Journal of Development Economics 161 (2023) 103031

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Student loans: Credit constraints and higher education in South Africa

Marc Gurgand[a,*], Adrien Lorenceau[b], Thomas Mélonio[b]

[a] Paris School of Economics, CNRS, ENS-PSL, 48 boulevard Jourdan, 75014 Paris, France
[b] Agence française de développement (AFD), France

ARTICLE INFO

JEL classification:
I22
O15

Keywords:
Higher education
Student loans
Credit constraint
South Africa

ABSTRACT

The empirical evidence that enrollment in higher education is constrained by access to credit is limited and usually indirect. We use a regression discontinuity design based on the fact that student loans are granted according to a score threshold at a South African credit institution (Eduloan) providing short-term loans at market conditions: we find that the credit constraint is substantial, as it reduces enrollment by more than 40 percentage points in a population of mostly middle-class applicants. However, this effect is entirely concentrated on women, and women granted a loan catch up with men's enrollment levels. This heterogeneity is not explained by lower incomes in the sample of women. It implies that women have lower access to credit, or that their options for managing without a credit are more limited than men's.

1. Introduction

Whereas primary and secondary education are almost universal in South Africa, higher education has become a severe problem in this emerging country. The enrollment rate in higher education was 19% in 2012 and 24% in 2018, low figures in comparison to the average of 29% (2012) and 36% (2018) in middle income countries.[1] Limited access is strongly concentrated on the Black African and Colored population and, generally, on the poor. This raises both efficiency and equity considerations that stand high on the political agenda.

While tuition fees are very high, typically representing between 15% and 40% of the average wage in the formal sector, or 40% to 90% of GDP per capita, wage returns to university degrees are also high. Families could leverage the wage returns to pay for the direct and indirect costs if they were able to borrow against future income (Becker, 1964). Therefore, credit constraint seems a natural explanation for this combination of high return, high cost, and low enrollment. However, although credit market imperfections are not unlikely, their 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 if the poor and/or minorities happen to lack the necessary academic qualifications or the taste for university studies, or if they are ill-informed about the benefits of education.

If credit constraint is a major problem, then a relevant policy would be to encourage the provision of student loans. As a matter of fact, the South African government has a very large public loans program, called the National Student Financial Aid Scheme (NSFAS), which may at least partly compensate for possible imperfections on the credit markets. This paper assesses the impact of another, smaller program: a private company, Eduloan, that provides short-term loans at market interest rates to pay for university fees.[2] Our sample is made up of potential students planning to enroll in a public university in one of the academic years between 2004 and 2007, and who applied to Eduloan for a loan to cover their fees. We compare the enrollment rates of individuals who obtain the loan with those of individuals who are denied it. 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.

We were able to match application and customer data from Eduloan with individual data on university students provided by the South African Ministry of Higher Education (HEMIS data). This allows us to observe loan requests, loan grants, and subsequent enrollment for a large sample of individuals. With this data, we can show that access to a loan substantially increases the probability of enrolling to 83%, from a level of about 41% in the population of unsuccessful applicants, thus doubling access. As expected, this effect tends to be stronger for families with lower incomes, indicating that they are more strongly constrained. Importantly, we find that this effect is entirely driven by female applicants. Although, in our sample of loan applicants, as well as in the general population, enrollment rates for women are slightly higher than those for men overall, women's enrollment is very low

* Corresponding author.
E-mail addresses: gurgand@pse.ens.fr (M. Gurgand), lorenceaua@afd.fr (A. Lorenceau), meloniot@afd.fr (T. Mélonio).

[1] Based on UNESCO data, http://data.uis.unesco.org/.
[2] Eduloan is now operating under the name Fundi (https://www.fundi.co.za/).

https://doi.org/10.1016/j.jdeveco.2022.103031
Received 30 December 2020; Received in revised form 19 May 2022; Accepted 9 December 2022
Available online 28 December 2022
0304-3878/

when they are not granted a loan. We are not aware of any previous mention of gender heterogeneity in this literature.

As the model is estimated on applicants to the loan, it describes the extent to which enrollment decisions are constrained for people that have limited liquidity available but potential positive returns to schooling. Our results apply to a middle-class population, Eduloan clients, that stands somewhere between the wealthiest who access bank loans, and the poorest who access the means-tested NSFAS program: their credit options are limited, but they value education as a driver of social mobility, and could have alternative options to credit to cope with the cost of education (such as working or consuming savings). The middle class is a major target for increasing participation in higher education, but the importance of credit constraint to them is an open question: we learn that a short-term credit is in fact not binding for men, whereas it is extremely so for women. Interestingly, the distribution of majors, qualifications, and type of university attended by the individuals enrolled in higher education in our sample is very close to that of the general population, so in that sense our population is not atypical.

We further find that higher female enrollment translates into more courses completed and credits earned. We also find that women denied a loan do not simply delay their enrollment decision: applicants granted a loan do enroll more either in the current year or in the next two years, and most of the impact is on the current year. One shortcoming of the data is that we do not observe enrollment in private institutions (which represent about 8% of enrollment nationally): applicants refused a loan could enroll in that sector, which we would miss. However, we are able to document the fact that transfers to cheaper institutions in response to not getting a loan are not observed in the data.

Another feature of our sample is that some of the borrowers cover their child's education, but others borrow for themselves and often target distance education and learn while working. This differs in part from the usual case in the literature where high school graduates apply for financial aid, but it covers an important segment of the South African higher education system, especially for the Black African population. Nevertheless, loan impacts, and the gender heterogeneity, do not differ strongly depending on the student's borrower status.

Although complier women (on whom the effect is identified) happen to be poorer than complier men, the gender difference remains *within* wage groups, so it does not simply reflect an income composition effect. To interpret this genuine gender difference, one must notice that we do not directly estimate credit constraints, but rather *binding* credit constraint, as some individuals may enroll even when they lack credit. As such, the outcome we observe (enrollment) results from a combination of credit constraint and preferences. Thus, the gender differences that we find can have two main sources: either both genders have similar utility values of schooling and debt but face different credit constraints in the absence of Eduloan; or they face similar constraints, but women have either lower returns to education or a higher cost of not being able to borrow, so that any given limit on their credit has a greater impact on them. We discuss those interpretations but have limited means to differentiate them strongly. The facts do not suggest that women have lower returns to education (quite the opposite), nor do we have evidence that constrained parents are less willing to make financial efforts to support girls' education. It remains possible either that women are more credit-constrained because banks are less willing to grant loans for female education; or that they are no more credit-constrained, but face a higher utility cost of attending university with limited liquidity available.

An abundant literature has examined financial constraints in access to education, especially to higher education. The robust correlation between parental income and children's education outcomes has received a number of interpretations: for instance, Cameron and Heckman (2001) and Carneiro and Heckman (2002) argue that such a link reflects long-run family factors correlated with income, such as early investment during childhood, rather than the influence of income per se or borrowing constraints. But consistent findings show an elasticity of enrollment with respect to grants and student aid,[3] and, although not systematically conclusive, substantial evidence has accumulated over the last decade to show that parents' income shocks often influence children's educational decisions or performance.[4] These findings are not explained by traits developed early in life, but they do not provide direct evidence of borrowing constraints either, because income effects may arise if education is demanded as a consumption good.

Most estimations of the role of credit constraints have taken indirect routes. For instance, Card (2001) 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 of a credit constraint for individuals of modest origins. Cameron and Taber (2004) reconsider Card's argument in a structural model and find no evidence of a credit constraint. Using a structural model, Keane and Wolpin (2001) estimate borrowing limits at college but find that they are not binding because students adjust working time and consumption rather than change their educational plans (something that may apply to men in our context). In a model of implicit contract between parent and child, Brown et al. (2012) infer that borrowing constraints can be important in some families. Lochner and Monge-Naranjo (2011) show how a number of stylized facts in the US can be rationalized by a model with borrowing constraints. Stinebrickner and Stinebrickner (2008) ask American students the hypothetical question of whether they would like to take out a loan at a fair interest rate. Generally, there is little agreement over the existence and importance of credit constraints, and the literature is strongly focused on the developed world.

This paper takes a very direct and transparent route to the empirical measurement of the existence and extent of credit constraints. It is very close to Solis (2017), who uses a regression discontinuity on an academic score in Chile to compare high school students who are granted a loan from a national program with those who are not, in a high-fees environment.[5] The findings are very similar: a doubling of the enrollment rate when a loan is granted. Melguizo et al. (2016) use a similar strategy in Colombia and find very significant, although smaller effects. As in this paper, both estimations are run on applicants to the scheme. Altogether, those papers confirm the existence and importance of credit constraints in emerging economies, where other interventions, in particular grant systems, are not as developed as in the US, to which most of the literature belongs. A recent paper by Bucarey et al. (2020) examines labor market returns to loans in the same Chilean experiment and finds no wage effect because, among compliers, access to university replaces high-quality vocational tertiary education. We have no means to assess long-term effects, but such a substitution would not happen here because our higher education outcome encompasses all types of institutions, vocational included.

However, Solis (2017) and Melguizo et al. (2016) do not provide evidence of heterogeneous exposure to borrowing constraints along

[3] See Kane (2007) for a survey and, for instance, Nielsen et al. (2010), Castleman and Long (2012), Steiner and Wrohlich (2012), Fack and Grenet (2015), Goldrick-Rab et al. (2015), or Duflo et al. (2017) for more recent evidence.

[4] Akee et al. (2010), Loken (2010), Coelli (2011), Lovenheim (2011), Dahl and Lochner (2012), Lovenheim and Reynolds (2013), Pan and Ost (2014), Hilger (2016), Manoli and Turner (2016), Bulman et al. (2017) and Bastian and Michelmore (2018). Further, Belley and Lochner (2007) have updated Carneiro and Heckman (2002) and find a stronger impact of income in the late 1990s.

[5] Rau et al. (2013) estimate the impact of the same program using a structural statistical model rather than regression discontinuity and find compatible results. Canton and Blom (2009) use data on actual loan provision in Mexico in a regression discontinuity setup, but they estimate impacts on academic performance conditional on enrollment.

gender lines, and the borrowing constraints literature has most often disregarded that dimension. In their Appendix A, Bucarey et al. (2020) find no differential effects by gender on how a loan encourages the substitution of vocational education in favor of university degrees, but they do find that women with a loan graduate from university at a higher rate than men do. There is also evidence that girls might be differentially affected by educational interventions at the secondary level in developing countries: for instance, Angrist et al. (2002) find stronger effects of a secondary education scholarship program on girls; Schultz (2004) finds stronger effects of Progresa on girls; and Duflo et al. (2017) find that a secondary education scholarship induces more girls to attend university. One interpretation is that girls receive less support from their families for education purposes. In our context, this would imply that girls have a lower total value of attending college when they cannot benefit from a loan. Alternatively, it is also possible that girls are discriminated against on the credit market (although they perform better at university and their wage returns are no lower than men's). This is a question clearly open to further investigation.

The paper is organized as follows. Section 2 provides a detailed description of how Eduloan operates in the South African higher education context. Section 3 presents the conceptual approach and empirical strategy. Section 4 describes the data. Section 5 presents the results, which are interpreted in Section 6, while their robustness is discussed in Section 7. Section 8 concludes.

2. The Eduloan scheme in the South African context

2.1. Higher education

Since the end of the apartheid regime in 1994, the South African higher education system has experienced profound changes. The government faced a challenging trade-off: to improve access for historically disadvantaged people while ensuring the development of the education system in keeping with international standards. In pursuit of the second of these objectives, it 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, whereas primary education is universal and secondary enrollment is more than 90%, enrollment in higher education reaches just 20%. Only 70% of higher education students are Black Africans although they represent 80% of the population. Moreover, the graduation rate at the undergraduate level is extremely low: between 15% and 20% depending on the qualification level and population group (Department of Education, 2016; Department of Higher Education and Training, 2019).[6] In this context, access to higher education, especially for the historically disadvantaged, remains an issue that is high on the South African political agenda.

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

An obvious source of constraint could be the "shared cost" principle implemented in the South African higher education system: since private returns to tertiary education are high, "users" are asked to finance it partially. 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 5251 million (Stumpf, 2008), for 744,000 students. The yearly average fee is thus about ZAR 7000,[7] with substantial variations 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).[8] These fees are to be compared to the average monthly wage in the formal sector, which was around ZAR 7500 in this period (Statistics South Africa, 2006), or to the GDP per capita, at about ZAR 38,000.[9] In the presence of credit constraints, such fees could well explain low enrollment and low graduation in spite of high returns.

In order to empower historically disadvantaged people and increase participation in higher education for the poorest, the government has implemented a contingent loan program (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 her salary; moreover, 40% of the loan can be converted into a bursary depending on the student's academic results. In 2004, the amounts lent ranged between ZAR 2000 and ZAR 25,000, and the program benefited 15% of students in public institutions (Stumpf, 2008), of whom 98% were historically disadvantaged.

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.[10] Informal money lenders also exist, but they charge very high interest rates. In the light of this financial environment, Eduloan holds a very specific market position.

2.2. Eduloan

Eduloan is a private financial company created in the mid-1990s. Its equity was provided by South African shareholders, and Eduloan also borrows from commercial banks to expand its activities. Three development finance institutions (International Finance Corporation, Deutsche Investitions und Entwicklungsgesellschaft, and Agence française de développement) initially participated in a risk-sharing mechanism with these commercial banks, but they did not provide any grant or soft loan. Loans offered by Eduloan are therefore not subsidized and are committed at market rates.

Eduloan provides loans to cover tuition fees for individuals planning to enroll in a public or private university in South Africa. The position of Eduloan in the student loan market is 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.

