalized to zero), for loan requests in the HEMIS perimeter (*i.e.* for a public university) for the years 2004-2007.[6] Each point represents the proportion of applicants that received a loan among individuals with values of E in a small range. In this graph and the following, we restrict the sample to a neighborhood of plus or minus 100 points around E_0 (the total range is about 400 points, but the information is very noisy at large values). To the left of E_0, the probability of obtaining a loan is very small, although not strictly zero. The probability of obtaining a loan increases with the score when the Empirica gets closer to the threshold. There is a very strong discontinuity past the threshold: the probability of obtaining a loan jumps from about 10% to about 50%. It then increases smoothly.

The discontinuous relationship between L and E at E_0 identifies the causal impact of the loan on enrollment if all other determinants (ε in the above statistical model) vary continuously with E, at least in the neighborhood of E_0. Individuals very close to the threshold have very different proportions of loan access but are otherwise extremely similar. As Lee and Lemieux (2010) convincingly argue, this strategy is in essence very similar to randomization, to the extent that individuals happen to have a few more points in E only by mere chance. This is very arguable in the case of the Empirica, because it is based on an unknown algorithm that depends on a number of variables.

This strategy has several limitations. First, identification is local: strictly speaking, it is relevant only for the population close to the threshold. In practice, we will see that, in our data, the population is fairly concentrated around E_0, so that we estimate a parameter that is valid for most of our sample. Second, as shown by Hahn *et al.* (2001), if the

[6] The value of E_0 remained constant over that period.

treatment effect is heterogeneous and if loan access is correlated with loan impact, then the estimated parameter is a local average treatment effect (LATE) in the sense of Imbens and Angrist (1994).[7] In our context, it is not clear that this is a strong limitation, because there is no reason why Eduloan agents should grant the loan on the basis of the likelihood of actually enrolling. The loan is guaranteed by the customer's current income, not their future income dependent on graduation. Therefore correlation between impact and loan access is not particularly to be expected.

To proceed with the estimation, let us first consider the first-step model describing the discontinuous relationship between loan access and the Empirica score:

$$L = g(E) + \delta D + u \qquad (1)$$

where $D = 1$ if ($E \geq E_0$), $g(E)$ is a continuous function of E (at least in the neighborhood of E_0), and δ measures the discontinuity jump. This simply fits the data in Figure 1. We can either estimate it on a large range of values of E and use flexible forms for g, or restrict the sample to the neighborhood of E_0 and estimate local linear regressions that approximate the function as linear. In both cases, specifications allow the function $g(E)$ to be different on the right and left of the discontinuity. Similarly, the structural equation can be written as:

$$Y = f(E) + \beta L + \varepsilon' \qquad (2)$$

Conditional on E, D is a valid instrument for L, so that this model can be estimated by instrumental variable. Here again, $f(E)$ can be allowed to have different shapes on the right and left of the discontinuity, and the model can be estimated on a large range or by local linear (instrumental) regression. In the latter case, bias is minimized when the sample is strongly restricted to the neighborhood of E_0, but precision is increased as the sample gets larger. Imbens and Lemieux (2007) suggest a cross-validation procedure to select the optimal bandwidth in terms of mean squared error. Depending on the specification, we find optimal bandwidths of +/-65 or +/-125 Empirica points around the threshold[8]. These are quite large bandwidths, reflecting the fact that the linear approximation is adequate in a large range. Nonetheless, we always present regressions for the full sample, using linear or quadratic functions for g or f, with different slopes on either side of the discontinuity, and local linear regressions for different bandwidths, including the optimal one.

Table 2 presents the estimation of equation 1: the increase in the proportion treated due to the discontinuity is estimated between 0.32 and 0.42 depending on the specification, which is always very significant. At the optimal bandwidth of +/-65 points, the effect is 0.39 and it is only slightly lower (0.36) using the full sample with quadratic functions. This ensures that the instrument will have identifying power.

We can also check that E_0 is not a threshold for variables other than the loan. Table 3 shows that there is no discontinuous change in the borrowers' wages, the choice of a public or a private institution, or the amount of the loan requested. This confirms that borrowers know neither their Empirica score nor the threshold, meaning they do not ask for larger loans when they know that their chances of being accepted are strong. Finally, Figure 2 plots the density of observations around E_0. First, there is no evidence of bunching to the right of the threshold, which would happen if individuals could manipulate their Empirica at the margin. Second, we can see that observations are concentrated around the threshold, so that, as mentioned earlier, local identification still involves a large fraction of our sample.

[7] This identifies an average of the causal loan effect on the population who would not get a loan when to the left of E_0 and who would get a loan when to the right of E_0.

[8] Optimal bandwidth is mentioned in table notes.

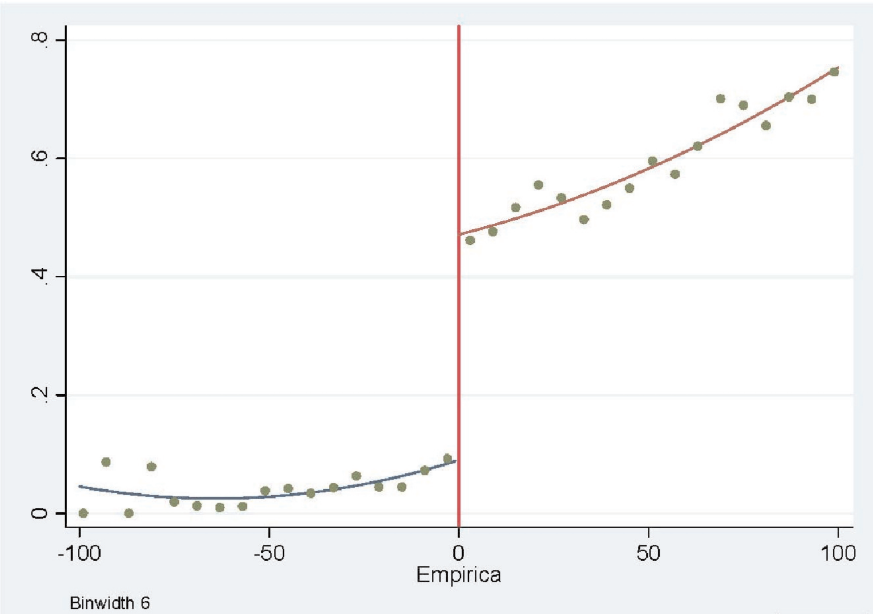
Table 2: Loan granted as a function of Empirica score (HEMIS perimeter, 2004-2007)

	Total sample		Local linear regression for various bandwidth			
			+/-125	+/- 100	+/- 65	+/-20
Above discontinuity point	0.3588095 *0.0176679*	0.4243667 *0.0125766*	0.38820428 *0.01410109*	0.37891166 *0.0152003*	0.38763807 *0.01810028*	0.32841626 *0.03198997*
Empirica	0.0015726 *0.0004512*	0.0006763 *0.000156*	0.00069272 *0.00016774*	0.00076202 *0.0001862*	0.00108194 *0.00025933*	0.00336206 *0.00132778*
Empirica x above	0.0094461 *0.0038847*	0.0010395 *0.0001877*	0.0018697 *0.00024076*	0.00201763 *0.0002957*	0.00100481 *0.00049735*	0.00354861 *0.00264491*
Empirica sq	0.0021081 *0.0005811*					
Empirica sq x above	-0.0207706 *0.0043605*					
Intercept	0.0854386 *0.0109713*	0.0726558 *0.007932*	0.07309332 *0.00815685*	0.07480028 *0.00850378*	0.08159359 *0.00960324*	0.10108989 *0.01716989*
# observations	9,980	9,980	8,533	7,719	6,012	2,340

Sample: loans requested for registration in a public University (""Hemis perimeter""). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Explained variable is a dummy that takes value 1 when a loan has been granted. Ordinary least squares. Robust standard errors in italics. Bandwidth of +/-65 around discontinuity point is the optimal bandwidth according to the cross-validation criteria.

Source: authors.

Figure 1: Share of loans granted as a function of Empirica score (2004-2007)



Source : STATA™.

