The distance elasticity at short distances - A study of the library choice of Oxford students

Ferdinand Rauch*, Kristiina Tuomikoski†

January 2019

Abstract

This paper presents a number of facts on the use of the Bodleian libraries by Oxford students. We pay particular attention to the importance of the distance between a student's home and a library on the choice of which library to use. This small scale distance elasticity is an important parameter for urban economics. We find a distance elasticity of around -0.3, closer to zero than observed in related studies.

JEL classification: F14, R12

Keywords: Library usage, Distance elasticity, Gravity model

We thank Frankie Wilson of the Bodleian Libraries for giving us access to the data used in this paper. We also thank numerous colleagues and seminar participants for comments.

*University of Oxford, ferdinand.rauch@economics.ox.ac.uk

†Bank of Finland

1 Introduction

This paper studies the way in which the distance between the home of a student and the location of a library influences the probability of a visit. This distance elasticity, closely related to the cost of distance, is a central parameter to model agglomeration economies and the shape of cities. It is also a central parameter for a planned retail business that tries to estimate and compare demand at various locations in a given city. Despite being of such importance, estimates of this elasticity are rare in the literature, likely because they require data that include addresses for customers and businesses that are hard to obtain. We use a dataset of library visits that includes students' addresses and allows us to measure this distance elasticity.

When we look for comparable evidence, we find distance elasticities estimates from gravity equation estimates in international trade. There are various reasons why we think that students' choice of a library could be compared to trade. First, by visiting a library, students consume a local service and accept a distance cost in order to do so. This is a form of trade. Second, recent theories of the gravity model have emphasized search costs of agents moving in space, see discussion below. Students' moving around Oxford can be seen as a more literal interpretation of this theory than goods shipped over oceans. Finally, the gravity framework has been applied to model international migration and migration within countries (Bakker et al 2018, Etzo 2008). Our exercise involves a form of short-distance migration of people.

On a larger scale, a meta-analysis of the distance coefficient (Disdier and Head 2008) shows that across the 159 papers included in their dataset, distance elasticity is consistently close to -1 and on average amounts to -0.93. There have been a few attempts to explain this persistence in the gravity estimates. Chaney (2013) builds a model based on input-output interactions between firms and shows that in such models there is a relationship between the firm distribution and the distance coefficient. Rauch (2016) shows that the gravity relationship with a distance coefficient of -1 should hold in any setting where force originating from a point covers all the possible area and behaves similarly at all distances from the origin. A distance coefficient close to -1 might therefore also be expected from the results of this study. However, the estimated distance elasticity of library visits, -0.29, is much smaller in absolute value than the traditional trade distance elasticity. This suggests that distance may enter the gravity model through different mechanisms than in trade or that economic gravity behaves differently at small scale, and could imply a non-linearity in the distance relationship.[1]

[1] Another systematic deviation of the distance coefficient has been documented by Lendle et al

A natural question arises to test the distance coefficient at smaller scales.[2] An existing attempt is Hillberry and Hummels (2008), who use data on individual firm to firm shipments from the 1997 U.S. Commodity Flow Survey that allows to describe distance frictions at a much smaller scale than in trade literature generally. They conclude that firm shipments are extremely localised: Shipment values within 4-mile zip codes are three times greater than values outside this range. This conclusion implies a steep distance coefficient at small scales, the opposite of what we conclude here. Another closely related paper is Davis et al (2018), who study restaurant choice using online reviews. They report travel time coefficients that are below -1, in some cases even below -2. Our data has a few advantage over theirs: (i) The addresses we use are measured and not inferred from reviews, (ii) the dataset we cover contains the universe of student, including those that never visit a library, and hence avoids selection problems and (iii) the means of transport used by students are restricted to walking and cycling, which simplifies the measurement of transport times. The distances we study are smaller than in both these studies. Our findings confirm that distance has a negative and statistically significant effect on student flows at this scale. However, the gradient is flatter than implied by these papers, which could imply non-linearities in the distance coefficient over scale. Finally, Couture (2016) estimates gains from variety and travel time reductions for the restaurant sector and argues that the first is more important.