Eduloan provides short- to medium-term loans (typically 12 to 24 months) at market rate (around 1% above the prime rate, which is the reference rate charged by commercial banks to households). It is important to note that there is no subsidy component in this loan, so that impacts can be interpreted as resulting only from a change in the credit constraint, not from implicit transfers. Eduloan has two main schemes. One, called PERSAL, is based on a special agreement with the government, whereby Eduloan is allowed to deduct the repayment of loans given to civil servants directly from their salaries. This is the larger of Eduloan's two portfolios, and one that has limited repayment risk. The other scheme is traditional lending to non-civil servants, and it is a much smaller portfolio. This paper uses data from the latter scheme (data from the former is not available). In order to be eligible, borrowers must be employed and have a minimum level of income;

[6] These figures were not substantially different at the time of our dataset, 2004–2008, although the share of Black Africans in higher education was closer to 60% at the time (Department of Education, 2010). We estimate the enrollment rate in higher education by taking the number of first-time undergraduates divided by the size of a generation.

[7] This is about 1000 current US dollars.

[8] Social Surveys (2009).

[9] About 5600 current US dollars.

[10] Commercial banks are estimated to have provided about 65,000 education loans in 2008 (Department of Higher Education and Training, 2010).

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 (typically their children).

Whether the loan is granted or not also depends in part on a credit score, called the *Empirica score*. It is calculated by a credit bureau (TransUnion), based on a nationwide banking history. Although details of the algorithm are not made public, we know that it includes information such as demographics, account data (the number and size of accounts), current debts, number of credit cards possessed, as well as financial and public delinquency. The final decision to grant a loan to an applicant is largely dependent on the applicant's Empirica score being above a certain threshold. This threshold is normalized to zero in this paper. The Empirica will thus be our forcing variable for the regression discontinuity identification strategy. However, because Eduloan agents also grant loans according to other considerations, this is a fuzzy design.

Loan applications work as follows. Eduloan has an office on most public university campuses. A student must first choose the university she wishes to attend and the courses she wishes to study. Once the university has accepted her application and provided her with the corresponding fee quotation, she can apply directly to Eduloan to cover part or all of the fees. If her 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.

3. Parameters of interest and empirical strategy

In this section, we use a simple model to define parameters of interest and discuss how we can provide evidence that credit market imperfection affects university attainment in a population. Papers like Solis (2017) and this one use a credit scheme as a quasi-experiment that provides some exogenous variation to the level of credit constraint. Although we can estimate the impact of that experiment, this does not directly answer questions of more general interest, such as the share of a population that is constrained in a given state of the economy. However, it measures the extent to which individuals with limited liquidity but potential returns to schooling have to give up education.

3.1. Conceptual framework

Assume that each agent has a value V^L of not going to college, and that the full value of going to college, $V^o(d)$, is conditional on the amount of debt d that the agent incurs to finance her education. Call d^* the optimal value of debt: generally, it is heterogeneous across agents in ways that depend on observed and unobserved characteristics (it would be $d^* = 0$ for instance, if the agent is rich enough and she faces a borrowing interest rate higher than the lending interest rate; it can vary with the intertemporal elasticity of substitution, or with the disutility of working while studying, etc.). The credit constraint is defined by a limit $\bar{d}$ on d that is also heterogeneous in the population. Agents are said to be credit-constrained when $\bar{d} < d^*$. The value of going to college given $\bar{d}$ is $V(\bar{d}) = \max_{d \le \bar{d}} V^o(d)$.

An agent faces a *binding* constraint if $V(d^*) > V^L$ and $V(\bar{d}) \le V^L$: she would go to college under perfect financial markets, but given her credit constraint $\bar{d}$, she is better off not attending. The question we consider here is not whether there are liquidity constraints in the economy, but whether those who face such constraints forego college rather than enrolling, albeit at the cost of a lower utility (for instance by adjusting work or consumption): this is a major brick of knowledge for schooling policies.[11]

In an economy where there is a distribution of $\bar{d}$, an important parameter of interest is:

$$P(V(d^*) > V^L) - P(V(\bar{d}) > V^L) \tag{1}$$

This measures the proportion of those wishing to go to college who do not enroll *because* of a binding constraint. In essence, parameter (1) is very difficult to evaluate, because $\bar{d}$ and $V(d^*) > V^L$ are hard or impossible to observe.

Imagine that some policy moves the liquidity constraint from $\bar{d}$ to $\bar{d}'$ in the population, with $\bar{d}' \ge \bar{d}$. We can define the policy parameter:

$$\Delta = P(V(\bar{d}') > V^L) - P(V(\bar{d}) > V^L)$$

which is the change in the share of the constrained population resulting from that move. It provides a lower bound to the proportion of constrained individuals in the economy under $\bar{d}$, as $\Delta \le P(V(d^*) > V^L) - P(V(\bar{d}) > V^L)$ by construction.

Similar to Solis (2017), the policy used in this paper is an exogenous (quasi-experimental) variation in liquidity constraints, identified by comparing individuals facing $\bar{d}$ and $\bar{d}'$ on either side of a loan eligibility threshold. In such a design, Δ can only be estimated on beneficiaries that applied to the scheme, i.e., individuals such that[12]:

$$\begin{cases} \bar{d}' > \bar{d} \\ V(\bar{d}') > V^L \end{cases}$$

Call this population "A" for applicant. Because non-A individuals are not affected by the exposure to $\bar{d}'$ instead of $\bar{d}$, it is easy to show that[13]:

$$\begin{aligned} \Delta &= P(A)[P(V(\bar{d}') > V^L|A) - P(V(\bar{d}) > V^L|A)] \\ &= P(A)P(V(\bar{d}) \le V^L|A) \end{aligned} \tag{2}$$

and the parameter in brackets is identified by the differences in enrollment rates resulting from the exogenous variation in the credit constraint faced by the applicant population (A).

Outside of "A" are individuals for whom $\bar{d} \ge \bar{d}'$, i.e., those who do not face such a strong constraint, either because they have access to bank loans or their own resources; or, at the other end of the spectrum, because they have access NSFAS, and it provides them with sufficient support.[14] Outside of "A" are also individuals for whom $V(\bar{d}') \le V^L$, i.e., returns to schooling are too low, even with a loan. Therefore, what we learn from this kind of empirical design is the value of relieving the credit constraint on a population that has potential positive returns to schooling but has limited liquidity available. This is not a trivial question, as we will learn that the constraint is in fact not binding for men, whereas it is extremely so for women.

3.2. Empirical strategy

Consider the following model, estimated over a sample of Eduloan applicants indexed by i:

$$Y_i = \alpha + \beta T_i + \varepsilon_i \tag{3}$$

[11] The distinction is important: for instance, Keane and Wolpin (2001) estimates that borrowing constraints are high among US students, although they are rarely binding because constrained students choose to enroll *and* work for a wage or consume less. The outcome considered in this paper, as in most of the literature, is enrollment rates and not utility losses resulting from liquidity constraints.

[12] In the regression discontinuity design, the identification is local (close to the eligibility threshold), but for now, we discuss the substantive aspects of this quasi-experiment.

[13] Use the fact that non-A individuals have either $V(\bar{d}') \le V^L$ and $\bar{d}' \ge \bar{d}$ or $V(\bar{d}') > V^L$ and $\bar{d}' = \bar{d}$.

[14] To give some context, NSFAS provided 153,795 loans in 2008 (Department of Higher Education and Training, 2010) when there were about 800,000 students in South Africa (Department of Education, 2010).

where Y is a dummy for enrollment in higher education in a given year and T is a dummy for obtaining a loan that same year (therefore facing the liquidity constraint $\bar{d}' > \bar{d}$); α and β are parameters to be estimated; and ϵ is a residual that contains unobserved determinants of enrollment other than the Eduloan loan. In terms of the above model, $\alpha = P(V(\bar{d}) > V^L|A)$ and $\beta = P(V(\bar{d}') > V^L|A) - P(V(\bar{d}) > V^L|A)$. Because ϵ may be correlated with T, however, the simple regression of enrollment on loan obtention does not provide a parameter with a causal interpretation.

In order to identify a causal impact, we use the regression discontinuity design (Imbens and Lemieux, 2007; Lee and Lemieux, 2010). We exploit the presence of the Empirica score (noted E)—the credit score that strongly influences Eduloan's decision to provide the loan. Call E_0 the threshold used to assess eligibility. The discontinuous relationship between T and E at E_0 identifies the causal impact of loan obtention on enrollment if all other determinants (ϵ) vary continuously with E, at least in the neighborhood of E_0. This strategy is in essence very similar to randomization, to the extent that individuals happen to have a few more or a few less points in E merely by 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. 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). This hypothesis will be confirmed by the manipulation test below.

The first-stage model describing the discontinuous relationship between loan obtention and the Empirica score is:

$$T_i = g(E_i) + \delta D_i + u_i \tag{4}$$

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 in granting loans.

Similarly, the structural equation can be rewritten as:

$$Y_i = \alpha + \beta T_i + f(E_i) + \epsilon'_i \tag{5}$$

where ϵ in (3) has been expressed as some continuous function of E, $f(E)$. In this model, D is a valid instrument for T, so that (5) can be estimated using an instrumental variable approach. Our baseline specifications will use local linear regressions that restrict the sample to the neighborhood of E_0, and approximate $f(E)$ (and $g(E)$) as linear functions, with different slopes on the right and on the left of the discontinuity point E_0. The bias is minimized when the sample is strongly restricted to the neighborhood of E_0, but the precision increases as the sample gets larger. In this context, Imbens and Kalyanaraman (2012) have provided a data-driven framework to determine the optimal bandwidth. We use the Calonico et al. (2014) version of the optimal bandwidth estimation (see Calonico et al., 2017). In most tables, we provide a set of different bandwidths next to the optimal one, as well as estimations on the full sample or large bandwidths, using a flexible specification for $f(E)$ (quadratic with different shapes on each side of the discontinuity).

This regression discontinuity (RD) strategy has well-known limitations. First, identification is local: strictly speaking, it is relevant only for the population close to the threshold. With respect to the discussion in the previous section, in practice we do not estimate a parameter that is valid for the whole population of Eduloan applicants, but rather for those who are at the margin of eligibility. In the optimal bandwidth estimation, we will be typically using 15% of the baseline sample. Table A.2 compares the characteristics of the full sample and the baseline optimal bandwidth sample on the set of variables that will be presented in Table 1 later on. The samples are very similar, in spite of the fact that the optimal sample is centered around lower values of the Empirica: in particular, the borrower's net salary is extremely close in the two samples, and the requested loan amount is ZAR 6756 vs. ZAR 6642; age and gender are also close. There is thus no reason to consider that the local estimation reflects impacts on a very particular population, and it may have some degree of external validity to the applicant population.

Table 1
Descriptive statistics on loan requests, 2004–2007.

	Loan requested for public institution			
	No loan granted		Loan granted	
	Mean	S.d.	Mean	S.d.
Male	0.46	0.50	0.48	0.50
Age	27.79	8.41	27.58	7.86
Monthly wage	6403	5045	7515	7360
Missing wage information	0.15	0.36	0.00	0.00
Requested loan value	7246	6101	6274	4569
Requested loan/monthly wage	1.54	1.66	1.04	0.80
Missing requested loan value	0.03	0.17	0.00	0.06
Student is the borrower	0.49	0.50	0.49	0.50
Empirica	24.93	57.85	77.46	53.29
Enrollment in public university	0.53	0.50	0.75	0.43
Nb courses registered (if enrolled)	7.14	4.38	6.82	4.13
Nb courses completed (if enrolled)	4.62	4.09	4.30	3.94
Credits completed (if enrolled)	0.46	0.37	0.44	0.37
Nb observations	4854		4801	
	Loan requested for private institution			
	No loan granted		Loan granted	
	Mean	S.d.	Mean	S.d.
Male	0.46	0.50	0.49	0.50
Age	26.57	9.30	26.21	8.55
Monthly wage	5900	4361	6730	4930
Missing wage information	0.15	0.35	0.00	0.00
Requested loan value	9876	8112	8425	6967
Requested loan/monthly wage	2.24	2.28	1.44	1.26
Missing requested loan value	0.07	0.25	0.06	0.24
Student is the borrower	0.35	0.48	0.36	0.48
Empirica	18.63	57.56	62.00	55.95
Enrollment in public university	0.11	0.31	0.18	0.39
Nb courses registered (if enrolled)	6.45	4.55	7.41	4.36
Nb courses completed (if enrolled)	4.15	4.13	4.72	4.31
Credits completed (if enrolled)	0.40	0.35	0.44	0.35
Nb observations	1582		763	

Notes: Eduloan and HEMIS data from 2004 to 2007. The unit of observation is loan requests per student per year; when several applications have been sent for a given student in the same year, we use the average requested loan value. Applications dated November/December are excluded, as in all baseline specifications. This table only uses data where the type of higher education institution targeted is identified: the upper panel is restricted to loan applications to a public university; the lower panel is restricted to loan applications to a private university.

Second, as shown by Hahn et al. (2001), if the treatment effect is heterogeneous and correlated with compliance, then the estimated parameter in the fuzzy RD design is a local average treatment effect (LATE) in the sense of Imbens and Angrist (1994), and is only valid for compliers. We will show counterfactual enrollment rates of compliers as well as compliers' characteristics, and compare them to the estimation sample; they are often not strongly different, although the wages of complier women are lower, something we will discuss below.

4. Data

This paper uses data from two distinct sources. Customer data from Eduloan describe loan applications and acceptance or rejection decisions and therefore provide information on the treatment variable. Administrative data (referred to as HEMIS data) provided by the Ministry of Education identify students enrolling in any public higher education institution, thus informing the outcome variable. These two datasets were matched using the national identification number.

4.1. Eduloan data

As a private credit company, Eduloan maintains customer files on both the whole set of applicants and its actual customers. It has provided us with two datasets. The first 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 the characteristics of the applicant such as the borrower's net salary, the institution applied for, the loan amount requested, her age, and so on. The second dataset contains actual customers, i.e., the students whose loan applications were 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 dataset, we can observe several application dates per applicant per year. These may be either duplicate administrative records for the same request or individuals who actually applied 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. However, because our outcome (university enrollment) is a yearly event, it is enough for us to know whether, for a given year, applications were sent and loans were obtained by a given applicant.