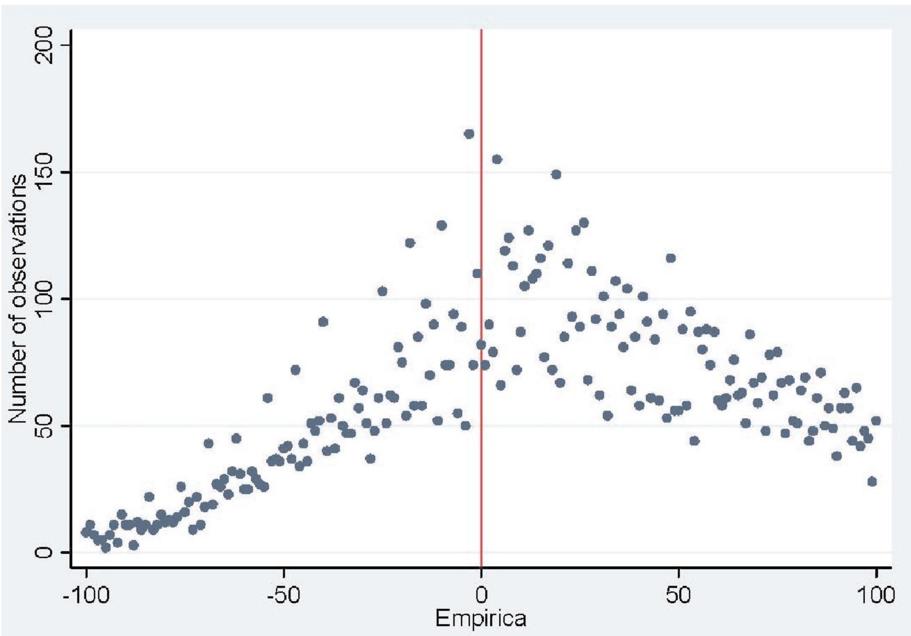
Table 3: Predetermined variables as a function of Empirica score (2004-2007)

	Log wage	Loan requested for public university (as opposed to private)	Log requested loan amount
Above discontinuity point	-0.0099474 *0.025977*	-0.0001706 *0.0191424*	0.0012409 *0.0354248*
Empirica	0.0016394 *0.0009854*	0.0025437 *0.0008259*	-0.0001394 *0.0014261*
Empirica x above	0.0193099 *0.0091447*	0.0164351 *0.0082171*	-0.0025797 *0.0140903*
Empirica sq	-0.0012238 *0.0010682*	-0.0016482 *0.0008673*	0.00000529 *0.001518*
Empirica sq x above	-0.0172864 *0.0094397*	-0.0188625 *0.0083395*	0.00912040 *0.0143757*
Intercept	8.5895280 *0.0215953*	0.7837317 *0.0163057*	8.5542550 *0.0298889*
# observations	13,886	12,453	14,243

Sample: loans requested with non-missing values for the relevant variable. The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Ordinary least squares. Robust standard errors in italics. Intercept of log wage and loan amounts not reported for confidentiality reasons.

Source: authors.

Figure 2: Density of the Empirica score (2004-2007)



Source : STATA™.

3.2 Baseline results: impact of loans on enrollment

Table 4 and Figure 3 show the reduced-form relation between enrollment and the Empirica score. The probability of being enrolled at a public university, for individuals who applied for a loan to study at such a university, increases precisely at the threshold E_0. This would not happen if the loan was not a causal determinant of enrollment, unless there were other determinants of enrollment that also changed discontinuously at E_0, something that we argued could be excluded in this environment. The effect is strong and very significantly estimated at 9 to 10 percentage points. Given that the threshold value is normalized to zero, the enrollment rate just to the left of the discontinuity is directly given by the constant: therefore, this reduced-form effect is to increase enrollment rates from about 50% to about 60%.

Table 5 presents estimates of equation 2. Ordinary least square estimation indicates that obtaining a loan increases enrollment by 20 percentage points. Instrumental variable estimation, using the discontinuity as an instrument, raises this effect to about 22 to 25 points. A stronger effect is found for the small +/-20 bandwidth, but this is the exception and this range is not the optimal one. As a result, we can claim that providing a loan to members of this population causally increases the probability that they will enroll in higher education from a level of 50% to 73%, at least for individuals close to the Empirica threshold. As expected, the results hardly change if we add control variables such as age, gender, required loan amount or monthly wage, because the instrument is not correlated to these variables. Including them does not systematically improve the precision of the estimation, so we present the simple regressions that are more transparent.

OLS estimation appears to be biased (precision is sufficient for a Hausman test to reject equality of the OLS and IV parameters), but the size of the bias is small. This implies that characteristics observed by Eduloan that determine loan acceptance are marginal determinants of the individual decision to enroll in this sample.

We do not find any significant difference when measuring loan impact separately for men and women. Nor does it seem to make a difference whether the borrower is the student himself or a relative. However, as shown in Table 6, the impact of the loan is different among the richest and the poorest borrowers. We do not have much statistical power when it comes to splitting the sample, but we can distinguish between the lowest wage quartile and the rest of the sample (higher panel) or between above and below the median (lower panel). Loan impact is about twice as large for the lowest quartile and about 70% higher when we compare samples across the median. Although the first comparison is only significant at the 10% level, and the second comparison is not significant at all, this is indicative of a plausible idea: that credit constraint is stronger for less wealthy families and that fewer financing alternatives exist at the bottom of the income distribution. One possibility is that commercial banks may be willing to grant loans to some of the richest individuals in our sample, thereby diminishing the impact of Eduloan activities on this specific population.

Two other outcome variables are shown in Table 7. The *number of courses registered for* takes value zero for the non-enrolled and whatever positive values for the enrolled, and similarly for credits completed. Because they enroll more frequently, applicants who get a loan tend to register for more courses on average (1.5 more courses, a 44% increase at the optimal bandwidth specification) and complete more credits than those who are rejected (around 8 percentage points, a 39% increase). In South Africa, one year of higher education represents 1.0 credit, so that a typical academic year is made up of 10 courses, each one worth 0.1 credit: our descriptive statistics recall the low completion rate of students, whether or not they obtain a loan. We cannot identify the impact of having a loan on educational outcomes *conditional on enrollment*.[9] However, we were able to show that increased enrollment resulting from loan access does translate into increased registration and credit completion, which is important from a policy point of view.

[9] If we compare individuals with and without a loan among the enrolled, we mix two effects. One is that the loan induces a different performance of *ex ante* similar people in the two groups, the other is that the loan induces enrollment of additional people, and these people may be different in terms of academic capacity or motivation. This is the usual selectivity problem, as faced by Canton and Blom (2004) for instance. Because we do not have an exogenous determinant of selection that would not have a direct influence on performance, we cannot control for selection without making arbitrary parametric assumptions. Bounds analysis only generates very large bounds here.

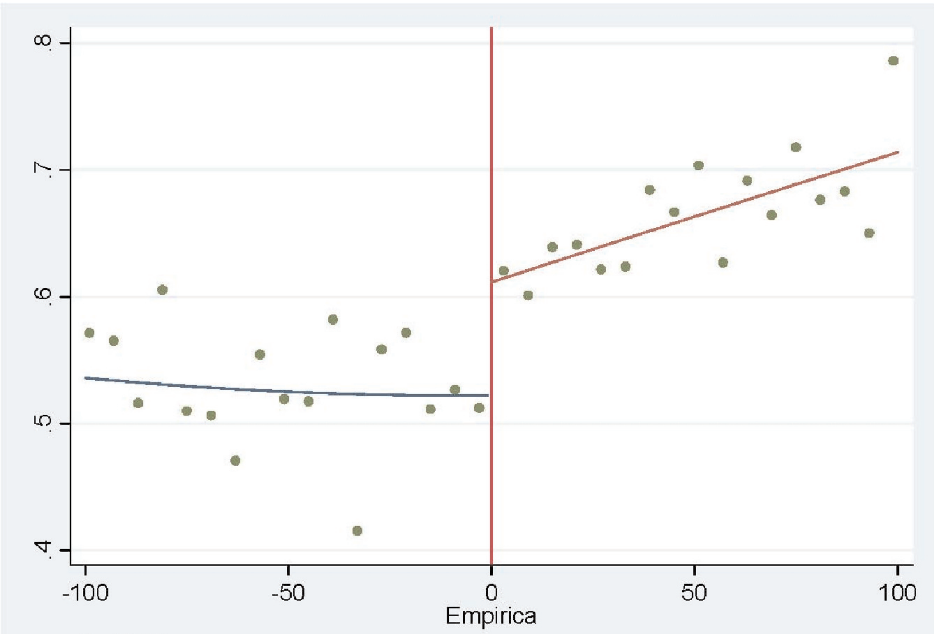
Table 4: University enrollment as a function of Empirica score (HEMIS perimeter, 2004-2007)