A large marketing and retail literature has evolved around questions of consumer behaviour across space (Dion and Cliquet 2013) and modelling the optimal location for retail firms (Cliquet 2013).[3] The distance elasticity estimate found in this paper also characterises local consumer demand and thus contributes to research in industrial organisation. Dion and Cliquet (2013) describe in their survey that distance remains a key factor in explaining consumer choice of stores and according to Brooks et al (2018): "location is arguably the most important decision affecting the success or failure of retail establishments". As an example, distance to stores seems to influence obesity (Robinson et al 2013). This kind of study faces some difficulty since it requires a dataset that includes the address of both the customers and the business in question,

(2016) for good purchased over the internet. For a discussion of the relationship of the distance coefficient and urbanisation see Rauch(2014).

[2]An analogy with physics can be made, where Jenke et al (2014) develop a methodology to accurately search for deviations from Newtonian gravity at small scales and conclude that gravity holds at least until the energy scale of 10^{-14}eV.

[3]In this literature gravity type modelling is an even older tradition than in trade: Cliquet (2013) cites Reilly's law of retail gravitation (1931) as the base of earliest models in marketing research whereas Tinbergen introduced gravity to trade in 1962.

which are hard to come by. Our model could be used to make predictions of demand for library services for locations in Oxford, and under some assumptions for other services.

This paper relates to agglomeration economies as discussed in the urban economics literature, and the consumer city (Glaeser et al 2001, Eizenberg et al 2015). While this literature has focused on production side externalities, Redding (2010) argues that consumption externalities might be equally important. He also suggests that empirical literature on spatial economics is scarce and that distinguishing between the sources of agglomeration forces, their contributions and identifying the scales at which they operate are key topics for the literature to investigate. Both the magnitude of the distance elasticity and the non-parametric relationship between distance and consumer flows can thus be useful for subsequent research. Our base estimation results uses a trade flow between a college and a library, and includes college and library fixed effects, in addition to other control variables for student characteristics. Our estimate of a distance elasticity of -0.29 on an exponential distance decay curve implies that a one percent increase in distance of a library to all students corresponds to loss of 0.0029 percent of students. Expressed differently, on average an 83 m increase in distance would correspond to a loss of about 20 students for a library per day.

Finally, this paper also describes university students' library use patterns. These descriptive statistics on student life could be of interest to library and academic administrators and the wide academic community.

This paper is organised as follows: Section 2 describes the data on student library visits and provides facts on students' use of library services, Section 3 provides detail on the application of the gravity model and Section 4 presents the main results. Section 5 concludes.

2 Data and Descriptive Statistics

The University of Oxford library network consists of 32 Bodleian libraries, 41 college libraries and 32 other University libraries (Bodleian Libraries 2017b). The Bodleian libraries are affiliated with the University and the mission of the group is to support the teaching, learning and research at the University and to preserve access to the unique collections the libraries hold (Bodleian Libraries 2017d). The group includes research, faculty and departmental libraries. Most of these libraries provide places for students to study, including study desk as well as infrastructure for electricity and wireless internet access, and many students see libraries as their natural work spaces. The principal library is called the Bodleian library. It is a legal deposit library: A copy of all material published in the UK must be submitted to the Bodleian library. It is the second largest library in the UK (Bodleian Libraries 2017a).

Our dataset consists of entry records into a subset of the Bodleian libraries in Oxford. This dataset was provided to us by the Bodleian Assessment unit. It has not been used in the academic literature, and so far only been used for internal studies of the Oxford library system. The data comes at the level of individual students, and so is confidential, but access to other researchers may be granted by the Bodleian. In addition to the entry and exit of students to the libraries, we have information on degree, college and year of students that allows us to identify their residence. This makes it possible to measure the distance between consumers' starting and end points even at small scale. Potential aggregation bias from observing interactions only at large scale can thus be avoided.

Nine of the 32 Bodleian libraries record visitor entries. These libraries are the *Law Library* (BLL), the *Old Bodleian Library complex* (BOD), which includes the Radcliffe Camera and the main library building, the *Social Science Library* (SSL), the *Radcliffe Science Library* (RSL), the *Cairns Library*, the *Weston Library* the *Oriental Institute Library*, *Sackler Library* and *Philosophy and Theology Faculties Library*. At these libraries visitors swipe their reader or university cards through entrance gates. The system records the time of entry and the unique card number of visitors and stores them in one of the entry databases managed by the Bodleian.