In most of the empirical analysis, we use data from 2004 to 2007; during this period, the Empirica threshold value for granting loans remained unchanged and generated a discontinuity. In 2008, Eduloan's activities were strongly impacted by the credit crunch following the financial crisis, and the threshold had almost no explanatory power. We use the 2008 data only for a robustness analysis.

4.2. 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 Higher Education Management Information System (HEMIS) was therefore created to collect accurate individual data on each and every student entering the public higher education system. The data contains information on all the courses and qualifications undertaken by a student throughout her studies in a 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 they obtained a loan from Eduloan and whether they enrolled and completed their credits in a public higher education institution during the relevant year.

4.3. Data limitations

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 small but not negligible: in 2008, about 8% of students were enrolled in the private sector (Department of Higher Education and Training, 2014).

Fortunately, loans are granted in order to pay fees to a specific institution, and they are paid directly to that institution by Eduloan when enrollment is effective. When a loan has been requested for a public institution, we therefore know whether granting the loan has indeed increased the likelihood that the applicant actually enrolls in that institution. Our data contains a variable for the type of institution for which the student has requested a loan, although this variable is missing for about 17% of observations. Where the information is available, a large majority of students (80%) applied to a public institution, compared to 20% for private ones.

Our baseline analysis will be restricted to applicants to public institutions, excluding loan requests for private or unknown institutions. We checked that this sample selection is independent from having an Empirica score on either side of the threshold (see Table 3 below). 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 to the effect on HEMIS perimeter individuals. But we do not make any claims about the population wishing to enter private institutions.

The other technical difficulty is to match dates between applications, loans granted, and enrollment. The academic year is the civil year in South Africa. The norm is that students register in January and then ask for a loan: 44% of our application dates are in January or February, and 53% in the first three months. But 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 per year. We consider that loans requested in year t have been granted whenever the same student has made one or more applications during year t and his or her request was approved during the same year. We do not have the data informing whether the loan was then actually disbursed; however, we can see from Table 1 that not all students granted a loan actually enrolled; therefore, loan grant and enrollment are separate things.

There is an ambiguity, however, when loan applications are made late in the year and a loan is granted at the beginning of the next year. We do not know whether it is intended to pay for late fees or 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 next one. As a result, our baseline estimation excludes individuals for whom the only application of the year was posted in November or December (we then keep 86% of our sample). As a robustness check, we will show that the results are not sensitive to inclusion of those observations.

Finally, it is worth mentioning that we had to drop some observations for which the national identification number was missing or obviously incorrect. Also, individuals with no credit history, thus no Empirica score, are excluded from the whole analysis.

4.4. 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 student who has applied for loans during a given year. As explained above, when the earliest application was made in November or December, the loans/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 and for a private institution. 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 (49% for public, 35% for private institutions) 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 to get 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 (UNISA) is dedicated to distance education.[15] As a matter of fact, in our data, students who are also the borrower are older, end up much more often in distance

[15] In 2008, 39% of South African students were registered in distance education (Department of Higher Education and Training, 2014).

education programs, are more often men, and have lower wages than those borrowers who are parents paying for their child. However, as we will mention below (Section 6.1 and Table A.6), there are no strong differences in the effect of a loan between those two groups.

Interestingly, the qualification types, the major field of studies and the type of institutions attended among those enrolled in higher education in our sample are extremely close to the proportions in the whole South African education system. For instance, Table A.1 shows that almost the same proportions are enrolled on Business and Management, STEM, and Humanities courses; we have only a few more undergraduates (85% vs. 82%), and a few more of them pursue a degree rather than a certificate, and a few less a Master or Doctorate (4% vs. 7%); also, almost the same share attend Technikons (as opposed to Universities) and UNISA. In that sense, the population enrolled in our sample is quite typical of the general population in higher education.

Borrowers declare wages that are relatively high by South African standards: their average monthly wage is between ZAR 6000 and 7500. This is to be compared with the average wage of the population in formal employment, which was around ZAR 7500 in this period (Statistics South Africa, 2006). Given that wages are usually skewed, and taking into account the existence of informal employment, it is very likely that our population of borrowers are 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 credit constraint in doing so. As a matter of fact, requested loan values represent on average one to two months' wages, an amount that households may find difficult to make available up-front, but which they 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 46% of applications. Loans are granted more often to borrowers who declare higher wages (by about ZAR 1000). However, the proportion of men, the proportion of students who are themselves the borrowers, and their ages do not differ much according to loan status.

When we consider loans requested for a public university, 75% of students who were granted a loan ended up 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 a 41% increase in 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. Another 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 high 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 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.[16]

[16] In South Africa, one year of higher education represents 1.0 credits, so that a typical academic year is made of 10 courses, each one worth 0.1 credits: our descriptive statistics recall the low completion rate of students, whether they get a loan or not.

Table 2
Loan granted as a function of Empirica score.

	Full sample	Bandwidth				
		±50	±30	±15	±10	**±20**
	(1)	(2)	(3)	(4)	(5)	(6)
Above discontinuity	0.354 (0.020)	0.360 (0.031)	0.329 (0.040)	0.348 (0.037)	0.358 (0.046)	0.328 (0.032)
Intercept	0.088 (0.015)	0.097 (0.017)	0.105 (0.022)	0.099 (0.020)	0.091 (0.026)	0.101 (0.017)
Linear in Empirica				x	x	x
Quadratic in Empirica	x	x	x			
Number of obs.	9655	4983	3336	1798	1225	2340

Notes: Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). The explained variable is a dummy for a loan being granted the same year a loan application was received; "Above discontinuity" is a dummy for when the Empirica score is above zero. Ordinary least squares estimation with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014).

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: this 7-point difference does not survive a causal estimation.[17] Also, looking at courses and credits, conditional on studying in a public university, those students do not appear different from the rest of the enrolled population.

5. Baseline results

5.1. Validity of the research design

Fig. 1 shows the probability of obtaining a loan, as a function of the Empirica score E, for loans requested for a public university over the years 2004–2007. The vertical bar marks the Empirica threshold, E_0 (normalized to 0).[18] Each point represents the proportion of applicants that obtained a loan among individuals within values of E in bins of size 5. On the left of E_0, the probability of obtaining a loan is small, although not zero. It then increases smoothly with the Empirica. There is a very strong discontinuity above the threshold: the probability of obtaining a loan jumps from about 10% to about 45%. This ensures that the instrument will have identifying power. Table 2 presents the estimation of Eq. (4). Column (1) corresponds to Fig. 1, using the full sample. Given the normalizations, the intercept measures the proportion of loans granted on the left of the discontinuity point, which is around 10%. The bandwidth in bold, column (6), indicates the optimal bandwidth for the local linear estimation (Calonico et al., 2014); it restricts the sample to ±20 Empirica points. We estimate that, above the threshold, the probability of obtaining a loan increases by 32.8 points. Other bandwidths around the optimal one are also presented. The results are very robust and strongly significant.

We can also check that E_0 is not a threshold for variables other than loan granted. Table 3 shows the coefficient of the discontinuity variable in local linear estimations for several predetermined characteristics and for an index of those characteristics (propensity score). Each estimation has its own optimal bandwidth and corresponding number of observations. There is no evidence of discontinuous change in the borrowers' gender, age, whether the borrower is the student, the loan

[17] See Table C.1 and further discussion in the Robustness section.

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

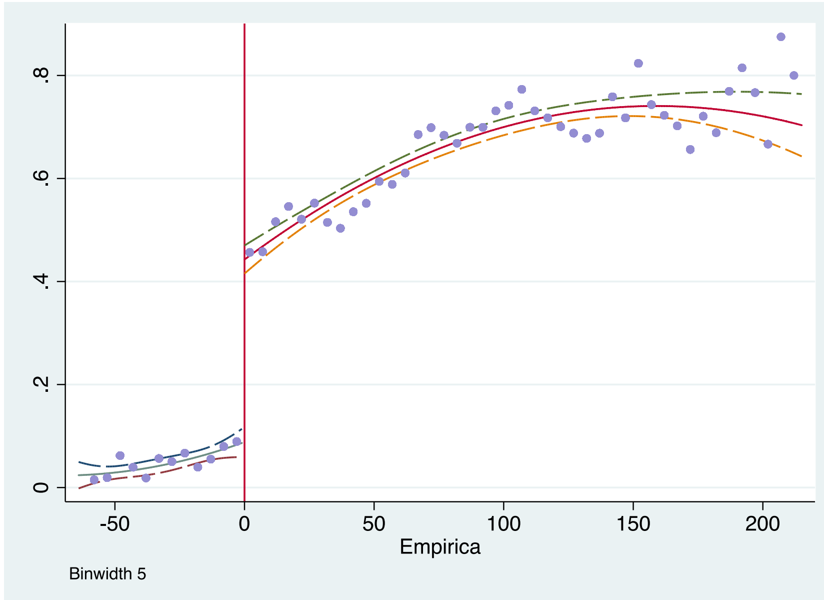

Fig. 1. Share of loans granted as a function of Empirica score (quadratic fit with 95% confidence intervals). *Notes:* Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). Each dot represents the proportion of loans granted the same year a loan application was received, within bins of 5 Empirica points. The quadratic fit uses the estimation in column 1 of Table 2.

Table 3
Placebo: Predetermined variables as a function of Empirica score.

	Coefficient on discontinuity variable (1)	Optimal bandwidth (2)	Number of obs. (3)
Applied public U.	0.003 (0.035)	±15	2316
Male	−0.012 (0.055)	±11	1331
Age	0.053 (0.856)	±13	1571
Student is the borrower	−0.028 (0.049)	±14	1693
Requested loan value	442.224 (568.484)	±13	1530
Monthly wage	761.386 (604.229)	±11	1173
Propensity score	0.005 (0.007)	±13	1347

Notes: Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). Each line is a separate ordinary least squares regression, the explained variable of which is given on the left-hand side. Column 1 gives the coefficient (robust standard errors in parentheses) of that regression on a dummy for when the Empirica score is above zero. There are no controls other than linear functions of the Empirica score (different on each side of the discontinuity). Each model is fit on its optimal bandwidth based on Calonico et al. (2014): the optimal bandwidth, defined with respect to the Empirica score, is given in column (2) and the corresponding number of observations in column (3). Propensity score is an index of the variables Male to Monthly wage, built from a regression of obtaining a loan on those variables.

amount requested, or the monthly wage or the index. Using the larger sample of individuals that asked for a loan for a public *or* a private institution, we do not see any discontinuity at the Empirica threshold in the choice of a public rather than a private one: our sample (public applicants) is thus not selected along the instrument.

Finally, we test for manipulation of the forcing variable around the threshold. Following McCrary (2008) and Cattaneo et al. (2018, 2020) devised a class of tests based on local polynomial estimation of the cumulative distribution function (CDF) of the forcing variable. Table 4 presents tests of equality of the densities on each side of the threshold, for different polynomial orders of the local approximation of the CDF. Bandwidth choice is data-driven, using the mean squared error criteria; this can be unrestricted, but one can also impose that the bandwidth size is identical on each side of the threshold. We present the two variants, as this generates different numbers of observations, and thus allows us to assess the robustness of the results for varying estimation samples. Column (3) shows the difference between the right-hand side and the left-hand side densities, and column (4) provides the p-value for a test of equality between the two. The differences between the two densities vary in sign with the specification, and they are occasionally significant. We believe that this instability is a matter of finite distance estimation. Ultimately, it is very difficult to believe that the Empirica (which is computed by a firm using an unknown algorithm, and based on a predetermined credit history) can be manipulated.

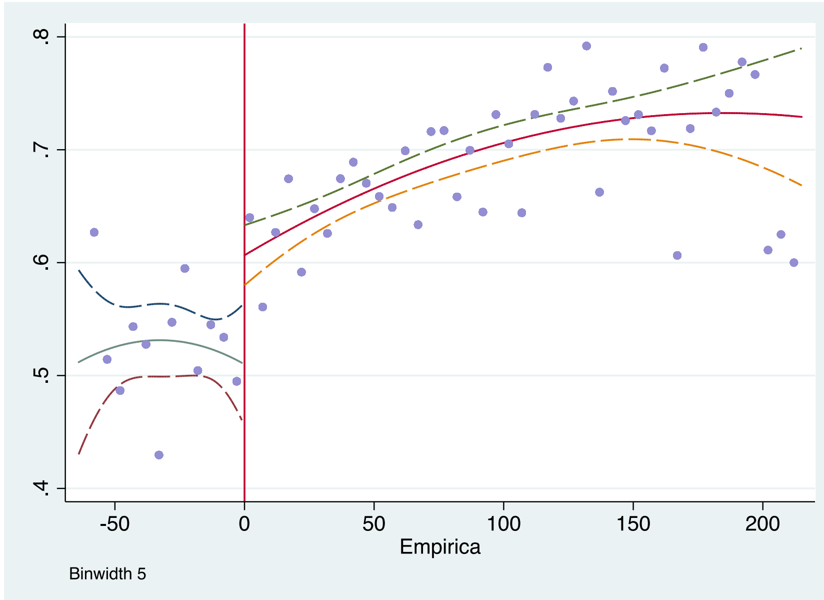

Fig. 2. Share of university enrollment as a function of Empirica score (quadratic fit on full sample with 95% confidence intervals). *Notes:* Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). Each dot represents the proportion of loan applicants that enrolled in a public university the same year a loan application was received, within bins of 5 Empirica points. The quadratic fit uses the estimation in column 1 of Table 5.

Table 4
Test of equality of Empirica densities on each side of the threshold.