	Total sample		Local linear regression for various bandwidth			
			+/-125	+/-100	+/- 65	+/-20
Above discontinuity point	0.0905 *0.0256*	0.1015 *0.0188*	0.0974 *0.0200*	0.0914 *0.0210*	0.0874 *0.0242*	0.1049 *0.0409*
Empirica	-0.0004 *0.0011*	-0.0001 *0.0004*	-0.0002 *0.0004*	-0.0001 *0.0004*	-0.0001 *0.0006*	-0.0014 *0.0027*
Empirica x above	-0.0031 *0.0106*	0.0008 *0.0004*	0.0011 *0.0004*	0.0011 *0.0005*	0.0012 *0.0007*	0.0025 *0.0035*
Empirica sq	0.0017 *0.0011*					
Empirica sq x above	-0.0002 *0.0108*					
Intercept	0.5174 *0.0217*	0.5216 *0.0163*	0.5177 *0.0166*	0.5203 *0.0171*	0.5222 *0.0191*	0.5042 *0.0313*
# observations	9,980	9,980	8,533	7,719	6,012	2,340

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Explained variable is a dummy that takes value 1 when a student is found enrolled at a public University in the same year as the loan request. Ordinary least squares. Robust standard errors in italics. Bandwidth of +/-125 around discontinuity point is the optimal bandwidth according to the cross-validation criteria.

Source: authors.

Figure 3: Proportion of university enrollment as a function of Empirica score (2004-2007)



Source : STATA™.

Table 5: University enrollment as a function of loan obtention (HEMIS perimeter, 2004-2007)

	Total sample		Total sample		IV for various bandwidth			
	OLS	OLS	IV	IV	+/-125	+/-100	+/- 65	+/-20
Loan	0.2035	0.2048	0.2523	0.2393	0.2508	0.2412	0.2254	0.3194
	0.0108	*0.0105*	*0.0705*	*0.0439*	*0.0508*	*0.0547*	*0.0616*	*0.1232*
Empirica	-0.0002	-0.0000	-0.0008	-0.0002	-0.0004	-0.0003	-0.0003	-0.0024
	0.0008	*0.0003*	*0.0012*	*0.0004*	*0.0004*	*0.0005*	*0.0007*	*0.0030*
Empirica x above	-0.0011	0.0004	-0.0055	0.0006	0.0007	0.0006	0.0010	0.0013
	0.0088	*0.0003*	*0.0109*	*0.0004*	*0.0004*	*0.0005*	*0.0007*	*0.0034*
Empirica sq	0.0009		0.0011					
	0.0010		*0.0011*					
Empirica sq x above	-0.0004		0.0051					
	0.0085		*0.0117*					
Intercept	0.5123	0.5174	0.4959	0.5042	0.4994	0.5023	0.5038	0.4719
	0.0120	*0.0093*	*0.0270*	*0.0191*	*0.0197*	*0.0205*	*0.0233*	*0.0418*
# observations	9,980	9,980	9,980	9,980	8,533	7,719	6,012	2,340

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Explained variable is a dummy that takes value 1 when a student is found enrolled at a public University in the same year as the loan request. In IV specification, the excluded instrument is a dummy for "above discontinuity point". Robust standard errors in italics. Bandwidth of +/-65 around discontinuity point is the optimal bandwidth according to the cross-validation criteria.

Source: authors.

Table 6: University enrollment as a function of loan obtention (HEMIS perimeter, 2004-2007)

	Heterogenous effects					
	Borrower wage below first quartile (IV)			Borrower wage above first quartile (IV)		
	Total sample	+/-125	+/- 65	Total sample	+/-125	+/- 65
Loan	0.4679	0.3622	0.4188	0.1731	0.2154	0.1520
	0.1762	*0.1250*	*0.1464*	*0.0751*	*0.0541*	*0.0663*
Empirica	-0.0028	-0.0008	-0.0019	0.0009	-0.0001	0.0009
	0.0024	*0.0009*	*0.0013*	*0.0015*	*0.0005*	*0.0008*
Empirica x above	-0.0221	0.0004	0.0013	0.0078	0.0005	-0.0000
	0.0225	*0.0008*	*0.0015*	*0.0143*	*0.0005*	*0.0009*
Empirica sq	0.0015			-0.0002		
	0.0021			*0.0014*		
Empirica sq x above	0.0274			-0.0093		
	0.0246			*0.0149*		
Intercept	0.4359	0.4636	0.4430	0.5446	0.5249	0.5530
	0.0506	*0.0377*	*0.0439*	*0.0360*	*0.0258*	*0.0309*
# observations	2,304	2,007	1,397	6,909	5,818	4,027
	Borrower wage below median (IV)			Borrower wage above median (IV)		
	Total sample	+/-125	+/- 65	Total sample	+/-125	+/- 65
Loan	0.2754	0.2559	0.2723	0.2050	0.2359	0.1604
	0.1212	*0.0815*	*0.1005*	*0.0837*	*0.0626*	*0.0751*
Empirica	-0.0004	0.0000	-0.0004	-0.0001	-0.0006	0.0007
	0.0018	*0.0006*	*0.0010*	*0.0018*	*0.0006*	*0.0010*
Empirica x above	-0.0051	0.0000	0.0005	0.0021	0.0009	0.0002
	0.0179	*0.0006*	*0.0010*	*0.0165*	*0.0006*	*0.0011*
Empirica sq	0.0003			0.0006		
	0.0016			*0.0017*		
Empirica sq x above	0.0054			-0.0026		
	0.0197			*0.0170*		
Intercept	0.4995	0.5054	0.4977	0.5247	0.5108	0.5460
	0.0416	*0.0296*	*0.0352*	*0.0433*	*0.0315*	*0.0376*
# observations	4,607	3,993	2,794	4,606	3,832	2,630

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Explained variable is a dummy that takes value 1 when a student is found enrolled at a public University in the same year as the loan request. IV estimation: the excluded instrument is a dummy for "above discontinuity point". Robust standard errors in italics.

Source: authors.

Table 7: University outcomes as a function of loan obtention (HEMIS perimeter, 2004-2007)

	Total sample OLS	Total sample IV	IV for various bandwidth +/-125	+/-100	+/- 65	+/-20
Number of courses registered						
Loan	1.0559 *0.1108*	1.85867 *0.6613*	1.51256 *0.4803*	1.7445 *0.5179*	1.4286 *0.5817*	1.6413 *1.1290*
Empirica	0.0043 *0.0072*	-0.0053 *0.0105*	-0.0030 *0.0039*	-0.0034 *0.0043*	-0.0020 *0.0061*	-0.0117 *0.0266*
Empirica x above	0.0407 *0.0791*	-0.0321 *0.0973*	0.0075 *0.0038*	0.0047 *0.0044*	0.0087 *0.0068*	0.0148 *0.0316*
Empirica sq	-0.0016 *0.0091*	0.0030 *0.0098*				
Empirica sq x above	-0.0136 *0.0766*	0.0775 *0.1050*				
Intercept	3.6460 *0.1155*	3.3748 *0.2471*	3.4166 *0.1814*	3.3944 *0.1896*	3.4452 *0.2148*	3.3191 *0.3727*
Credits completed						
Loan	0.0602 *0.0084*	0.1191 *0.0495*	0.0846 *0.0363*	0.1002 *0.0392*	0.0717 *0.0436*	0.1734 *0.0851*
Empirica	0.0004 *0.0005*	-0.0003 *0.0008*	-0.0002 *0.0003*	-0.0002 *0.0003*	0.0002 *0.0004*	-0.0030 *0.0020*
Empirica x above	0.0034 *0.0061*	-0.0019 *0.0074*	0.0006 *0.0003*	0.0004 *0.0003*	0.0003 *0.0005*	0.0023 *0.0025*
Empirica sq	-0.0004 *0.0007*	-0.0000 *0.0007*				
Empirica sq x above	0.0006 *0.0059*	0.0073 *0.0079*				
Intercept	0.2354 *0.0087*	0.2156 *0.0182*	0.2187 *0.0136*	0.2168 *0.0141*	0.2267 *0.0158*	0.1876 *0.0274*
# observations	9,980	9,980	8,533	7,719	6,012	2,340

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. The sample is NOT restricted to individuals enrolled at University. Explained variable are for the same academic year as the loan request. In IV specification, the excluded instrument is a dummy for "above discontinuity point". Robust standard errors in italics. Bandwidth of +/-125 around discontinuity point is the optimal bandwidth according to the cross-validation criteria for both variables.