The Bodleian libraries provided us with data for library entries from 4 libraries: BLL, BOD, SSL and RSL. These libraries have recorded entries since 2013, so the time in the dataset spans from 1st of August 2013 to 31st August 2016. The entry system of these 4 libraries records the students' university card numbers directly making it possible to identify students' colleges by matching the university card number with information held by the university student records. We could not use the data for the other libraries, since the data recording system differs, and for the other libraries we were unable to match the entry data with student records. The available data on BLL, BOD, SSL and RSL give a good representation of the Bodleian libraries. BOD is the main library of the group and is also the largest and most central of the Bodleian libraries. RSL and SSL rival or even exceed the BOD in daily student visits. These four libraries also represent a good variety of subjects: BLL, SSL and RSL specialise in materials for law, social sciences and natural sciences respectively. The Radcliffe Camera, part of the main Bodleian library complex, is used by the History Faculty and the BOD manages several special and rare collections.

To match Oxford University student identifiers from library entries, we search for card number information in the University student records database. This database gives us information on students' college, department, division and year of study. A limitation is that the student record database only contains information on current University of Oxford students. Thus if the card number was not found within the Student Records, we use student information from the Bodleian internal database if available.

Oxford University has an undergraduate entry of around 3200 students each year. We observe that in the first term of 2014, 2200 first year undergraduate students enter at least one of our four libraries. 1400 enter only one of the four, 570 enter two of the four, 130 enter three of the four, and only 15 students enter all four libraries. These numbers show that our four libraries include observations on a majority of Oxford students, and a sizable fraction of students visit multiple libraries, even within the restricted sample we observe. This shopping behaviour is what allows us to infer a distance coefficient. This number also suggests that zeros make up around a third of first year students in their first term. This suggests that while there is a sizable number of zero flows, our dataset leaves much observable variation.

At Oxford, colleges provide student accommodation that does not need to be on the main site. However, by tradition, first year students are typically given rooms on the main college site (University of Oxford 2017a). For 31 out of the 35 colleges and permanent private halls first year undergraduates are housed on the main sites (Oxford University Student Union 2016). We therefore restrict the dataset to first year students for which we measure their room address with great confidence. We measure all pairwise distances between these four libraries and all Oxford colleges. We construct this distance matrix by recording the shortest walking distance in meters between a college and a library as suggested by Google Maps. The addresses of the libraries and colleges proposed by Google Maps were used.

There are about 1,951,000 observations in the raw dataset of which about 1,346,000 observations are student entries. Trimming to only first year undergraduate students leaves 430,000 library entries from 12,695 unique card numbers. We aggregate these data further to aggregate flows from colleges to libraries so that the cleaned dataset contains 70,675 daily college-library student flows.our universities are the best in the world.

Table 1 describes the average distance from colleges to libraries. From the 4 main libraries, the BOD is the most central library with a 612 m average distance to colleges. SSL is the most distant library, on average 1027 m away from colleges. The table also shows the daily average number of visits, and the average number of colleges whose members show up at each library. We show in columns (5) and (6) that the SSL is the

Table 1: Average distances to libraries

	(1) Mean (m)	(2) Std. Dev.	(3) Min	(4) Max	(5) Day visits	(6) Day colleges
BOD	612	398	120	1800	412.1	21
RSL	714	276	210	1700	254.2	20.9
BLL	971	349	133	1920	216.6	13.8
SSL	1027	390	120	2000	513.1	19.6

most popular library with 513 unique entries per day. The BLL is least busy with 217 average daily entries. Comparing the libraries using the average daily student flow per college, we see that all four libraries receive daily visits from over 13 colleges, and so there is healthy variation in distance every day.

Libraries attract students from the academic divisions for which they hold specialised materials. Figure 1 shows that BLL and SSL are mostly used by social science students, humanities students favour the BOD and RSL is most popular with mathematical, physical and life science students. The figure also shows that this specialisation is not complete, and all libraries accommodate students from other subjects. Figure 2 plots the average number of entries into the libraries by hour and reveals interesting differences in the most popular entry times. 10 am is the peak entry hour for BLL, 11 am for RSL, 12 pm for BOD and 1 pm for SSL. BLL looks like the only library that is used at very early hours, as there are entries even between 1 am and 7 am. The BOD usage pattern seems to have two peaks: the first in the morning at 8 am after which visits peak again between 12 pm and 1 pm. RSL also tends to get most visits early during the day as the most popular entry hours are between 9 am and 1 pm, whereas SSL tends to get more entries later in the afternoon, between 1 pm and 4 pm. These patterns are likely to be related to the timing of classes and lectures that might be organised within the libraries.