Order polynomial for CDF specification (1)	Bandwidth selection (2)	Diff. in densities (3)	*p*-value (4)	Bandwidths (5)	Number of obs. (6)
1	Specific to side	−0.001	0.213	7–30	2398
1	Same both sides	−0.002	0.015	9–9	1110
2	Specific to side	0.001	0.073	28–82	5907
2	Same both sides	0.000	0.519	32–32	3516
3	Specific to side	0.000	0.775	41–67	5620
3	Same both sides	0.000	0.616	49–49	4920
4	Specific to side	0.001	0.269	53–136	8349
4	Same both sides	0.000	0.897	56–56	5437
5	Specific to side	−0.002	0.218	33–113	7174
5	Same both sides	−0.003	0.135	35–35	3769

Notes: Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). This table tests the hypothesis that the density of the Empirica score is continuous at the cutoff point, using Cattaneo et al. (2018, 2020). The density estimation is run separately on each side of the discontinuity and uses a local polynomial approximation of the cumulative distribution function. We run tests for different orders of that approximation, and each line is a different test: column 1 gives the order used for each test run separately. Columns 2 and 3 show the test statistic and its *p*-value respectively (the test is positive when the estimated density on the right-hand side is higher). The local approximation is run on a bandwidth based on the mean square error of each density separately: column 4 gives the number of observations used, given the optimal bandwidth for each test.

5.2. Overall impact of loans on enrollment

Table 5 and Fig. 2 show the reduced-form relationship between enrollment and the Empirica score. The probability of being enrolled in a public university, for individuals who applied for a loan to study at such a university, increases precisely at the threshold E_0. Fig. 2 shows the full sample, with a quadratic fit (each dot is an average for bins of 5 Empirica points). Table 5 presents different samples, including the optimal bandwidth sample on the last column. The effect at the discontinuity is strong and very significant: starting at around 50% on the left-hand side of the discontinuity (intercept), enrollment increases by 14.8 percentage points on the right-hand side for the optimal bandwidth estimation (column (6)). The exact point estimates vary with the specification, but they remain in a 10 to 20 point range.[19]

Table 6 presents estimates of the structural Eq. (5). The ordinary least squares (OLS) estimation indicates that obtaining a loan increases enrollment by 20 percentage points. The instrumental variable estimation, using the discontinuity dummy as an instrument, raises this effect to about 27 points when using the full sample and a quadratic fit for the running variable. A somewhat stronger effect is found when the instrumental variable estimation is restricted to the optimal bandwidth (±11 points, column (7)): the effect is 41.9 percentage points, and this order of magnitude is quite stable across specifications.

All in all, we estimate that providing a loan to members of this population causally increases the probability that they will enroll in

[19] Naturally, identification is local (valid for the population around the threshold): Table A.2 compares the characteristics of the full baseline sample with that of the optimal bandwidth sample. They are not strongly different; in particular, monthly wages of the borrower are very similar and so are requested loan values. There are modest differences in terms of gender and borrower status, but in general the local sample does not seem a very distorted one.

Table 5
University enrollment as a function of Empirica score.

	Full sample	Bandwidth				
		±50	±30	±15	±10	**±11**
	(1)	(2)	(3)	(4)	(5)	(6)
Above discontinuity	0.097 (0.031)	0.121 (0.040)	0.135 (0.050)	0.234 (0.072)	0.162 (0.057)	0.148 (0.054)
Intercept	0.510 (0.028)	0.496 (0.031)	0.484 (0.039)	0.436 (0.058)	0.477 (0.044)	0.482 (0.042)
Linear in Empirica					x	x
Quadratic in Empirica	x	x	x	x		
Number of obs.	9655	4983	3336	1798	1225	1331

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). The explained variable is a dummy for enrollment in a public university the same year a loan application was received; "Above discontinuity" is a dummy for when the Empirica score is above zero. Ordinary least squares estimation with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014).

Table 6
University enrollment as a function of loan obtention.

	Full sample		Bandwidth (2SLS)				
	OLS	2SLS	±50	±30	±15	±10	**±11**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan granted	0.203 (0.011)	0.273 (0.087)	0.337 (0.110)	0.411 (0.153)	0.663 (0.215)	0.452 (0.159)	0.419 (0.155)
Intercept	0.512 (0.013)	0.486 (0.035)	0.463 (0.040)	0.441 (0.053)	0.373 (0.078)	0.436 (0.056)	0.443 (0.055)
$E(y_0 \mid Comp)$		0.516	0.447	0.381	0.219	0.369	0.407
Linear in Empirica						x	x
Quadratic in Empirica	x	x	x	x	x		
Number of obs.	9655	9655	4983	3336	1798	1225	1331

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). The explained variable is a dummy for enrollment in a public university the same year a loan application was received; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Ordinary least squares estimation (column 1) and two-stage least squares (columns 2–7) with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. The instrument for 2SLS is a dummy for when the Empirica score is above zero. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014). $E(y_0 \mid Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

higher education from a level of about 41% among the compliers to 83%, thus doubling access. 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 those variables (as demonstrated in Table 3). Including them does not systematically improve the precision of the estimation, so we present the simple regressions that are more transparent.

5.3. Heterogeneous impacts

These are average effects, but there is substantial heterogeneity along two dimensions: gender and wage of the borrower. Table 7 presents each of the three regressions (first stage, reduced form, and 2SLS) for men and women separately. For each gender, we use the reduced form optimal bandwidth and impose it on the two other estimations (first stage and 2SLS) for clarity. Tables A.3 and A.4 reproduce Tables 3 and 4 for the men and women samples separately, to check for imbalance on the predetermined covariates and bunching at the threshold: there is no evidence of either.

Male–female results are strikingly contrasted. First, the first stage is much stronger for women: on the left-hand side of the discontinuity, the share with a loan is similar, just below 10%; but being on the right-hand side of the Empirica threshold increases women's probability of obtaining a loan much more (by 42 percentage points compared to 24 percentage points for men—the p-value for the test of equality is 0.014). Second, the reduced form effect on enrollment is very small and non-significant for men, whereas it is strong and very significant for women. Fig. 3 illustrates those reduced forms for men and women separately. Figs. A.1 and A.2 in the Appendix A show smoother graphs using the full data, and also provide specifications that use placebo cutoff values for the Empirica (from −15 to +15): although those graphs bear the risk of false positives, the only specification that is clearly significant is for women with the true threshold (0).

Overall, the effect on enrollment of obtaining a loan is entirely concentrated on women: the impact is small and undistinguishable from zero for men (column (3)), whereas loan access increases enrollment by 47.8 percentage points for complier women (column (6)): their enrollment rate moves from about 38% to about 86% when they obtain a loan.[20] We are not aware of comparable findings in the literature, but

[20] The LATE from the full sample, 0.419, may not seem consistent with the LATEs from the two subsamples, 0.052 and 0.478. But it is. The fact is that

Table 7
University enrollment as a function of loan obtention by gender (optimal bandwidth samples).

	Men			Women		
	First stage (Loan) (1)	Red. form (Enrollment) (2)	2SLS (Enrollment) (3)	First stage (Loan) (4)	Red. form (Enrollment) (5)	2SLS (Enrollment) (6)
Above discontinuity	0.240 (0.050)	0.013 (0.067)		0.422 (0.055)	0.202 (0.066)	
Loan granted			0.052 (0.278)			0.478 (0.158)
Intercept	0.096 (0.026)	0.531 (0.051)	0.526 (0.073)	0.098 (0.032)	0.462 (0.051)	0.415 (0.065)
$E(y_0\|Comp)$			0.600			0.381
Number of obs.	896	896	896	876	876	876

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. In columns 1 and 4, the explained variable is a dummy for a loan being granted the same year a loan application was received. In the other columns, the explained variable is a dummy for enrollment in a public university the same year a loan application was received. "Above discontinuity" is a dummy for when the Empirica score is above zero; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. The models in columns 1, 2, 4, and 5 are estimated by ordinary least squares, and the models in columns 3 and 6 are estimated by two-stage least squares with "Above discontinuity" as an instrument. No controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. For each gender, the sample is restricted to the optimal bandwidth sample for the reduced form based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

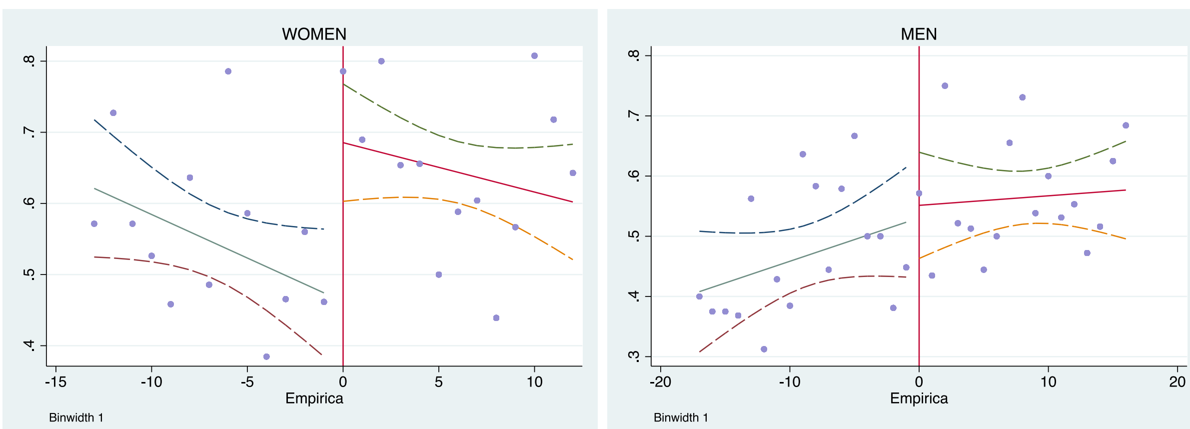

Fig. 3. Share of university enrollment as a function of Empirica score by gender (linear fit on optimal bandwidth with 95% confidence intervals). *Notes:* Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded), separately by gender of the applicant student, restricted to Empirica scores belonging to the optimal bandwidth for each gender. Each dot represents the proportion of loan applicants that enrolled in a public university the same year a loan application was received, within bins of 1 Empirica point. The linear fits use the estimations in columns 2 and 5 of Table 7 respectively.

this has not been typically looked at. Table A.5 shows a set of different specifications for the structural equation, with different bandwidths, and, as for Table 6, results are consistent across specifications. For men, point estimates are higher when we do not use the optimal bandwidth, but still much lower than those for women, and not significant.

Table 8 further decomposes the population by borrower's wage, again presenting first stage, reduced form, and 2SLS, all at each sample's reduced form optimal bandwidth. We separate the samples into applicants whose borrower's wage is above the median of our full baseline sample of applicants, and those whose wage is below the median. For women, the impact of the loan is about four times as large among the poorest, as compared to the richest (where it is not statistically significant). The large effect that we found earlier for women is thus largely driven by our sample's poorest households. A comparable income gradient is found by Solis (2017) (for men and women together). For men, the impact of Eduloan is almost zero on the richest, and again much larger on the poorest, although in this case not significant. All this is indicative of a plausible fact: that credit constraint is stronger for less wealthy families and that fewer financing alternatives exist at the bottom of our 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. Another is that the poorer applicants have to abandon their higher education project when they do not obtain a loan, whereas richer households have the margin to make sacrifices on consumption for instance, or have savings of their own.

Three more outcome variables are shown in Table 9: the number of courses registered for, the number of courses among them for which credit was obtained, and the value of the credits obtained.[21] Each of

a LATE in a population is the weighted sum of LATEs in subpopulations with weights that depend on the population shares *and* on the relative size of the first stages. Further, the optimal bandwidths are different in each specification (±11 for the full sample, ±13 for women and ±17 for men), which blurs slightly the consistency.

[21] As mentioned earlier, the HEMIS accounting system normalizes credits to 1 per full-time academic year equivalent, so that a completed full-time year would show a credit of 1.

Table 8
University enrollment as a function of loan obtention: By gender and borrower wage (optimal bandwidth samples).

	Women					
	Wage below median			Wage above median		
	First stage (Loan) (1)	Reduced form (Enrollment) (2)	2SLS (3)	First stage (Loan) (4)	Reduced form (Enrollment) (5)	2SLS (6)
Above discontinuity	0.376 (0.096)	0.304 (0.118)		0.464 (0.078)	0.089 (0.093)	
Loan granted			0.810 (0.348)			0.193 (0.199)
Intercept	0.104 (0.049)	0.476 (0.093)	0.392 (0.138)	0.166 (0.055)	0.549 (0.077)	0.517 (0.105)
$E(y_0\|Comp)$			0.380			0.550
Number of obs.	278	278	278	460	460	460
	Men					
	Wage below median			Wage above median		
	First stage (Loan) (1)	Reduced form (Enrollment) (2)	2SLS (3)	First stage (Loan) (4)	Reduced form (Enrollment) (5)	2SLS (6)
Above discontinuity	0.213 (0.072)	0.049 (0.093)		0.324 (0.079)	0.003 (0.104)	
Loan granted			0.232 (0.432)			0.010 (0.323)
Intercept	0.094 (0.040)	0.491 (0.071)	0.469 (0.106)	0.105 (0.045)	0.515 (0.081)	0.514 (0.110)
$E(y_0\|Comp)$			0.596			0.516
Number of obs.	492	492	492	359	359	359

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student and the wage of the borrower; the median wage is computed over the full sample of male and female students (i.e., the baseline 9655 observations). In columns 1 and 4, the explained variable is a dummy for a loan being granted the same year a loan application was received. In other columns, the explained variable is a dummy for enrollment in a public university the same year a loan application was received. "Above discontinuity" is a dummy for when the Empirica score is above zero; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. The models in columns 1, 2, 4, and 5 are estimated by ordinary least squares, and the models in columns 3 and 6 are estimated by two-stage least squares with "Above discontinuity" as an instrument. No controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. For each gender and wage group, the sample is restricted to the optimal bandwidth sample for the reduced form based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

Table 9
University outcomes as a function of loan obtention, by gender (optimal bandwidth samples).