Source: authors.

4. Robustness

4.1 The 2008 credit crunch

In 2008, the financial crisis induced a restriction in credit that impacted financial institutions, including Eduloan. As a result, fewer loans were granted that year, especially to people above the Empirica threshold, as illustrated in Figure 4 for years 2007 and 2008. We can thus compare individuals on the right of the Empirica score before and after 2008 and use this as different identifying information to check the robustness of our initial results.

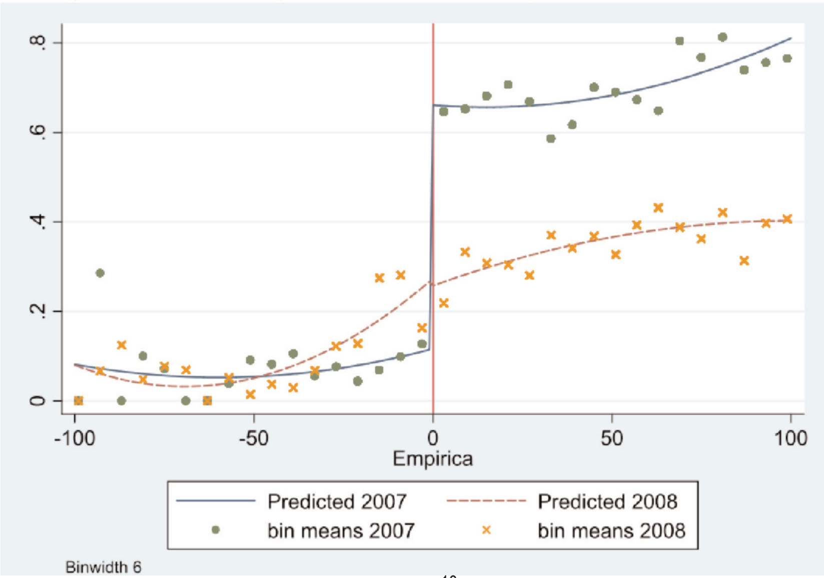
Figure 5 shows the reduced-form relationship between enrollment rates and the Empirica threshold: the discontinuity that is apparent in 2007 disappears in 2008, and this mirrors perfectly the structure of loan access in Figure 4. We can fit this data with a model that interacts functions f(S) in equation 2 with years:

$$Y = f_{2007}(S) + f_{2008}(S) + \theta D + \beta L + \varepsilon'$$

In this regression we can allow D to be present in the regression because L is now instrumented by the interaction between D and year 2007: we thus use a different identification restriction. As a result, this also gives an opportunity to check that being on the right-hand side of the discontinuity has no impact on enrollment when it has no impact on loans: we expect $\theta = 0$.

Table 8 presents this estimation for 2007 and 2008.[10] Although they are based on a different type of information, coefficients on loans are only slightly smaller than in the baseline estimation, but they are very comparable and significant. Also, the coefficient on D is small and non-significant, which confirms our baseline identification hypothesis.

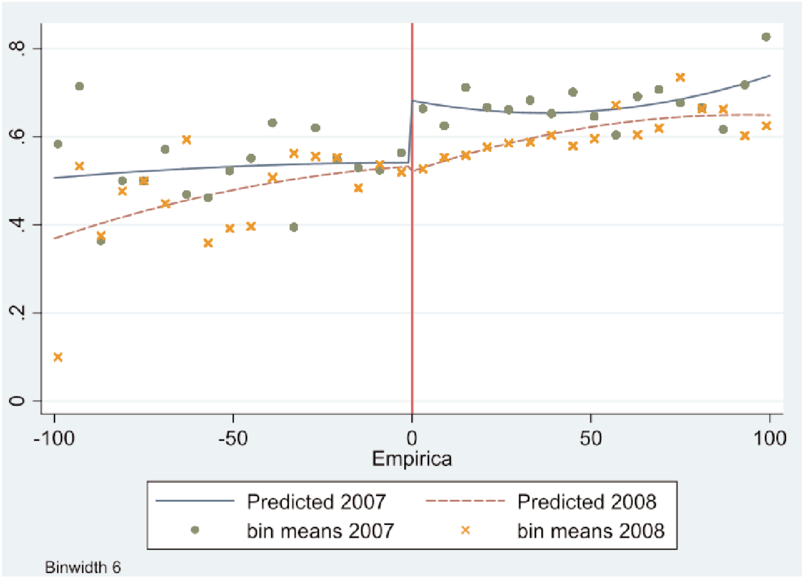
Figure 4: Share of loans granted as a function of Empirica score (2007 and 2008)



Source : STATA™.

[10] In 2008, applications made after April are excluded because the Empirica threshold E_0 was increased.

Figure 5: Proportion of university enrollment as a function of Empirica score (2007 and 2008)



Source : STATA™.

Table 8: Difference in characteristics between enrolled with and without access to loan (HEMIS perimeter, 2004-2007)

	Difference	s.e.	p-value	Population mean
Borrower wage below first quartile	0.25	0.30	0.40	0.25
Borrower wage below median	0.01	0.38	0.97	0.5
Male	0.10	0.27	0.71	0.45
Age	1.19	4.42	0.79	26.53
Student is the client	-0.16	0.50	0.75	0.43
Non-White	0.21	0.16	0.21	0.90
NASFAS beneficiary	0.36	0.23	0.12	0.12
Distance education	-0.12	0.27	0.65	0.34
Student at technical university	0.27	0.25	0.27	0.05

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Bootstrapped standard errors; p-values are for the null that "difference" is zero. See text for computation.

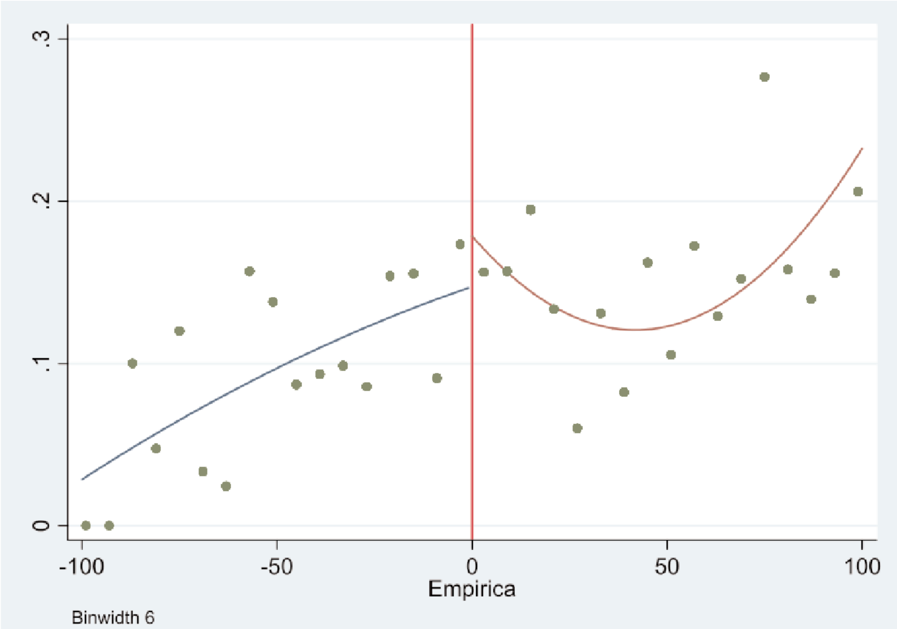
Source: authors.

4.2 Sample variants

As mentioned earlier, the sample used until now has been restricted to loans requested to pay public university fees (HEMIS perimeter), but only when information on the kind of university was actually available. There are 2,664 observations for which either the field was not completed or the abbreviation or acronym used did not refer to an institution we could clearly identify. This sample may contain a number of loans in the HEMIS perimeter, and the corresponding population may be specific. As a robustness check, we would like to include this population. However, this means including an unknown proportion of loans requested for private institutions as well.