Plotting daily entries over time in Figure 3 reveals strong termly pattern in the usage of libraries. The academic year in Oxford is divided into three terms of 9 weeks: Michaelmas, Hilary and Trinity. Michaelmas generally spans from 1st of October to 17th of December and Hilary begins 7th January and ends around 25th of March. Trinity is the last term of the year beginning around the 20th of April and ending the 6th of July and is the term during which most of university exams are organised. The 4 libraries and most noticeably the RSL and BLL experience visitor spikes during the first days of terms. For RSL and BLL the highest visitor peaks occur in the first term of the year, Michaelmas. BOD and SSL, on the other hand, experience their busiest days

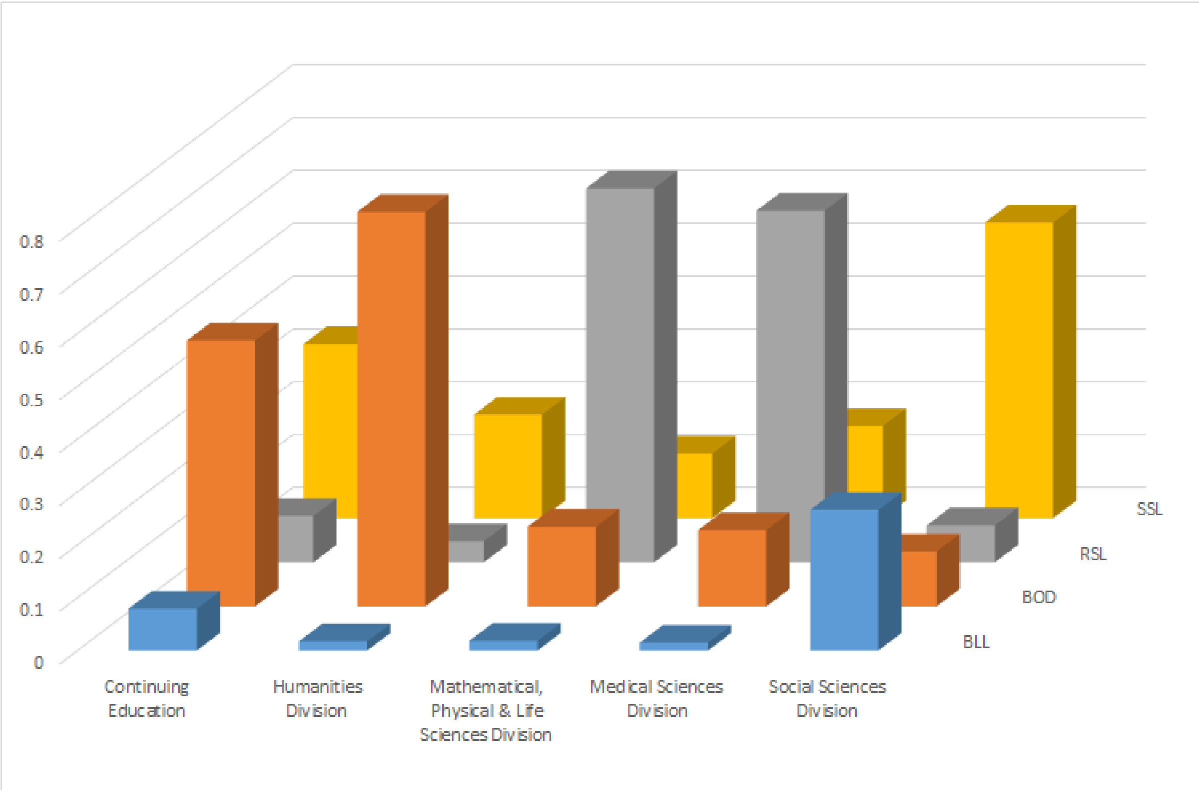


Figure 1: Student entries by division

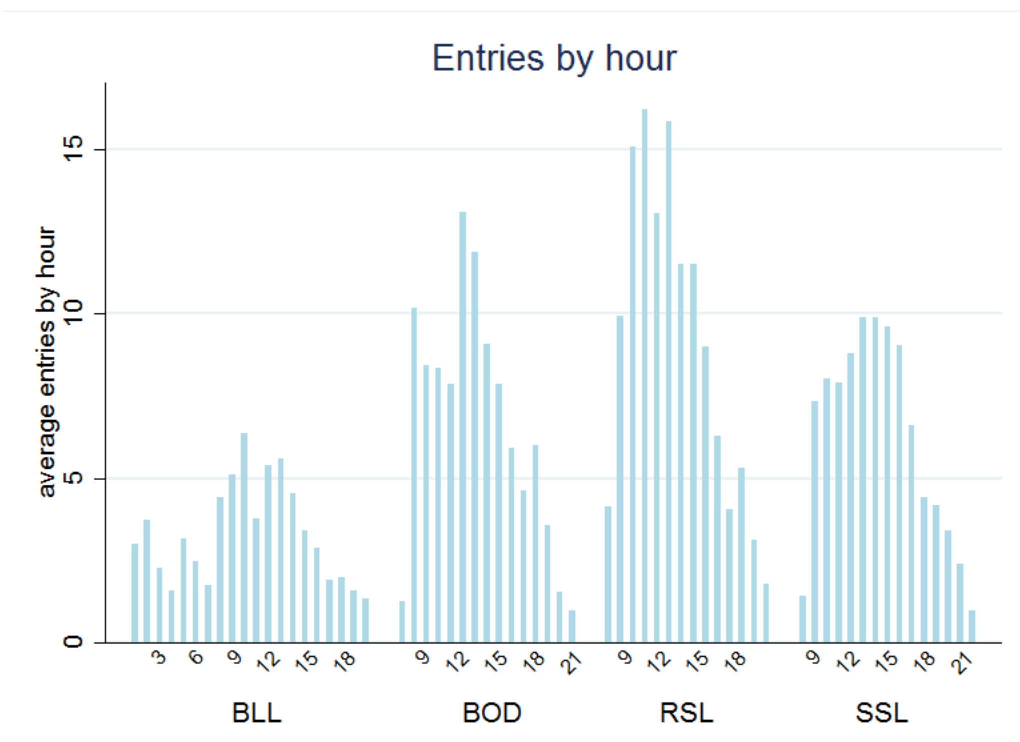


Figure 2: Entry hours