	Men			Women		
	Nb courses registered (1)	Nb courses completed (2)	Credits completed (3)	Nb courses registered (4)	Nb courses completed (5)	Credits completed (6)
Loan granted	−0.782 (2.366)	−1.223 (1.946)	0.074 (0.164)	5.201 (1.731)	3.410 (1.208)	0.264 (0.114)
Intercept	3.671 (0.677)	2.236 (0.536)	0.186 (0.045)	2.507 (0.651)	1.345 (0.450)	0.165 (0.042)
$E(y_0\|Comp)$	3.773	2.383	0.138	1.393	1.108	0.140
Number of obs.	745	695	745	751	876	948

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. In column 1, the explained variable is the number of courses the student registered in at a public university the same year a loan application was received; in column 2, it is the number of those courses that were completed by the end of the academic year; in column 3, it is the number of credits granted during the academic year. Each of those variables is set to zero when the student is not enrolled in a public university. "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Two-stage least squares with "Above discontinuity" as an instrument. No controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. For each gender and outcome variable, the sample is restricted to the optimal bandwidth sample based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

Table 10
Means of predetermined and outcome variables, by gender.

	Full sample (1)	Optimal bandwidth sample (2)	Compliers (3)
	Women		
Age	26.92	27.54	28.17
Student is the borrower	0.44	0.46	0.55
Requested loan value	6942	6748	5008
Monthly wage	6828	6809	4788
Enrollment in public university	0.66	0.61	.
Nb courses completed (if enrolled)	4.78	4.44	.
Nb courses registered (if enrolled)	7.15	6.99	.
Credits completed (if enrolled)	0.48	0.46	.
Number of obs.	5123	876	876
	Men		
Age	28.55	29.30	24.59
Student is the borrower	0.54	0.59	0.65
Requested loan value	6547	6504	5017
Monthly wage	7197	6944	6995
Enrollment in public university	0.62	0.53	.
Nb courses completed (if enrolled)	4.03	4.05	.
Nb courses registered (if enrolled)	6.71	6.56	.
Credits completed (if enrolled)	0.41	0.41	.
Number of obs.	4532	896	896

Notes: Eduloan and HEMIS data, restricted to loan applications to a public university from 2004 to 2007; the sample is split by the gender of the applicant student. The unit of observation is loan request per student per year; when several applications have been sent for a given student in the same year, we use the average requested loan value. Applications dated November/December are excluded, as in all baseline specifications. The optimal bandwidth samples are those used in the baseline specifications (see Table 7); complier means are estimated on those same samples, see Appendix B.

those variables is set to zero for the non-enrolled and takes the reported positive value for the enrolled (we cannot identify the impact of having a loan on educational outcomes *conditional on enrollment*[22]).

Mechanically, because they enroll more frequently, female applicants who get a loan tend to register for more courses on average (5.2 more courses, from 1.4 among non-treated compliers); because men do not enroll more frequently, the loan effects are all small and not significant. It might be posited that the marginal individuals who are induced to enroll would in fact fail massively, so that loan access does not translate into increased completion. Yet, this is far from the case: loan access increases women's number of courses completed and, as a result, also their number of credits. They pass 3.4 more courses and earn .26 more credits, which represents about a quarter of a full-time year; it raises credits in this population to about 0.4. Given that not all women granted a loan actually enroll and given also the significant amount of dropouts and failures in the South African higher education system, this appears to be a noticeable impact from a policy perspective.

6. Interpretations

6.1. Interpretations of gender differences

Why is the credit constraint only apparent for women? One possibility could be that female applicants have different characteristics than male applicants, which are confounded with gender. Let us first consider this interpretation.

Table 10 contrasts male and female characteristics (1) in our sample of applicants for public universities, (2) in the subsample used for optimal estimation, and (3) in the complier population (the latter only for predetermined variables). Generally, men are older, they are more often borrowing for themselves (as opposed to parents borrowing for their child), but the borrower for men has only slightly higher wages. Those differences are similar in the full sample and in the optimal bandwidth sample.[23]

Specifically, the fact that women are less often borrowers is unlikely to explain much of the gender contrast: Table A.6 shows that the impact of a loan is slightly higher for female students when they are not the borrower than when they are. But the effects in both groups are much smaller and not statistically significant for male students. Thus, the stronger effect on women does not just hide a composition effect driven by borrower status.

When looking at complier populations, a stronger contrast becomes apparent: the borrower now has much higher wages for male than for female compliers (+46% on average). Also (Table 7), the counterfactual enrollment rate in the absence of a loan for the male compliers (0.60) is much higher than for the female compliers (0.38).[24] This seems to point to the fact that, although women are not poorer than men in the *full* sample, the marginal women who take out a loan are mostly from the lower income group. However, just as for borrower status, it is important to note that *the overall male–female contrast in loan impact is not driven by an income composition effect*: we have seen in Table 8 that the loan impact is higher for women than for men even *within* wage groups. Of course, the stronger overall female effect is driven by the poorest of them, who are overrepresented in the complier group, but the same impacts would not be found among men of similar income.

[22] 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 those people may be different in terms of academic capacity or motivation. This is the usual selectivity problem, as faced by Canton and Blom (2009) 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.

[23] We have already noted that the full sample and the optimal bandwidth sample do not differ much in the baseline estimation — see Table A.2.

[24] See Appendix B for the computation of those counterfactual outcomes on compliers.

We thus believe that our findings most likely reflect a genuine gender contrast. As emphasized in Section 3.1, we do not directly assess if there is a credit constraint, but rather if this constraint is binding in terms of educational decisions. Therefore, what we observe (enrollment) is the outcome of a combination of credit constraint and preferences. Women and men can thus differ in either dimension. Therefore, we can think of two polar (but not exclusive) possible differences between them:

1 They face similar credit constraints, but women have either lower returns to education or higher costs of being credit-constrained, so that a given limit on their credit has more impact on enrollment decisions for them.
2 They have similar utility values of schooling and debt, but in the absence of an Eduloan loan, they face very different credit constraints even with similar borrower wages;

In terms of our conceptual framework, the available amount of debt is $\bar{d}$ in the absence of an Eduloan loan, and $\bar{d}' \geq \bar{d}$ with such a loan. Students with $V(\bar{d}) > V_L$ enroll without Eduloan: according to Table 7, this is estimated at 60% for male compliers and 38% for female compliers. An Eduloan loan *causes* enrollment when $V(\bar{d}) < V_L$ *and* $V(\bar{d}') > V_L$: this hardly happens for men but is true for 47.8% of women.

In the first polar interpretation above, men and women face similar distributions of $\bar{d}$ and $\bar{d}'$, but $V(.)$ and V_L are such that it is much more frequent for men that $V(\bar{d}) > V_L$ because, for example, their returns to schooling are higher, or the utility cost of living with only debt $\bar{d}$ is lower for them. Effects via returns to education are not likely: Salisbury (2016) estimates higher wage returns to education in South Africa for women than for men; and although this is estimated conditional on working, his data also implies that higher education has a greater positive impact on the labor market participation of women.

More likely, men who face limited borrowing options can more easily compensate by working while studying, for instance, or by decreasing their consumption; this may be less easy for women, for example because they have to bear more domestic responsibility, so that relaxing the constraint from $\bar{d}$ to $\bar{d}'$ makes a much greater difference in utility for women than for men, and changes their decision toward more enrollment.

The second polar interpretation as mentioned above is that $V(.)$, the value of higher education as a function of debt, and V_L, the value of not going to university, are distributed similarly between men and women; but men access a higher level of liquidity without the loan, $\bar{d}$, so that for them $V(\bar{d})$ is often above V_L: they thus have high counterfactual enrollment rates. And $V(\bar{d}')$ is close to $V(\bar{d})$, so the loan impact on enrollment is negligible. For women, $\bar{d}$ is much lower in this interpretation, they have low enrollment rates without Eduloan, and the loan impact on enrollment is high. This story also fits the main findings.

There are two main reasons why women may be more credit-constrained: they are more rationed by banks; or, when parents are the borrower, parents are more likely to pay for boys' education "out of their pocket", whereas they would only support girls if a loan is available. Table A.6 provides hints on the latter interpretation: if parents were more reluctant to pay for girls' education in the absence of a loan, enrollment rates of women relative to men without a loan would be lower when the student is not the borrower than when she is. Looking at the counterfactual enrollment rates ($E(y_0|Comp)$), this is not what we observe: the counterfactual enrollment is similar for women, whether the borrower is the student or not.

Women may thus be more rationed by banks. This could be out of pure discrimination or because banks anticipate lower repayment capacity. We have mentioned above that women's returns to education are higher than men's, both in terms of wages and participation. Further, the data also show that their graduation rates are much higher than men's.[25] Therefore, one could expect that financing women's higher education is not particularly risky for banks. Pure discrimination is a more likely possibility.

A last segment of the population of both men and women calls for comment. Among applicants to Eduloan, there remain an estimated 14% of women and 35% of men who do not enroll in public higher education despite asking for and obtaining a loan.[26] There may be several explanations for this: their ex ante plan was conditional on other expected circumstances than just the loan, and those circumstances did not happen; they enrolled in a private institution (this issue is discussed below); or they are very early dropouts, and thus do not appear in the HEMIS data.[27] The third interpretation is quite likely given the low rates of graduation and the high dropout rates in South African higher education. In that perspective, it is not surprising that this is a rarer event for women: consistent with that, Table 10 illustrates that, conditional on enrollment, women register for more courses and complete more courses than men.

Finally, a striking difference between men and women is the first stage impact of the Empirica score: passing the threshold increases loan access much more for women (+42 points) than for men (+24 points) (Table 7). This implies that the information contained in the Empirica score is complementary to the characteristics of women as potential repayers of a loan. We can hint that such a complementarity exists in general: for this, we regress a dummy for being granted a loan on age, borrower monthly wage, requested loan value, and whether the student is the borrower (but not gender). From this regression, we can form a propensity score for obtaining a loan: it measures how confident Eduloan is that a given subject will repay the loan. If we run the first stage regression on individuals below this propensity score median, we find that passing the Empirica threshold increases loan access by 22 points (p-value = 0.000); but for individuals above the propensity score median, it increases loan access by 44 points (p-value = 0.000). Therefore, the information contained in the Empirica score seems to be complementary with the information already available. One possibility is therefore that women are perceived to be more reliable repayers. We do not have a direct measure of this, but two facts make it plausible: they graduate more often, and they have higher wage returns to schooling as well as a larger effect of higher education on participation. As such, Eduloan would be more likely to grant loans to women, but only once it is reassured by the Empirica score information.

6.2. Longer-term impacts

So far, we have provided robust evidence that, for women, access to credit (or lack thereof) actually distorts enrollment decisions. But the very consequences can be discussed further. For instance, students deprived of a loan may enter private institutions if they are less expensive: as we do not observe enrollment in such institutions in our data, enrollment in *some* university may be higher than we estimate in the control group. We will discuss this in the next section. Another possibility is that students without a loan "only" delay enrollment: they may, for instance, work to save money and enroll later. In that case, enrollment in the control group would be higher after some period of time than during the year considered, and in the longer run, the loan impact would be smaller than assumed so far.

[25] The number of female graduates in 2020 is 1.7 times higher than male graduates (Department of Higher Education and Training, 2022), whereas the proportion enrolled in higher education four years earlier is only 1.4 times higher (Department of Higher Education and Training, 2019).

[26] Those never-takers are counted as the residual of those who enroll anyway (60% and 38%) plus those who enroll when granted a loan (5% and 48%) among compliers.

[27] HEMIS data are primarily meant to govern the amount of public subsidy, so they reflect actual enrollment during the year and exclude early dropouts.

Table 11
Current and following years enrollment as a function of loan obtention, women (baseline optimal bandwidth samples).

	Applicants at t...		
	... enrolled at t or $t+1$	... enrolled at $t+1$ but not at t	... enrolled at t or $t+1$ or $t+2$
	(1)	(2)	(3)
Loan granted	0.461 (0.155)	−0.017 (0.086)	0.532 (0.210)
$E(y_0\|Comp)$	0.388	0.007	0.375
Number of obs.	876	876	559

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 in columns 1 and 2, and from 2004 to 2006 in column 3 (applications dated November/December excluded). The explained variable in column 1 is a dummy for enrollment in a public university the same year or the year after a loan application was received; the explained variable in column 2 is a dummy for enrollment in a public university the year after a loan application was received, but not the same year; the explained variable in column 3 is a dummy for enrollment in a public university the same year or the year after or two years after a loan application was received. "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Two-stage least squares with no controls other than linear functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. The instrument for 2SLS is a dummy for when the Empirica score is above zero. For comparison purposes, the bandwidth used is the optimal bandwidth in Table 7, column 6. $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

If this was the case, enrollment in year $t+1$ among the non-treated would compensate for the lag in t and there would be a smaller (or no) impact of enrolling either in t or in $t+1$. To figure this out, Y being the dummy for enrollment and T the dummy for loan, we could estimate:

$$P(Y_t = 1 \text{ or } Y_{t+1} = 1|T_t = 1) - P(Y_t = 1 \text{ or } Y_{t+1} = 1|T_t = 0) \quad (6)$$

This can be identified based on the regression discontinuity as before. Notice that we can rewrite this parameter as:

$$\begin{aligned} &P(Y_t = 1|T_t = 1) - P(Y_t = 1|T_t = 0) \\ +\ &P(Y_t = 0 \text{ and } Y_{t+1} = 1|T_t = 1) - P(Y_t = 0 \text{ and } Y_{t+1} = 1|T_t = 0) \quad (7) \end{aligned}$$

The first term is the short-term treatment effect estimated so far. The second term measures the treatment impact on the share who do not enroll in t but do enroll later on. We test here if this second impact is negative.