The Appendix shows that if we do so, and estimate the regression discontinuity model using both the known public and unknown samples, we obtain a lower bound to the true parameter. This is true provided that loan access has no causal effect on enrollment in a public university for those who wished to enter a private institution. This is expected, given the fact that fee payments are delivered directly by Eduloan.

Figure 6: Proportion of public university enrollment for individuals who requested a loan for a private university, as a function of Empirica score (2004-2007)



Source : STATA™.

It is confirmed by Figure 6: this figure uses the sample of loan requests known to be for a private institution (2,473 observations) and plots the reduced form of enrollment in a public university as a function of the Empirica discontinuity. There is no evidence that loan status has any impact on enrollment in a public university.[11]

In this context, it is intuitive that pooling public and non-public loan demands will provide an average of: (1) the true effect on HEMIS perimeter demands and (2) a zero effect; thus a lower bound to the true effect. As detailed in the Appendix, this argument is complicated by the fact that the two subsamples may have different discontinuity impacts in the first-stage regression, but we show that the lower bound rule remains. Results are presented in Table 9 and they show significant effects, to the order of 0.18, compared to our baseline estimates of about 0.23 (Table 5). We are thus confident of the presence of an impact and its order of magnitude.

A second restriction to our baseline sample has been to exclude observations with loan requests made in November or December, because we are unsure whether they refer to the current year or to the following year. The sample change is rather marginal, as the number of observations is increased by only 12% if we keep late requests. With such data, we expect some enrollment measurements to correspond to the wrong year. According to the same argument as above, the impact has to be zero for a (small and unidentified) share of the sample, because the wrong year outcome variable will not be sensitive to loan access. Including November and December applications, we thus estimate a lower bound. Table 10 shows that coefficients are only slightly lower than our baseline estimates.

To sum up, data limitations imply that, strictly speaking, our baseline estimation may have external validity limitation, even if we restrict our universe to loan requests to Eduloan to attend public universities. When we enlarge the sample, we can only estimate bounds to the parameter of interest. Nevertheless, these bounds do confirm the order of magnitude of the effects and they are not significantly different from our baseline point estimates.

[11] Remember that a small share of individuals who applied for a loan for a private university end up enrolled in a public university. Figure 6 shows that this is unrelated to loan status.

Table 9: Loan impact on credit, conditional on enrollment (OLS)

	Total sample	+/-125	+/- 65
Above discontinuity point	0.0032	-0.0144	-0.0182
	0.0257	*0.0198*	*0.0239*
Empirica	0.0000	-0.0001	0.0005
	0.0011	*0.0004*	*0.0006*
Empirica x above	0.0011	0.0004	-0.0003
	0.0115	*0.0004*	*0.0007*
Empirica sq	-0.0007		
	0.0012		
Empirica sq x above	0.0062		
	0.0117		
Intercept	0.4363	0.43445	0.4455
	0.0221	*0.0166*	*0.0192*
# observations	9 980	8 533	6 012

Sample: loans requested for registration in a public University, restricted to individuals enrolled in a public university. The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations.

Source: authors.

Table 10: University enrollment as a function of loan obtention, difference-in-difference (HEMIS perimeter, 2007-2008)

	Total sample		IV for various bandwidth			
	IV	IV	+/-125	+/- 100	+/- 65	+/-20
Loan	0,1862	0,1898	0,2049	0,2280	0,2473	0,3313
	0.1115	*0.0958*	*0.0981*	*0.1001*	*0.1001*	*0.2191*
Above discontinuity point	0,0127	-0,0050	-0,0126	-0,0146	-0,0127	-0,0565
	0.0407	*0.0378*	*0.0369*	*0.0372*	*0.0389*	*0.0772*
Year 2007	*0,0349*	*0,0176*	*0,0144*	*0,0271*	*0,0329*	*0,0150*
	0.0356	*0.0276*	*0.0282*	*0.0288*	*0.0324*	*0.0625*
Empirica x 2007	*-0,0003*	*0,0001*	*0,0000*	*0,0001*	*0,0000*	*-0,0007*
	0.0020	*0.0007*	*0.0007*	*0.0008*	*0.0011*	*0.0048*
Empirica x above x 2007	0,0005	0,0004	0,0005	0,0001	-0,0002	0,0012
	0.0020	*0.0007*	*0.0007*	*0.0008*	*0.0013*	*0.0057*
Empirica sq x 2007	-0,0047					
	0.0195					
Empirica sq x above x 2007	0,0069					
	0.0201					
Empirica x 2008	-0,0005	0,0011	0,0000	0,0001	0,0000	-0,0007
	0.0016	*0.0005*	*0.0007*	*0.0008*	*0.0011*	*0.0048*
Empirica x above x 2008	0,0014	-0,0005	0,0005	0,0001	-0,0002	0,0012
	0.0018	*0.0006*	*0.0007*	*0.0008*	*0.0013*	*0.0057*
Empirica sq x 2008	-0,0183					
	0.0170					
Empirica sq x above x 2008	0,0169					
	0.0168					
Intercept	0,4852	0,5080	0,5073	0,4947	0,4856	0,4777
	0.0297	*0.0210*	*0.0218*	*0.0230*	*0.0268*	*0.0448*
# observations	7,145	7,145	6,185	5,597	4,383	1,748

Sample: loans requested for registration in a public University ("Hemis perimeter"). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. Explained variable is a dummy that takes value 1 when a student is found enrolled at a public University in the same year as the loan in request. In IV specification, the excluded instrument is a dummy for "above discontinuity point x year 2007". Robust standard errors in italics.

Source: authors.

5. Discussion

5.1 Enrollment in the private sector

We have shown that when an individual plans to enter a public university and requests a short-term loan from Eduloan to pay the fees, he or she is more likely to enroll in a public university when the loan is granted. We cannot strictly exclude that an individual whose demand is turned down will decide to enroll in the private sector instead, because our data contains no information on private enrollment.[12] To the extent that our main question concerns the existence of a liquidity constraint and the estimation of how many individuals are constrained in a population, our conclusion is robust: a large number of individuals who had an explicit plan to enter some kind of university had to change this plan one way or another because they did not obtain short-term credit to pay the fees of that university.

It is more debatable whether this liquidity constraint results in an equivalent decrease in the number of individuals that actually enter higher education. If private institutions are less expensive than public universities,[13] it could be rational for some individuals to turn to a private institution when they are refused a loan by Eduloan, provided the cost is sufficiently low to avoid the liquidity constraint, and the quality is sufficiently high to make this choice a second best. If such behavior (unobserved by us) was present, this would reduce the loan impact *in terms of overall enrollment in higher education*. We do know from Figure 6 that the reverse does not hold true: individuals who apply to a private institution, and are not granted a loan for that, do not turn more often to a public university. But this could just be because this is a more expensive option.

However, we have a way to check whether individuals refused a loan by Eduloan tend to choose a less costly university instead. South Africa has a famous distance learning institution, which was open to black and colored people under apartheid: the University of South Africa (UNISA). In our data, 31% of all loan demands for a public university (HEMIS perimeter) are made for UNISA. Its lower cost is reflected in the size of the loans requested: the average loan request is ZAR 8,051 for other public universities but only ZAR 3,885 for UNISA. Table 11 examines individuals that requested a loan for a public university *other than* UNISA. It checks whether those who were refused a loan eventually enroll at UNISA. To do so, we simply use the same regression discontinuity design as before to estimate the causal effect of a loan on this new outcome ("being registered at UNISA"). We find no evidence of such behavior.

If switching to a less costly institution were optimal for many individuals when a loan for a public university is refused, then we would expect at least some of them to switch to UNISA and others to enter a private university. As we find no evidence of the former (in spite of the fact that UNISA is a well-known and popular institution), we do not expect the latter to be a major source of bias on the enrollment impact of loans.

[12] As a matter of fact, there are a few individuals who have filed loan requests for both public and private institutions. When this is the case, the year-loan request observation has been classified as private, in order to remain on the safe side.

[13] Anecdotal evidence tends to indicate this is the case, although there is substantial heterogeneity.