during Trinity term. BOD is one of the closest libraries to the Examination Schools, where the exams are held, which probably contributes to its popularity during Trinity term. In theory, libraries refuse entry to students if they are at full capacity. This could potentially affect our results. However, as is clear from this figure, the four libraries run well below peak capacity most of the time, and so this constraint is hardly ever binding.

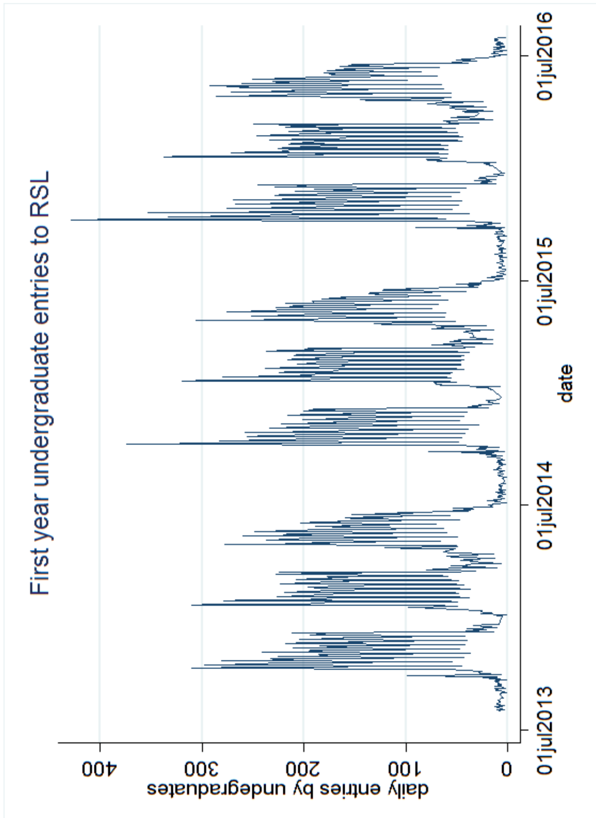
There were two main shocks to the Bodleian library system during this period. First, October 2014 marks the opening of the Weston library after three years of reconstruction during which it was completely closed to students. Second, BLL had big renovations while remaining partially open. Major reconstruction works began in September 2015 and finished by November 2016. As Figure 3 shows, the visits to BLL during 2015-2016 academic year look much lower than in previous years suggesting the renovation works had a noticeable effect on law library visits.