The results for women are presented in Table 11.[28] Column (1) estimates Eq. (6); for comparison purposes, we report estimates at the optimal bandwidth (±13) of the baseline estimation for women.[29] Receiving a loan increases the probability of enrolling either in the current year or in the following year by 46.1 percentage points. This is hardly different from the impact of a loan on enrolling in the current year only, which is 47.8 percentage points (Table 7). As a matter of fact, column (2) of Table 11, which estimates the second line of Eq. (7), indicates only a small and insignificant negative impact of receiving a loan on waiting a year before enrolling (and these are very rare events in the absence of a loan). In the last column, we extend the time window to an additional year, and, if anything, the coefficients are higher (but the difference is not significant). As a result, there is no evidence that not obtaining a loan simply delays enrollment for women: it does decrease final enrollment, at least in a two- or three-year window.

6.3. Enrollment in the private sector

We have shown that, when a woman plans to enter a public university and asks Eduloan for a short-term loan to pay the fees, she is more likely to enroll in a public university when the loan is granted. We cannot strictly exclude that an individual whose request is turned down may decide to enroll in the private sector instead, because our data contain no information on private enrollment.[30] 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 in one way or another because they did not obtain short-term credit to pay the fees for 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. As mentioned earlier, the private higher education sector is small but not negligible (8% of enrollment). If private institutions are less expensive than public universities,[31] 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 escape the liquidity constraint and the quality is sufficiently high to make this choice a second best. If such behavior (unobserved by us) were present, this would reduce the loan impact *in terms of overall enrollment in higher education*.

We cannot directly measure this, but we have a way to check whether individuals turned away by Eduloan tend to choose a less costly university instead. South Africa has a famous distance learning institution, which was open to Black Africans and Colored people under apartheid: the University of South Africa (UNISA). In our data, 31% of all loan requests for a public university are made for UNISA. Its lower cost is reflected in the size of the loans requested: the average loan request is ZAR 7431 for other public universities but only ZAR 4142 for UNISA. Table 12 examines women who requested a loan for a public university *other than* UNISA. It checks whether those who were refused a loan eventually enrolled 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 shifting to a less costly institution were optimal for many individuals when a loan for a public university is refused, then we would

[28] The results for men all remain small and not statistically significant.

[29] Although we add a period, the sample remains that of applicants between 2004 and 2007 as in earlier tables (we do not use 2008 for loan applications, but we do observe enrollment for that year).

[30] 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.

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

Table 12
Enrollment at UNISA as a function of loan obtention, when applicants did not ask for a loan there, women.

	Full sample		Bandwidth (2SLS)				
	OLS	2SLS	±50	±30	±15	±10	±**16**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan granted	0.005	−0.045	−0.028	−0.039	−0.015	−0.026	−0.010
	(0.007)	(0.063)	(0.073)	(0.102)	(0.097)	(0.128)	(0.091)
Intercept	0.038	0.056	0.061	0.060	0.049	0.073	0.047
	(0.008)	(0.024)	(0.027)	(0.037)	(0.035)	(0.042)	(0.033)
$E(y_0\|Comp)$		0.062	0.044	0.046	0.016	0.069	0.007
Linear in Empirica					x	x	x
Quadratic in Empirica	x	x	x	x			
Number of obs.	3651	3651	1878	1283	703	487	742

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university *except UNISA*, from 2004 to 2007 (applications dated November/December excluded) and female applicant students. The explained variable is a dummy for enrollment *at UNISA* the same year a loan application was received; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Ordinary least squares estimation (column 1) and two-stage least squares (columns 2–7) with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. The instrument for 2SLS is a dummy for when the Empirica score is above zero. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

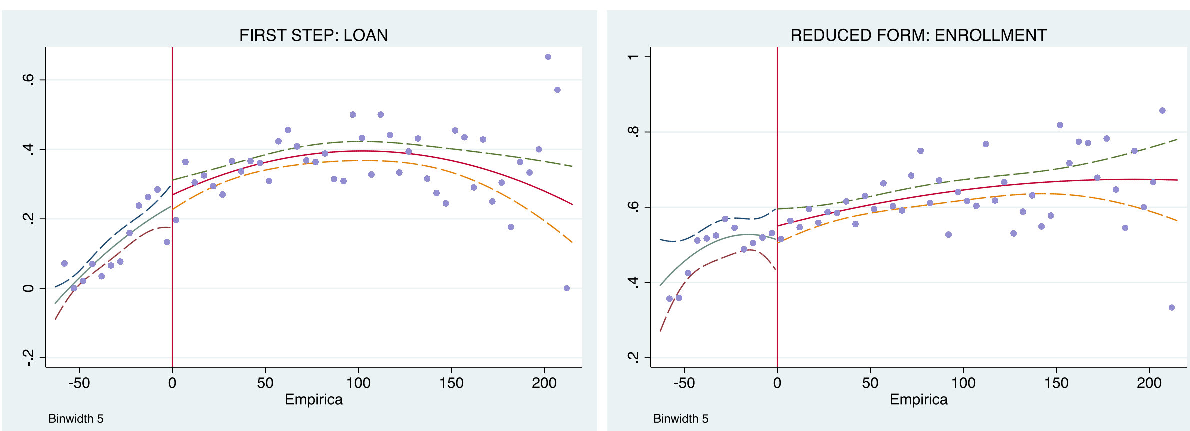

Fig. 4. Loan and university enrollment as a function of Empirica score in 2008 (quadratic fit on full sample with 95% confidence intervals). *Notes:* Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded). In the left-hand panel (resp. right-hand panel), each dot represents the proportion of loans granted (resp. the proportion of loan applicants that enrolled in a public university) the same year a loan application was received, within bins of 5 Empirica points. The quadratic fits use the estimation in columns 2 and 5 of Table 13 respectively.

expect at least some of them to shift 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.

7. Robustness

7.1. Placebo: The 2008 credit crunch

In 2008, the financial crisis led to 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 the left-hand panel of Fig. 4, which presents the probability of receiving a loan as a function of the Empirica score in 2008. This is in strong contrast to the data from 2004 to 2007 that is used in the rest of the paper. It thus provides a placebo test: as passing the threshold no longer increases loan access discontinuously in this data, it should not increase enrollment in higher education either.

The right-hand panel of Fig. 4 illustrates that there is indeed no increase in enrollment when there is no increase in loan access. Table 13 estimates the full models, separately for men and women on the 2008 data, using the optimal bandwidths. It confirms that the identification strategy passes this placebo test.

7.2. Sample variants

The sample used until now has been restricted to loans requested to pay public university fees, but only when information on the kind of university was actually available. There are 2509 observations in 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.

Appendix C shows formally that pooling public and non-public loan requests will provide an average of: (1) the true effect on HEMIS

Table 13
University enrollment as a function of loan obtention in 2008, by gender (optimal bandwidth samples).

	Men			Women		
	First stage (Loan) (1)	Reduced form (Enrollment) (2)	2SLS (3)	First stage (Loan) (4)	Reduced form (Enrollment) (5)	2SLS (6)
Above discontinuity	−0.029 (0.101)	−0.015 (0.130)		0.170 (0.101)	0.006 (0.130)	
Loan granted			0.516 (4.376)			0.033 (0.640)
Intercept	0.171 (0.077)	0.443 (0.098)	0.355 (0.674)	0.097 (0.051)	0.619 (0.076)	0.616 (0.127)
Number of obs.	244	244	244	342	342	342

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university in 2008 (applications dated November/December excluded); the sample is split by the gender of the applicant student. In columns 1 and 4, the explained variable is a dummy for a loan being granted the same year a loan application was received. In the other columns, the explained variable is a dummy for enrollment in a public university the same year a loan application was received. "Above discontinuity" is a dummy for when the Empirica score is above zero; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. The models in columns 1, 2, 4, and 5 are estimated by ordinary least squares, and the models in columns 3 and 6 are estimated by two-stage least squares with "Above discontinuity" as an instrument. No controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. For each gender, the sample is restricted to the optimal bandwidth sample for the reduced form based on Calonico et al. (2014).

Table 14
University enrollment as a function of loan obtention, by gender (optimal bandwidth samples).

	A: Applicants to public university and unknown university	
	Women	Men
Loan granted	0.348 (0.170)	−0.108 (0.288)
Intercept	0.467 (0.062)	0.562 (0.069)
Number of obs.	1238	1478
	B: Applicants to public university including Nov. and Dec. applications	
	Women	Men
Loan granted	0.510 (0.164)	0.144 (0.209)
Intercept	0.415 (0.065)	0.489 (0.063)
Number of obs.	908	1102

Notes: Eduloan and HEMIS data, restricted to loan applications from 2004 to 2007. In panel A, the sample is restricted to applications for a public university or for a university that is not identified in the data, and applications dated November/December are excluded. In Panel B, the sample is restricted to loan applications for a public university, but applications dated November/December are included. Each sample is split by the gender of the applicant student. The explained variable is a dummy for enrollment in a public university the same year a loan application was received; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Two-stage least squares with no controls other than linear functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. The instrument for 2SLS is a dummy for when the Empirica score is above zero. Each estimation uses the optimal bandwidth sample based on Calonico et al. (2014).

perimeter demands and (2) a zero effect (given that loan access has no causal effect on enrollment in a public university for those who asked for a loan for a private institution); thus a lower bound to the true effect.

Table 14, Panel A, estimates the impact of loan obtention on enrollment in public universities, pooling the sample of known HEMIS applicants and applicants to an undetermined university. For women, it shows an effect of 0.348, which is a lower bound to our baseline impact of 0.478 (see Table 7). We are thus confident of the presence of an impact and its order of magnitude for women. The impact is still close to zero for men.

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 coming year. The sample change is rather marginal, as the number of observations is increased by only 14% 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 outcome variable will not be sensitive to loan access in the next year. Table 14, Panel B, shows that the coefficient for women (0.510) is in fact slightly higher than our baseline estimate, but notice that point estimates are using different optimal bandwidths across samples.

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

8. Conclusion

This paper provides simple and robust evidence that giving loans to potential female students who apply for them strongly affects their actual enrollment in higher education. It implies that in the absence of the scheme offered by Eduloan, the borrowing constraint in the South African economy would be strongly binding for that class of students. This is compatible with earlier findings by Solis (2017) in a comparable context, although he does not point to the gender heterogeneity.

This strong impact seems at odds with much of the literature. One important difference between our findings and the mostly US-based evidence, apart from methodology, is that either credit markets for human capital investment are more present (as analyzed by Lochner and Monge-Naranjo, 2011) or the large range of subsidies to education that exist in the US compensate more for credit market constraints than they do in developing economies. In our context, the poorest are covered by the NSFAS program, but middle-class students may face strong constraints. 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 few so far to fill the gap.

Table A.1
Type of studies in our sample and in the South African higher education system.

	Our sample	South Africa
Business and Management	31%	29%
STEM	28%	28%
Education, Humanities, and Social Sciences	39%	43%
Unreported	2%	.
Undergrad certificates	28%	34%
Undergrad degrees	56%	48%
Postgrad	7%	8%
Master/Doctorate	4%	7%
Occasional	4%	3%
Technikons	19%	18%
UNISA	32%	33%

Notes: Our sample: Eduloan and HEMIS data, restricted to loan applications enrolled in a public university, 2004 to 2007. South African higher education system: Department of Education (2010).

Table A.2
Descriptive statistics on loan requests, full and estimation sample, 2004–2007.

	Full baseline sample		Optimal bandwidth sample	
	Mean	S.d.	Mean	S.d.
Male	0.47	0.50	0.44	0.50
Age	27.68	8.14	28.53	8.48
Monthly wage	7001	6418	6967	5111
Missing wage information	0.08	0.26	0.12	0.32
Requested loan value	6756	5405	6642	5488
Requested loan/monthly wage	1.27	1.29	1.27	1.49
Missing requested loan value	0.02	0.13	0.03	0.17
Student is the borrower	0.49	0.50	0.53	0.50
Empirica	51.05	61.52	0.75	6.49
Enrollment in public university	0.64	0.48	0.57	0.50
Credits completed (if enrolled)	0.45	0.37	0.43	0.36
Nb courses registered (if enrolled)	6.95	4.24	6.73	4.21
Number of obs.	9655		1331	

Notes: Eduloan and HEMIS data, restricted to loan applications to a public university, from 2004 to 2007. The unit of observation is loan request per student per year; when several applications have been sent for a given student in the same year, we use the average requested loan value. Applications dated November/December are excluded, as in all baseline specifications. The optimal bandwidth sample is restricted to observations with an Empirica score between −11 and +11, which is the optimal bandwidth for the baseline reduced form estimation (see Table 5).

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. However, such a policy should be considered with caution in view of the increasing student debt issue in the US and may necessitate some form of insurance scheme, which is not analyzed here.

Finally, the striking gender difference that is found in our context has not been documented either in the US or in developing countries, and understanding its origin will require additional research in the future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

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 for their work on the field; and to Jean Skene and Dr. Qhobela from the Department of Education (DoE) for providing us with the HEMIS data. Thanks to Jocelyn Vass, Fabian Arends, Mariette Visser from HSRC for very valuable discussions and advice, and to Geneviève Michaud for assistance with the data. Special thanks also to the AFD team in South Africa. We are grateful to Clément de Chaisemartin, Eric Maurin, Simon Jaeger, the Editor, Aprajit Mahajan, two anonymous referees, and seminar participants at Pompeu Fabra University, Cape Town University, Columbia University, the University of Bergen, the Paris School of Economics, the European Development Network Workshop, EALE/SOLE, and the Econometric Society World Congress for their useful comments. The authors gratefully acknowledge the financial support of the Agence française de développement (AFD), France. The views expressed herein are those of the authors only.

Appendix A. Tables and figures

See Tables A.1–A.6 and Figs. A.1 and A.2.