Table 11: University enrollment as a function of loan obtention (HEMIS + Unknown perimeter, 2004-2007)

	Total sample	IV for various bandwidth				
	IV	IV	+/-125	+/- 100	+/- 65	+/-20
Above discontinuity point	0.1791 *0.0729*	0.1877 *0.0432*	0.1934 *0.0512*	0.1832 *0.0557*	0.1675 *0.0640*	0.1780 *0.1340*
Empirica	0.0004 *0.0010*	0,0002 *0.0003*	0.0001 *0.0004*	0.0002 *0.0004*	0.0003 *0.0006*	0.0002 *0.0027*
Empirica x above	0.0026 *0.0088*	0.0003 *0.0003*	0.0003 *0.0003*	0.0004 *0.0004*	0.0004 *0.0006*	0.0003 *0.0030*
Empirica sq	0.0000 *0.0009*					
Empirica sq x above	-0.0025 *0.0097*					
Intercept	0.5245 *0.0246*	0.5201 *0.0168*	0.5180 *0.0176*	0.5202 *0.0185*	0.5249 *0.0213*	0.5225 *0.0407*
# observations	12,643	12,643	10,946	9,970	7,869	3,118

Sample: loans requested for registration in a public University ("Hemis perimeter") or under unknown status (either public or private University). The unit of observation is loan demand per year. Loans requested in November/December are excluded as in all baseline estimations. In the second panel, explained variable is a dummy that takes value 1 when a loan has been granted. In IV specification, the excluded instrument is a dummy for "above discontinuity point". Robust standard errors in italics.

Source: authors.

5.2 External validity

We have already mentioned the limitation on external validity due to the specific population of Eduloan applicants. On the one hand, we do not include people for whom higher education is not an option anyway, and on the other, we probably do not observe the poorest among the rest of them.

Eduloan provides its customers with short-term loans with a limited grace period (most of the time 2 or 3 months). This implies that Eduloan only alleviates short-term constraints, which only constitute part of financial credit constraints. Deeper solvency problems, which can only be solved through an increase in future income and longer grace periods, are not identified in our study, since students who cannot repay at least part of their loan while studying are not granted a loan. The fact that even a simple smoothing payment mechanism has a very significant impact on university enrollment suggests that the credit constraint must be very strong in South Africa. Eduloan, as simple a mechanism as it may seem, is nonetheless a unique system in the developing world, with no known equivalent in Africa, for example.

This general result is all the more striking because South Africa is a highly financialized emerging country. Its credit-to-GDP ratio stood at 88% in 2009, much higher than that of Burkina Faso (15%), Cameroon (23%), Nigeria (26%), Ghana (32%) or Kenya (35%). This indicates a level of financial development close to that of other emerging countries such as Thailand or Vietnam (between 90% and 100% according to the IMF). Since the level of financial development is correlated to GDP per capita and South Africa is a relatively rich country in the developing world, we would expect credit constraints to be even more significant in most of the rest of the world. Therefore, it is most likely that many students in low-income or middle-income countries are also strongly affected by credit constraints, which prevent them from pursuing studies commensurate with their abilities. In this context, the development of education loan mechanisms would appear to be suitable.

5.3 Cost-efficiency

Are student loan mechanisms expensive? As mentioned above, although Eduloan is a private company, it has several partnerships with international donors. At least three development finance institutions have partnerships with Eduloan, but none have ever subsidized it. As an example, Agence Française de Développement guaranteed 50% of the amount borrowed by Eduloan from a South African bank, against an annual fee.

Such a risk-sharing agreement between a development agency and a local bank has no direct cost for the donor unless the borrower goes bankrupt or asks for some form of debt cancellation or restructuring. If there was a market for such guarantees, the annual cost of this "development project" would be:

$$c = A(p^* - p)$$

where c is the cost for the development agency, A the amount guaranteed (commonly 50% of the total loan), p^* the annual market price of such a guarantee (a form of interest rate) and p the annual price actually paid by the local institution. While p^* is not an observable parameter, we can calculate orders of magnitude. Broadly speaking, it must be a function of the borrower's bankruptcy risk and the return expected by the local bank (and the donor agency) on its off-balance sheet commitments. If we assume that Eduloan is comparable to the average South African SME, market rates for SMEs give an indirect indication of possible p^* values.

In South Africa, the Central Bank sets a repurchase rate (also called "Repo", comparable to the US Federal Funds rate or the European Central Bank refinancing rate) and a Prime overdraft rate (also called "Prime"). The Prime rate is 3 to 4% higher than the Repo and is a reference rate for households and SMEs, which generally borrow money at the prime rate plus one or two percentage points. Since commercial banks then have a 5% margin over the refinancing rate when they lend to customers, the usual 1% bank operating costs imply a market price of 4% for risk coverage.

If p^* were equal to 4% in our example, p would have to be smaller than 4%, by at least 0.5%[14] and most likely bigger than 1% (because development banks also have operating costs). We can therefore assume that p stands somewhere between 1% and 3.5%, so that c should be in the interval [0.005A; 0.03A]. This seems a modest cost on public money (roughly 2% of loan amounts according to this estimation) for a "program" that increases university enrollment by 50%.

[14] Otherwise there would not be any real interest for the local bank and no development finance institution would be necessary.

Conclusion

Having access to customer data from a private credit company (Eduloan) enables us to provide direct evidence on the impact of credit constraint on higher education enrollment in South Africa, whereas most of the related literature relies on indirect or subjective evidence. Eduloan uses a threshold to grant its loans, which allows us to implement a robust identification strategy based on a regression-discontinuity design. The causal impact of access to credit is estimated for a relevant population, that is, the individuals wishing to borrow and to attend a South African university.

We show that our sample is strongly constrained by liquidity and that obtaining a loan raised the probability of enrollment by about 23 percentage points between 2004 and 2007, a 50% increase, and raised the number of academic credits completed by borrowers by roughly 40%. We also find that effects are stronger for the poorer part of our sample, which confirms the idea that this constraint is more binding for that population. Therefore, although South Africa is a highly financialized country, liquidity constraint matters in terms of access to higher education. It may be even stronger in many other low-and middle-income countries where financial markets are more incomplete.

One important difference between our findings and the mostly US-based evidence is that in the US credit markets for human capital investment are more present (as analyzed by Lochner and Monge-Naranjo, 2011) or the wide range of subsidies to education alleviate credit market constraints more than they do in the developing world. To that extent, the mixed evidence from most of the literature is a poor guide for higher education policy in the developing world and this paper is one of the very few so far to fill the gap.

On the policy side, our findings tend to support State- or donor-sponsored loan schemes, at least in developing countries, as they are likely to offer both efficiency and equity benefits. Several such schemes do exist already but, to our knowledge, they have not been evaluated in terms of impact. This would be desirable in order to confirm the generality of our conclusions.

Acemoglu D. & J-S. Pischke (2001), “Changes in the wage

Bibliography

structure, family income, and children's education", *European Economic Review*, Vol. 45, pp. 890-904.

Attanasio O.P. & K.M. Kaufmann (2009), "School Choices, Subjective Expectations, and Credit Constraints", NBER Working Paper No. 15087.

Branson N., M. Leibbrandt & T. L. Zuze (2009), "What are the returns of tertiary education and who benefits?" in "*Responding to the Educational Needs of Post-school Youth*", N. Cloete (Ed.), Centre for Higher Education Transformation.

Brown M., J.K. Scholz & A. Seshadri (2009), "A new test of borrowing constraints for education", NBER Working Paper No. 14879.

Cameron S.V. & J.J. Heckman (2001), "The Dynamics of Educational Attainment for Black, Hispanic, and White Males", *Journal of Political Economy*, Vol. 109, No. 3, pp. 455-499.

Cameron S.V. & C. Taber (2004), "Estimation of Educational Borrowing Constraints Using Returns to Schooling", *Journal of Political Economy*, Vol. 112, No. 1, pp. 132-182.

Canton E. & A. Blom (2004), "Do student loans improve accessibility to higher education and student performance? An impact study of the SOFES program in Mexico", CPB Discussion paper 33, The Hague.

Card D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, Vol. 69, No. 5 (Sep., 2001), pp. 1127-1160.

Carneiro P. & J.J. Heckman (2002), "The Evidence on Credit Constraints in Post-Secondary Schooling", *The Economic Journal*, Vol. 112, No. 482, pp. 705-734.

Department of Education (2009), Education Statistics in South Africa 2007, Department of Education, South Africa.

Hahn J., P. Todd & W. Van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design", *Econometrica*, Vol. 69, No. 1, pp. 201-210.