3 Methodology

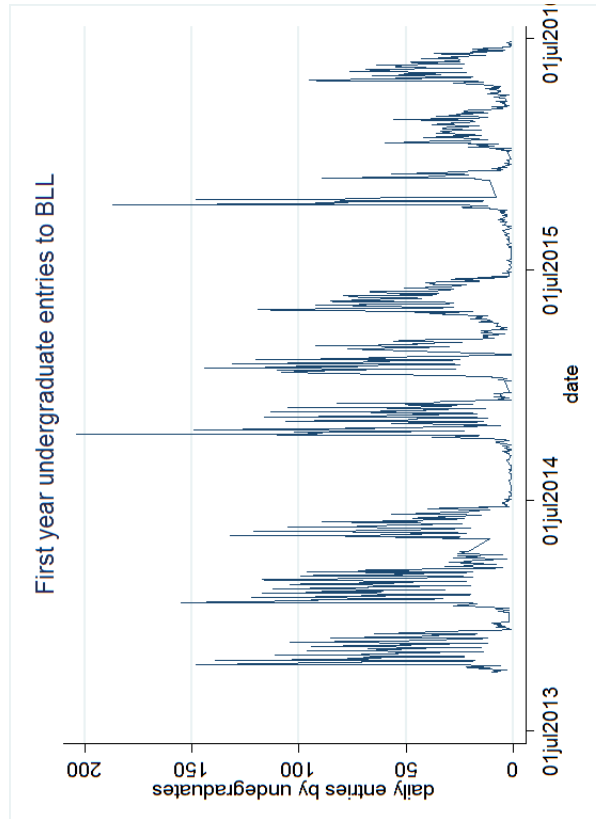
Student i visiting library l gets utility u_{il} from doing so. Numerous factors influence function u_{il}. First, there are a number of subject specific factors by which students prefer the library of their chosen division. These include that students may attend lectures in buildings closer to their subject library and wait there between lectures, they may have stronger social networks in their area, or they may consult books related to their subject there. The five subject divisions are the *continuing education division*, the *humanities division*, the *mathematical, physical and life sciences division*, the *medical sciences division* and the *social sciences division*. Second, there are time related factors, such as day of the week, or hour of day. Third, some effects relate to the origin college of students. Some colleges may have better college libraries, or provide better amenities such that students are less likely to leave college than others. Finally, there are individual effects, such as some students may prefer the atmosphere of one library over another, some libraries are closer to coffee shops than others, differences in comfort of chairs, desks, electricity supply or internet speed. These include the distance cost each student faces relative to their starting position. Taking these factors into account, we can write for the probability that student i visits library l:

$$\ln u_{ilt} = \alpha + \beta_1 \mathbb{1}[d_l = d_i] + \beta_2 \ln D_{c(i)l} + \theta_{c(i)} + \phi_l + \mu_t + \epsilon_{ilt}, \tag{1}$$

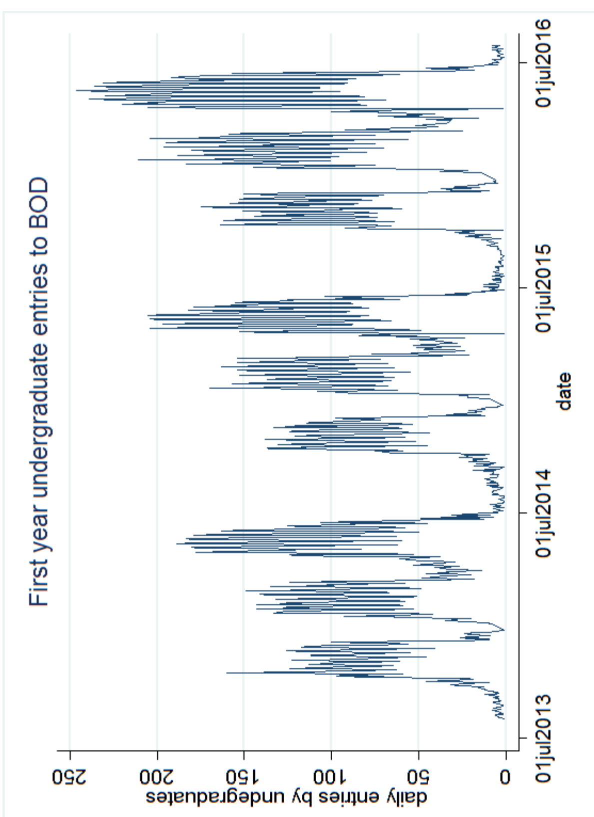
where $\mathbb{1}[d_l = d_i]$ indicates when the division of student i matches library l, $\theta_{c(i)}$ is