Appendix B. Computation of complier characteristics

$Y(1)$ is counterfactual enrollment in a public university when a loan is granted and $Y(0)$ when it is not. E is the Empirica score, E_0 being the identifying threshold, and $D = 1$ if $(E \geq E_0)$. $T = 1$ when a loan is granted (treatment). 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. Adapting Abadie (2002) to the regression discontinuity design, we have

$$E(Y(0)|C) = \frac{E^+[(1-T)Y|E_0] - E^-[(1-T)Y|E_0]}{-P_C}$$

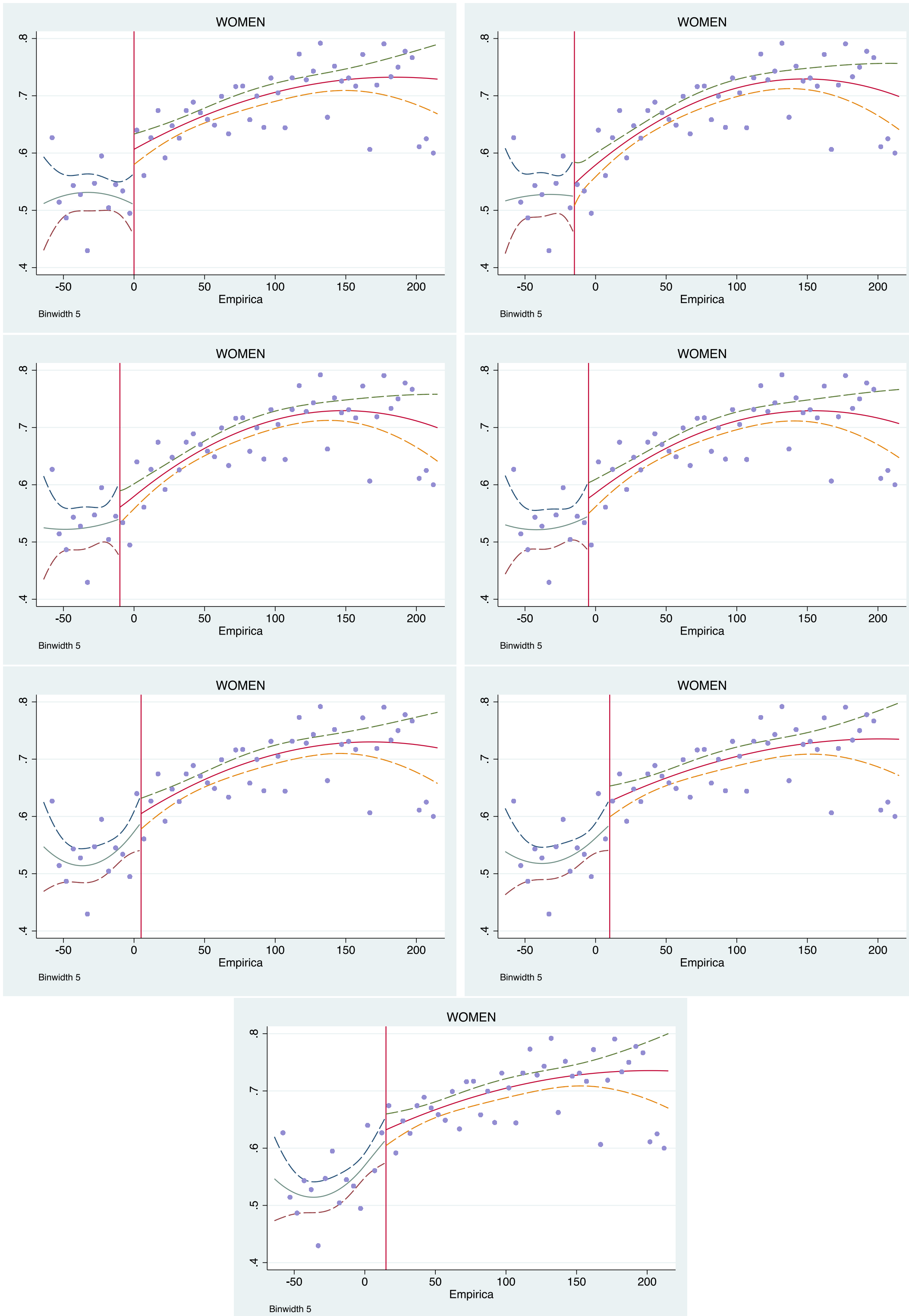

Fig. A.1. Share of university enrollment as a function of Empirica, women: Placebo cutoffs (the ''true'' cutoff is zero). *Notes:* Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded), separately by gender of the applicant student, quadratic fit. Each dot represents the proportion of loan applicants that enrolled in a public university the same year a loan application was received, within bins of 5 Empirica points.

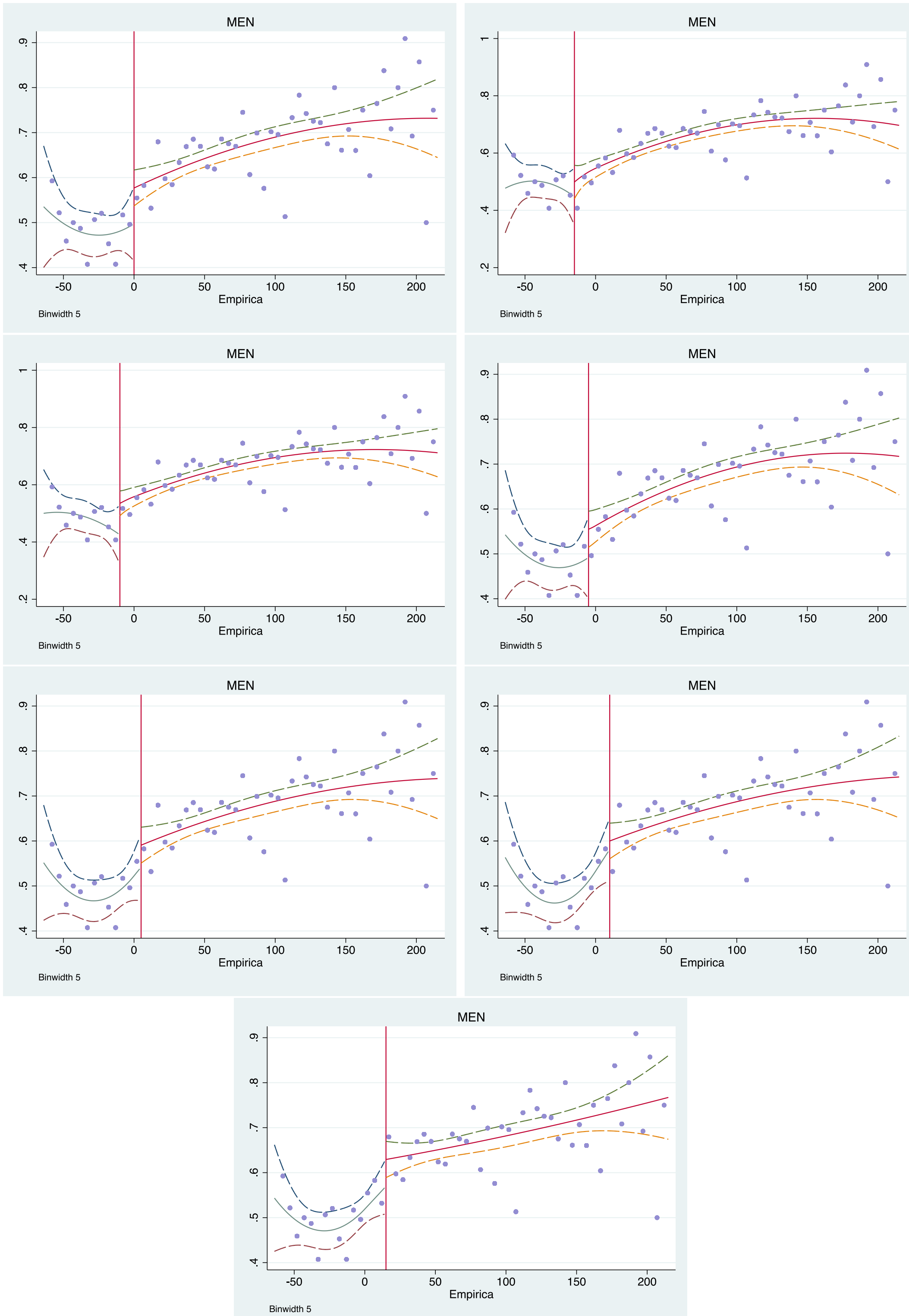

Fig. A.2. Share of university enrollment as a function of Empirica, men: Placebo cutoffs (the "true" cutoff is zero). *Notes:* Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded), separately by gender of the applicant student, quadratic fit. Each dot represents the proportion of loan applicants that enrolled in a public university the same year a loan application was received, within bins of 5 Empirica points.

Table A.3
Placebo: Predetermined variables as a function of Empirica score, by gender.

	Coefficient on discontinuity variable (1)	Optimal bandwidth (2)	Number of obs. (3)
	Women		
Applied public U.	0.064 (0.047)	±14	1202
Age	−1.047 (1.140)	±13	876
Student is the borrower	−0.048 (0.064)	±15	997
Requested loan value	−67.180 (577.450)	±20	1258
Monthly wage	1053.577 (709.816)	±13	780
Propensity score	0.018 (0.014)	±11	650
	Men		
Applied public U.	−0.018 (0.050)	±17	1175
Age	0.740 (1.095)	±18	948
Student is the borrower	−0.034 (0.064)	±18	948
Requested loan value	516.610 (811.760)	±15	782
Monthly wage	621.625 (903.093)	±12	557
Propensity score	−0.009 (0.010)	±13	588

Notes: Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. Each line is a separate ordinary least squares regression, the explained variable of which is given on the left-hand side. Column 1 gives the coefficient (robust standard errors in parentheses) of that regression on a dummy for when the Empirica score is above zero. There are no controls other than linear functions of the Empirica score (different on each side of the discontinuity). Each model is fit on its optimal bandwidth based on Calonico et al. (2014): the optimal bandwidth, defined with respect to the Empirica score, is given in column (2) and the corresponding number of observations in column (3). Propensity score is an index of the variables Male to Monthly wage, built from a regression of obtaining a loan on those variables.

Table A.4
Test of equality of Empirica densities on each side of the threshold, by gender.

Order polynomial for CDF specification (1)	Test statistic (2)	*p*-value (3)	Number of obs. (4)
	Women		
1	−1.786	0.074	692
2	1.638	0.101	3019
3	1.278	0.201	3478
4	1.184	0.237	4024
5	−0.546	0.585	4471
	Men		
1	−1.285	0.199	881
2	−0.824	0.410	1911
3	−1.338	0.181	2533
4	0.139	0.890	3801
5	−1.092	0.275	3375

Notes: Eduloan data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. This table tests the hypothesis that the density of the Empirica score is continuous at the cutoff point, using Cattaneo et al. (2018, 2020). The density estimation is run separately on each side of the discontinuity and uses a local polynomial approximation of the cumulative distribution function. We run tests for different orders of that approximation, and each line is a different test: column 1 gives the order used for each test run separately. Columns 2 and 3 show the test statistic and its *p*-value respectively (the test is positive when the estimated density on the right-hand side is higher). The local approximation is run on a bandwidth based on the mean square error of each density separately: column 4 gives the number of observations used, given the optimal bandwidth for each test.

Table A.5
University enrollment as a function of loan obtention by gender.

	Women						
	Full sample		Bandwidth (2SLS)				
	OLS	2SLS	±50	±30	±15	±10	**±13**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan granted	0.205	0.289	0.405	0.520	0.763	0.600	0.478
	(0.015)	(0.108)	(0.129)	(0.164)	(0.261)	(0.191)	(0.158)
Intercept	0.529	0.496	0.464	0.406	0.378	0.416	0.415
	(0.017)	(0.046)	(0.052)	(0.070)	(0.112)	(0.078)	(0.065)
$E(y_0\|Comp)$		0.511	0.440	0.335	0.253	0.302	0.381
Linear in Empirica					x	x	x
Quadratic in Empirica	x	x	x	x			
Number of obs.	5123	5123	2702	1836	997	691	876
	Men						
	Full sample		Bandwidth (2SLS)				
	OLS	2SLS	±50	±30	±15	±10	**±17**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan granted	0.200	0.248	0.198	0.127	0.140	0.219	0.052
	(0.016)	(0.145)	(0.202)	(0.336)	(0.262)	(0.284)	(0.278)
Intercept	0.491	0.474	0.471	0.499	0.510	0.456	0.526
	(0.019)	(0.055)	(0.063)	(0.085)	(0.073)	(0.081)	(0.073)
$E(y_0\|Comp)$		0.521	0.470	0.512	0.558	0.479	0.600
Linear in Empirica						x	x
Quadratic in Empirica	x	x	x	x	x		
Number of obs.	4532	4532	2281	1500	801	534	896

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. The explained variable is a dummy for enrollment in a public university the same year a loan application was received; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. Ordinary least squares estimation (column 1) and two-stage least squares (columns 2–7) with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. The instrument for 2SLS is a dummy for when the Empirica score is above zero. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

where C denotes compliers and P_C their proportion in the population. However, because there are two multiplicative discontinuities in a regression of $(1-T)Y$ on the whole sample, the local linear regression is not well specified to approximate the numerator. For a simpler approach, note that $E^+[(1-T)Y|E_0] = E^+[(1-T)Y|T=0,E_0]P_N$ with P_N the proportion of never-takers, because only never-takers are untreated on the right-hand side of the discontinuity, and $(1-T)Y = 0$ for the others. Similarly, $E^-[(1-T)Y|E_0] = E^-[(1-T)Y|T=0,E_0](P_C+P_N)$ because compliers are also untreated on the left-hand side. We can thus rewrite:

$$E(Y(0)|C) = -E^+[(1-T)Y|T=0,E_0]\frac{P_N}{P_C} + E^-[(1-T)Y|T=0,E_0]\frac{P_C+P_N}{P_C}$$

We can use a discontinuity model of the form $Y = g(E) + \delta D + u$ estimated on the untreated only ($T=0$) using local linear regression. In this regression, $g(0)$ identifies $E^-[(1-T)Y|T=0,E_0]$ and δ identifies $E^+[(1-T)Y|T=0,E_0] - E^-[(1-T)Y|T=0,E_0]$. P_C and P_N are directly recovered from the first-stage discontinuity regression. The estimation of $E(Y(0)|C)$ follows. Of course, this uses the independence assumption that $Y(0)$ is orthogonal to D conditional on E, which is discussed at length in the paper.