Imbens G. & J. Angrist (1994), "Identification and estimation of local average treatment effects", *Econometrica*, vol. 62, pp. 467-475.

Imbens G. & T. Lemieux (2007), "Regression Discontinuity Designs: A Guide to Practice", *Journal of Econometrics*, Vol. 142, No. 2, pp. 807-828.

Keane P.K. & K.I. Wolpin (2001), "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", *International Economic Review*, Vol. 42, No. 4, pp. 1051-1103.

Keswell M. & L. Poswell (2004), "Returns to Education in South Africa: A retrospective Sensitivity Analysis of The Available Evidence", *South African Journal of Economics*, Vol. 72, No. 4, pp. 834-860.

LANG K. (1993), “Ability Bias, Discount Rate Bias and the Returns to Education”, unpublished discussion paper, Boston University.

LANG K. & P. RUUD (1986), “Returns to Schooling, Implicit Discount Rates and Black-White Wage Differentials”, *The Review of Economics and Statistics*, Vol. 68, No. 1, pp. 41-47.

LEE D. & T. LEMIEUX (2010), “Regression Discontinuity Designs in Economics”, *Journal of Economic Literature*, Vol. 48, No. 2, pp. 281-355.

LOCHNER L. & A. MONGE-NARANJO (2011), “The Nature of Credit Constraints and Human Capital”, *American Economic Review*, forthcoming.

MAURIN E. (2002), “The Impact of Parental Income on Early Schooling Transitions: A Re-examination Using Data over Three Generations”, *Journal of Public Economics*, Vol. 85, No 3, pp. 301-332.

SOCIAL SURVEYS (2009), “Access to Public Higher Education in South Africa: Financial Constraints and Consequences”, mimeo, Social Surveys, South Africa.

STATISTICS SOUTH AFRICA (2006), “Quarterly Employment Statistics, June 2006, September 2006, December 2006”.

STINEBRICKNER R. & T. STINEBRICKNER (2008), “The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study”, *The American Economic Review*, Vol. 98, No. 5, pp. 2163-2184.

STUMPF R. *et al.* (2008), *Tuition Fees: Higher Education Institutions in South Africa*, HESA Report, Higher Education South Africa.

Appendix: a lower bound to the estimator when we mix HEMIS and non-HEMIS loan requests.

We are interested in the parameter $E[Y(1) - Y(0)|E = E_0, H = 1]$ where $Y(1)$ is counterfactual enrollment when a loan is granted and $Y(0)$ when it is not. E is the Empirica score, E_0 being the identifying threshold, and $H = 1$ if the individual asked a loan for a HEMIS (*i.e.* public) institution and $H = 0$ otherwise. The parameter is defined for the HEMIS population and the problem stems from the fact that we do not observe H (or not fully so). We use the notation $E^+[.|E = E_0] = \lim_{E \to E_0^+} E(.|E)$ for the right-hand-side limit to the threshold and similarly with minus for the left-hand-side. Following Hahn, Todd & Van der Klaw (2001), let us think in terms of the Wald estimator. We can compute empirically:

$$W = \frac{E^+[Y|E_0] - E^-[Y|E_0]}{E^+[L|E_0] - E^-[L|E_0]}$$

where Y is observed outcome and L is observed loan status (obtained or not). The Wald estimator W is the parameter we compute using any of the 2SLS methods devised in the text when we pool HEMIS and non-HEMIS loan demands. For the students who applied to a private university, the public enrollment variable is always equal to zero in our data, that is formally: $Y(0) = Y(1) = 0$ if $H = 0$. Then:

$$E(Y|E) = P(H = 1|E)E[Y|E, H = 1] + 0$$

by construction and, $P(H = 1|E)$ being continuous in E_0:

$$E^+[Y|E_0] - E^-[Y|E_0] -$$

$$P(H = 1|E_0) \cdot E[Y(1) - Y(0)|E_0, H = 1] \times (E^+[L|E_0, H = 1] - E^-[L|E_0, H = 1])$$

In addition we have:

$$E^+[L|E_0] - E^-[L|E_0] =$$

$$P(H = 1|E_0) \times \left(E^+[L|E_0, H = 1] - E^-[L|E_0, H = 1]\right) +$$

$$(1 - P(H = 1|E_0)) \times \left(E^+[L|E_0, H = 0] - E^-[L|E_0, H = 0]\right)$$

Replacing in the first equation, it is straightforward to show that:

$$W = E[y(1) - y(0)|E_0, H = 1] \times$$

$$\left[1 + \frac{1 - P(H = 1|E_0)}{P(H = 1|E_0)} \frac{E^+[L|E_0, H = 0] - E^-[L|E_0, H = 0]}{E^+[L|E_0, H = 1)] - E^-[L|E_0, H = 1)]}\right]^{-1}$$

The term in brackets is clearly positive and higher than 1 so that we can write that:

$$W =< E[y_1 - y_0| E_0, H = 1]$$

which in turn means that W estimates a lower bound to the parameter of interest.

Série Documents de travail / Working Papers Series
Publiés depuis janvier 2009 / published since January 2009

Les numéros antérieurs sont consultables sur le site : http://recherche.afd.fr

Previous publications can be consulted online at: http://recherche.afd.fr

N° 78 « L'itinéraire professionnel du jeune Africain » Les résultats d'une enquête auprès de jeunes leaders Africains sur les « dispositifs de formation professionnelle post-primaire »
Richard Walther, consultant ITG, Marie Tamoifo, porte-parole de la jeunesse africaine et de la diaspora
Contact : Nicolas Lejosne, département de la Recherche, AFD - janvier 2009.

N° 79 Le ciblage des politiques de lutte contre la pauvreté : quel bilan des expériences dans les pays en développement ?
Emmanuelle Lavallée, Anne Olivier, Laure Pasquier-Doumer, Anne-Sophie Robilliard, DIAL - février 2009.

N° 80 Les nouveaux dispositifs de formation professionnelle post-primaire. Les résultats d'une enquête terrain au Cameroun, Mali et Maroc
Richard Walther, Consultant ITG
Contact : Nicolas Lejosne, département de la Recherche, AFD - mars 2009.

N° 81 *Economic Integration and Investment Incentives in Regulated Industries*
Emmanuelle Auriol, Toulouse School of Economics, Sara Biancini, Université de Cergy-Pontoise, THEMA,
Comments by : Yannick Perez and Vincent Rious - April 2009.

N° 82 Capital naturel et développement durable en Nouvelle-Calédonie - Etude 1. Mesures de la « richesse totale » et soutenabilité du développement de la Nouvelle-Calédonie
Clément Brelaud, Cécile Couharde, Vincent Géronimi, Elodie Maître d'Hôtel, Katia Radja, Patrick Schembri, Armand Taranco, Université de Versailles - Saint-Quentin-en-Yvelines, GEMDEV
Contact : Valérie Reboud, département de la Recherche, AFD - juin 2009.

N° 83 *The Global Discourse on "Participation" and its Emergence in Biodiversity Protection*
Olivier Charnoz. - July 2009.

N° 84 *Community Participation in Biodiversity Protection: an Enhanced Analytical Framework for Practitioners*
Olivier Charnoz - August 2009.

N° 85 Les Petits opérateurs privés de la distribution d'eau à Maputo : d'un problème à une solution ?
Aymeric Blanc, Jérémie Cavé, LATTS, Emmanuel Chaponnière, Hydroconseil
Contact : Aymeric Blanc, département de la recherche, AFD - août 2009.

N° 86 Les transports face aux défis de l'énergie et du climat
Benjamin Dessus, Global Chance.
Contact : Nils Devernois, département de la Recherche, AFD - septembre 2009.

N° 87 Fiscalité locale : une grille de lecture économique
Guy Gilbert, professeur des universités à l'Ecole normale supérieure (ENS) de Cachan
Contact : Réjane Hugounenq, département de la Recherche, AFD - septembre 2009.

N° 88 Les coûts de formation et d'insertion professionnelles - Conclusions d'une enquête terrain en Côte d'Ivoire
Richard Walther, expert AFD avec la collaboration de Boubakar Savadogo (Akilia) et de Borel Foko (Pôle de Dakar)
Contact : Nicolas Lejosne, département de la Recherche, AFD - octobre 2009.

N° 89 Présentation de la base de données. Institutional Profiles Database 2009 (IPD 2009)
Institutional Profiles Database III - Presentation of the Institutional Profiles Database 2009 (IPD 2009)
Denis de Crombrugghe, Kristine Farla, Nicolas Meisel, Chris de Neubourg, Jacques Ould Aoudia, Adam Szirmai
Contact : Nicolas Meisel, département de la Recherche, AFD - décembre 2009.

N° 90 Migration, santé et soins médicaux à Mayotte
Sophie Florence, Jacques Lebas, Pierre Chauvin, Equipe de recherche sur les déterminants sociaux de la santé et du recours aux soins UMRS 707 (Inserm - UPMC)
Contact : Christophe Paquet, département Technique opérationnel (DTO), AFD - janvier 2010.

N° 91 Capital naturel et developpement durable en Nouvelle-Calédonie - Etude 2. Soutenabilité de la croissance néo-calédonienne : un enjeu de politiques publiques
Cécile Couharde, Vincent Géronimi, Elodie Maître d'Hôtel, Katia Radja, Patrick Schembri, Armand Taranco Université de Versailles – Saint-Quentin-en-Yvelines, GEMDEV
Contact : Valérie Reboud, département Technique opérationnel, AFD - janvier 2010.

N° 92 *Community Participation Beyond Idealisation and Demonisation: Biodiversity Protection in Soufrière, St. Lucia*
Olivier Charnoz, Research Department, AFD - January 2010.

N° 93 *Community participation in the Pantanal, Brazil: containment games and learning processes*
Participation communautaire dans le Pantanal au Brésil : stratégies d'endiguement et processus d'apprentissage
Olivier Charnoz, département de la Recherche, AFD - février 2010.

N° 94 Développer le premier cycle secondaire : enjeu rural et défis pour l'Afrique subsaharienne
Alain Mingat et Francis Ndem, IREDU, CNRS et université de Bourgogne
Contact : Jean-Claude Balmès, département Education et formation professionnelle, AFD - avril 2010

N° 95 Prévenir les crises alimentaires au Sahel : des indicateurs basés sur les prix de marché
Catherine Araujo Bonjean, Stéphanie Brunelin, Catherine Simonet, CERDI - mai 2010.

N° 96 La Thaïlande : premier exportateur de caoutchouc naturel grâce à ses agriculteurs familiaux
Jocelyne Delarue, Département de la Recherche, AFD - mai 2010.

N° 97 Les réformes curriculaires par l'approche par compétences en Afrique
Francoise Cros, Jean-Marie de Ketele, Martial Dembélé, Michel Develay, Roger-François Gauthier, Najoua Ghriss, Yves Lenoir, Augustin Murayi, Bruno Suchaut, Valérie Tehio - juin 2010.

N° 98 Les coûts de formation et d'insertion professionnelles - Les conclusions d'une enquête terrain au Burkina Faso
Richard Walther, Boubakar Savadogo, consultants en partenariat avec le Pôle de Dakar/UNESCO-BREDA.
Contact : Nicolas Lejosne, département de la Recherche, AFD - juin 2010.

N° 99 *Private Sector Participation in the Indian Power Sector and Climate Change*
Shashanka Bhide, Payal Malik, S.K.N. Nair, Consultants, NCAER
Contact : Aymeric Blanc, Research Department, AFD - June 2010.

N° 100 Normes sanitaires et phytosanitaires : accès des pays de l'Afrique de l'Ouest au marché européen - Une étude empirique
Abdelhakim Hammoudi, Fathi Fakhfakh, Cristina Grazia, Marie-Pierre Merlateau.
Contact : Marie-Cécile Thirion, département de la Recherche, AFD - juillet 2010.

N° 101 Hétérogénéité internationale des standards de sécurité sanitaire des aliments : Quelles stratégies pour les filières d'exportation des PED ? - Une analyse normative
Abdelhakim Hammoudi, Cristina Grazia, Eric Giraud-Héraud, Oualid Hamza.
Contact : Marie-Cécile Thirion, département de la Recherche, AFD - juillet 2010.

N° 102 Développement touristique de l'outre-mer et dépendance au carbone
Jean-Paul Ceron, Ghislain Dubois et Louise de Torcy.
Contact : Valérie Reboud, AFD - octobre 2010.

N° 103 Les approches de la pauvreté en Polynésie française : résultats et apports de l'enquête sur les conditions de vie en 2009
Javier Herrera, IRD-DIAL, Sébastien Merceron, Insee - novembre 2010.
Contact : Cécile Valadier, département de la Recherche

N° 104 La gestion des déchets à Coimbatore (Inde) : frictions entre politique publique et initiatives privées
Jérémie Cavé, Laboratoire Techniques, Territoires et Sociétés (LATTS), CNRS - décembre 2010.

N° 105 Migrations et soins en Guyane - Rapport final à l'Agence Française de Développement dans le cadre du contrat AFD-Inserm
Anne Jolivet, Emmanuelle Cadot, Estelle Carde, Sophie Florence, Sophie Lesieur, Jacques Lebas, Pierre Chauvin
Contact : Christophe Paquet, département Technique opérationnel (DTO), AFD - décembre 2010.

N° 106 Les enjeux d'un bon usage de l'électricité : Chine, Etats-Unis, Inde et Union européenne
Benjamin Dessus et Bernard Laponche avec la collaboration de Sophie Attali (Topten International Services), Robert Angioletti (Ademe), Michel Raoust (Terao)
Contact : Nils Devernois, département de la Recherche, AFD - février 2011.

N° 107 Hospitalisation des patients des pays de l'Océan indien - Prises en charges spécialisées dans les hôpitaux de la Réunion
Catherine Dupilet, Dr Roland Cash, Dr Olivier Weil et Dr Georges Maguerez (cabinet AGEAL)
En partenariat avec le Centre Hospitalier Régional de la Réunion et le Fonds de coopération régionale de la Réunion
Contact : Philippe Renault, AFD - février 2011.

N° 108 *Peasants against Private Property Rights: A Review of the Literature*
Thomas Vendryes, Paris School of Economics - February 2011.

N° 109 Le mécanisme REDD+ de l'échelle mondiale à l'échelle locale - Enjeux et conditions de mise en oeuvre
ONF International
Tiphaine Leménager, département de la Recherche, AFD - mars 2011.

N° 110 L'aide au Commerce : état des lieux et analyse
Aid for trade: A survey
Mariana Vijil, Marilyne Huchet-Bourdon et Chantal Le Mouël
AGROCAMPUS OUEST, INRA, Rennes - avril 2011.

N° 111 Métiers porteurs : le rôle de l'entrepreneuriat, de la formation et de l'insertion professionnelle
Sandra Barlet et Christian Baron, GRET
Nicolas Lejosne, département de la Recherche, AFD (lejosnen@afd.fr) - avril 2011.

N° 112 Charbon de bois et sidérurgie en Amazonie brésilienne : quelles pistes d'améliorations environnementales ?
L'exemple du pôle de Carajas
Ouvrage collectif sous la direction de Marie-Gabrielle Piketty, Cirad, UMR Marchés,
Contact : Tiphaine Leménager, département de la Recherche, AFD (lemenagert@afd.fr) - avril 2011.

N° 113 Gestion des risques agricoles par les petits producteurs Focus sur l'assurance-récolte indicielle et le warrantage
Guillaume Horréard, Bastien Oggeri, Ilan Rozenkopf sous l'encadrement de :
Anne Chetaille, Aurore Duffau, Damien Lagandré
Contact : Bruno Vindel, département des Politiques alimentaires, AFD - mai 2011.

N° 114 Analyse de la cohérence des politiques commerciales en Afrique de l'Ouest
Jean-Pierre Rolland, Arlène Alpha, GRET
Contact : Jean-René Cuzon, département PSP, AFD (cuzonjr@afd.fr) - juin 2011

N° 115 L'accès à l'eau et à l'assainissement pour les populations en situation de crise :
comment passer de l'urgence à la reconstruction et au développement ?
Julie Patinet (Groupe URD) et Martina Rama (Académie de l'eau),
sous la direction de François Grünewald (Groupe URD)
Contact : Thierry Liscia, département du Pilotage stratégique et de la Prospective, AFD (lisciat@afd.fr)

N° 116 Formation et emploi au Maroc - Etat des lieux et recommandations
Jean-Christophe Maurin, division Education et Formation professionnelle, AFD
Thomas Melonio, département de la Recherche, AFD