To estimate the average characteristics X of the compliers, we can make use of the fact that predetermined X's can also be assumed orthogonal to D conditional on E, just as much as $Y(0)$. This hypothesis has been tested in Table 3. Therefore, we can apply the same strategy to estimate $E(X|C)$.

The intuition for this estimation is that Y for $T=0$ is always $Y(0)$, which is assumed continuous in E. Therefore, any discontinuity at D in the Y of the untreated must come from the fact that the population shifts suddenly from compliers+never-takers to never-takers only. If the $Y(0)$ is different between compliers and never-takers, this will be captured by δ. The same intuition holds for the characteristics X that need to change at the threshold, if populations become different.

Appendix C. 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 in a public university 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 requested 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 in part of the sample. 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 et al. (2001), we describe the regression discontinuity estimator as a Wald estimator.

When H is not observed (we then pool HEMIS and non-HEMIS demands), the Wald estimator we can compute is:

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

where Y is observed outcome and T is observed loan status (obtained or not).

Table A.6
University enrollment as a function of loan obtention by borrower status (optimal bandwidth samples).

	Women					
	Student is borrower			Student is not borrower		
	First stage (Loan) (1)	Reduced form (Enrollment) (2)	2SLS (3)	First stage (Loan) (4)	Reduced form (Enrollment) (5)	2SLS (6)
Above discontinuity	0.336 (0.063)	0.122 (0.079)		0.458 (0.078)	0.223 (0.094)	
Loan granted			0.363 (0.225)			0.488 (0.218)
Intercept	0.114 (0.033)	0.508 (0.060)	0.467 (0.080)	0.107 (0.048)	0.487 (0.076)	0.435 (0.100)
$E(y_0\|Comp)$			0.432			0.427
Number of obs.	616	616	616	439	439	439
	Men					
	Student is borrower			Student is not borrower		
	First stage (Loan) (1)	Reduced form (Enrollment) (2)	2SLS (3)	First stage (Loan) (4)	Reduced form (Enrollment) (5)	2SLS (6)
Above discontinuity	0.252 (0.067)	0.011 (0.091)		0.242 (0.080)	0.027 (0.099)	
Loan granted			0.042 (0.360)			0.112 (0.410)
Intercept	0.079 (0.031)	0.497 (0.068)	0.493 (0.091)	0.112 (0.046)	0.574 (0.076)	0.561 (0.115)
$E(y_0\|Comp)$			0.655			0.474
Number of obs.	498	498	498	370	370	370

Notes: Eduloan and HEMIS data, restricted to loan applications for a public university, from 2004 to 2007 (applications dated November/December excluded); the sample is split by the gender of the applicant student. In columns 1 and 4, the explained variable is a dummy for a loan being granted the same year a loan application was received. In the other columns, the explained variable is a dummy for enrollment in a public university the same year a loan application was received. "Above discontinuity" is a dummy for when the Empirica score is above zero; "Loan granted" is a dummy for a loan being granted the same year a loan application was received. The models in columns 1, 2, 4, and 5 are estimated by ordinary least squares, and the models in columns 3 and 6 are estimated by two-stage least squares with "Above discontinuity" as an instrument. No controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. For each gender, the sample is restricted to the optimal bandwidth sample for the reduced form based on Calonico et al. (2014). $E(y_0|Comp)$ estimates the mean counterfactual enrollment in the absence of a loan in the population of compliers, see Appendix B.

We have:

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

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

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

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

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

However, as Y measures enrollment in a public university, we expect that passing the E_0 threshold for individuals that applied to a private university will not affect enrollment in a public university. This is all the more likely given that fee payments are delivered directly by Eduloan to the university; but it could still be the case that applicants who did not receive a loan for a private university are more likely to move to a public one. Table C.1 tests this on the subset of applicants for which we know that $H=0$. It shows that, at least in that sample, we cannot reject $E^+[Y|E_0,H=0] = E^-[Y|E_0,H=0]$.[32] Under this condition, it is straightforward to show that W is a lower bound to the parameter of interest. Indeed, we then have:

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

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

In addition:

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

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

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

Replacing:

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

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

The term within brackets is clearly positive and higher than 1, so

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

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

[32] It is a fact that a small share of individuals who applied for a loan for a private university end up enrolled in a public university. Table C.1 shows that this is unrelated to loan status.

Table C.1
Public university enrollment as a function of Empirica score, applicants to a private university, by gender.

	Women					
	Full sample	Bandwidth				
		±50	±30	±15	±10	±**16**
	(1)	(2)	(3)	(4)	(5)	(6)
Above discontinuity	−0.000 (0.002)	0.002 (0.004)	−0.013 (0.009)	−0.005 (0.006)	−0.001 (0.010)	−0.005 (0.005)
Intercept	0.145 (0.024)	0.175 (0.036)	0.108 (0.046)	0.128 (0.043)	0.142 (0.053)	0.135 (0.042)
Linear in Empirica				x	x	x
Quadratic in Empirica	x	x	x			
Number of obs.	1236	751	495	273	174	288
	Men					
	Full sample	Bandwidth				
		±50	±30	±15	±10	±**15**
	(1)	(2)	(3)	(4)	(5)	(6)
Above discontinuity	0.003 (0.002)	0.009 (0.004)	0.009 (0.008)	0.002 (0.006)	0.019 (0.007)	0.002 (0.006)
Intercept	0.145 (0.026)	0.204 (0.041)	0.180 (0.052)	0.155 (0.051)	0.213 (0.058)	0.155 (0.051)
Linear in Empirica				x	x	x
Quadratic in Empirica	x	x	x			
Number of obs.	1109	680	459	245	175	245

Notes: Eduloan and HEMIS data, restricted to loan applications for a private university, from 2004 to 2007 (applications dated November/December excluded). The explained variable is a dummy for enrollment in a public university the same year a loan application was received; "Above discontinuity" is a dummy for when the Empirica score is above zero. Ordinary least squares estimation with no controls other than functions of the Empirica score (different on each side of the discontinuity). Robust standard errors in parentheses. Bandwidth is defined with respect to the Empirica score. The last column uses the optimal bandwidth sample based on Calonico et al. (2014).

References

Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. J. Amer. Statist. Assoc. 97 (457), 288–292.

Akee, R., Copel, W., Keeler, G., Angold, A., Costello, E.J., 2010. Parents' incomes and children's outcomes: A quasi-experiment using transfer payments from casino profits. Am. Econ. J.: Appl. Econ. 2 (1), 65–118.

Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. Amer. Econ. Rev. 92, 1535–1558.

Bastian, J., Michelmore, K., 2018. The long-term impact of the earned income tax credit on children's education and employment outcomes. J. Labor Econ. 36 (4), 1127–1163.

Becker, G.S., 1964. Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. Columbia University Press, NY.

Belley, Ph., Lochner, L., 2007. The changing role of family income and ability in determining educational achievement. J. Hum. Cap. 1 (1), 37–89.

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

Brown, M., Scholz, J.K., Seshadri, A., 2012. A new test of borrowing constraints for education. Rev. Econom. Stud. 79 (2), 511–538.

Bucarey, A., Contreras, D., Munoz, P., 2020. Labor market returns to student loans for university: Evidence from Chile. J. Labor Econ. 38 (4).

Bulman, G., Fairlie, R., Goodman, S., Isen, A., 2017. Parental Resources and College Attendance: Evidence from Lottery Wins. mimeo.

Calonico, S., Cattaneo, M., Farrell, M., Tiniunik, R., 2017. Rdrobust: Software for regression discontinuity designs. Stata J. 17 (2), 372–404.

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

Cameron, S., Heckman, J., 2001. The dynamics of educational attainment for black, hispanic, and white males. J. Polit. Econ. 109 (3), 455–499.

Cameron, S., Taber, C., 2004. Estimation of educational borrowing constraints using returns to schooling. J. Polit. Econ. 112 (1), 132–182.

Canton, E., Blom, A., 2009. Student support and academic performance: Experiences at private universities in Mexico. Educ. Econ. 18 (1), 49–65.

Card, D., 2001. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica 69 (5), 1127–1160.

Carneiro, P., Heckman, J., 2002. The evidence on credit constraints in post-secondary schooling. Econ. J. 112 (482), 705–734.

Castleman, B., Long, B., 2012. Looking beyond College Enrollment: The Causal Effect of Need-Based Grants on College Access, Persistence, and Graduation. NBER Working Paper 19306.

Cattaneo, M., Jansson, M., Ma, X., 2018. Rddensity: Manipulation testing based on density discontinuity. Stata J. 18 (1), 234–261.

Cattaneo, M., Jansson, M., Ma, X., 2020. Simple local polynomial density estimators. J. Amer. Statist. Assoc. 115 (531), 1449–1455.

Coelli, M., 2011. Parental job loss and the education enrollment of youth. Labour Econ. 18, 25–35.

Dahl, G., Lochner, L., 2012. The impact of family income on child achievement: Evidence from the earned income tax credit. Amer. Econ. Rev. 102 (5), 1927–1956.

Department of Education, 2010. Education Statistics in South Africa 2008. Department of Education, South Africa.

Department of Education, 2016. Education Statistics in South Africa 2014. Department of Education, South Africa.

Department of Higher Education and Training, 2010. Report of the Ministerial Committee on the Review of the National Student Financial Aid Scheme.

Department of Higher Education and Training, 2014. Statistics on Post-School Education and Training in South Africa 2012. Department of Higher Education and Training, South Africa.

Department of Higher Education and Training, 2019. Statistics on Post-School Education and Training in South Africa 2017. Department of Higher Education and Training, South Africa.

Department of Higher Education and Training, 2022. Statistics on Post-School Education and Training in South Africa 2020. Department of Higher Education and Training, South Africa.

Duflo, E., Dupas, P., Kremer, M., 2017. The Impact of Free Secondary Education: Experimental Evidence from Ghana. mimeo.

Fack, G., Grenet, J., 2015. Improving college access and success for low-income students: Evidence from a large need-based grant program. Am. Econ. J.: Appl. Econ. 7 (2), 1–32.

Goldrick-Rab, S., Harris, D., Kelchen, R., Benson, J., 2015. Need-Based Financial Aid and College Persistence: Experimental Evidence from Wisconsin. IRP Discussion Paper 1393-12.

Hahn, J., Todd, P., Van der Klaauw, B., 2001. Identification and estimation of treatment effects with a regression discontinuity design. Econometrica 69 (1), 201–210.

Hilger, N., 2016. Parental job loss and children's long-term outcomes: Evidence from 7 million fathers' layoffs. Am. Econ. J.: Appl. Econ. 8 (3), 247–283.

Imbens, G., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–475.

Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econom. Stud. 79 (3), 933–959.

Imbens, G., Lemieux, T., 2007. Regression discontinuity designs: A guide to practice. J. Econometrics 142 (2), 807–828.

Kane, Th., 2007. Public intervention in postsecondary education. In: Hanushek, Eric A., Welch, Finis (Eds.), Handbook of the Economics of Education. Elsevier Sci, Amsterdam.

Keane, P.K., Wolpin, K.I., 2001. The effect of parental transfers and borrowing constraints on educational attainment. Internat. Econom. Rev. 42 (4), 1051–1103.

Keswell, M., Poswell, L., 2004. Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. South Afr. J. Econ. 72 (4), 834–860.

Lee, D., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.

Lochner, L., Monge-Naranjo, A., 2011. The nature of credit constraints and human capital. Amer. Econ. Rev. 101 (6), 2487–2529.

Loken, K., 2010. Family income and children's education: Using the norwegian oil boom as a natural experiment. Labour Econ. 17 (1), 118–129.

Lovenheim, M., 2011. The effect of liquid housing wealth on college enrollment. J. Labor Econ. 29 (4), 741–771.

Lovenheim, M., Reynolds, C., 2013. The effect of housing wealth on college choice: Evidence from the housing boom. J. Hum. Resour. 48 (1), 1–35.

Manoli, D., Turner, N., 2016. Cash-on-Hand & College Enrollment: Evidence from Population Tax Data and Policy Nonlinearities. NBER Working Paper 19836.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. J. Econometrics 142, 698–714.

Melguizo, T., Sanchez, F., Velasco, T., 2016. Credit for low-income students and access to and academic performance in higher education in Colombia: A regression discontinuity approach. World Dev. 80, 61–77.

Nielsen, H., Sørensen, T., Taber, C., 2010. Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform. Am. Econ. J.: Econ. Policy 2 (2), 185–215.

Pan, W., Ost, B., 2014. The impact of parental layoff on higher education investment. Econ. Educ. Rev. 42, 53–63.

Rau, T., Rojas, E., Urzúa, S., 2013. Loans for Higher Education: Does the Dream Come True? NBER Working Paper 19138.

Salisbury, T., 2016. Education and inequality in South Africa: Returns to schooling in the post-apartheid era. Int. J. Educ. Dev. 46, 43–52.

Schultz, T., 2004. School subsidies for the poor: Evaluating the mexican progresa poverty program. J. Dev. Econ. 74, 1999-250.

Social Surveys, 2009. Access to Public Higher Education in South Africa: Financial Constraints and Consequences. mimeo, Social Surveys, South Africa.

Solis, A., 2017. Credit access and college enrollment. J. Polit. Econ. 125 (2), 562–622.

Statistics South Africa, 2006. Quarterly Employment Statistics, June 2006, September 2006, December 2006.

Steiner, V., Wrohlich, K., 2012. Financial student aid and enrollment in higher education: New evidence from Germany. Scand. J. Econ. 114 (1), 124–147.

Stinebrickner, R., Stinebrickner, T., 2008. The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. Am. Econ. Rev. 98 (5), 2163–2184.

Stumpf, R., 2008. Tuition Fees: Higher Education Institutions in South Africa. HESA Report, Higher Education South Africa.