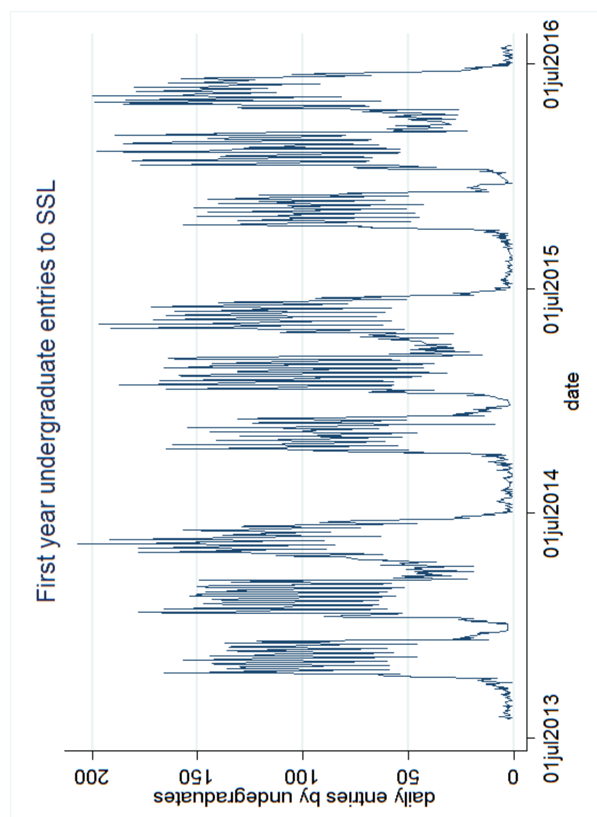
(a) RSL



(b) BLL



(c) BOD



(d) SSL

Figure 3: First year undergraduate entries over time

a college fixed effect, ϕ_l a library fixed effect that captures elements of a library that influence all students in the same way, and $D_{c(i)l}$ the distance between a college and a library. Term ϵ_{ilt} can be seen as a random variable that captures individual preference differences for libraries that can vary over time. Once we label the observable constant terms $[d_l = d_i]$, ϕ_l and μ_t by C, the utility ratio for two libraries becomes:

$$\frac{\ln u_{il1t}}{\ln u_{il2t}} = \frac{C + \beta_2 \ln D_{c(i)l1} + \epsilon_{il1t}}{C + \beta_2 \ln D_{c(i)l2} + \epsilon_{il2t}}. \tag{2}$$

The trade off between two libraries becomes the choice between a distance cost and an individual preference shock. Consider two extreme cases: if all students have the same preference ϵ_{ilt}, then distance cost becomes the only relevant varying factor in this trade off. Every student would only visit their closest library and we would not observe variation in the data. If a few students still enter a distant library occasionally due to some omitted error term, and so we find some variation in the data, we would estimate a very steep distance coefficient. There would be many zero flows, as for most students their college library, which is not part of our dataset, is their closest library. If on the other hand β_2 is zero, and students don't consider distance an important factor, the whole variation would rest on ϵ_{ilt}. Since this is orthogonal to distance accounted for by $D_{c(i)l}$, we would not estimate a distance coefficient that is significantly different from zero in this case. A distance coefficient estimate that is neither zero nor very steep indicates a service with some scope for individual preferences.

This discussion gives some indication for the comparability of distance coefficients across different types of local services. While the cost of distance may be the same for a consumer when consuming very different services, the scope for individual utility variation ϵ over stores may differ across industries, which would imply different distance coefficients. In similar settings a conventional assumption is that preference shocks follow a Frechet distribution (Eaton and Kortum 2002, Ahlfeldt et al 2015), in which case the log distance specification as above is the expected specification. As we show in a non-parametric specification below, the log distance functional form assumption is a good approximation in our case.

A simple descriptive graph shows that distance is an important consideration in student's decision which library to visit. Figure 4 shows observed distances against random distances. In this graph, the grey line shows the distribution of college-library distances in Oxford, while the black line shows the distribution of distances we actually observe. As expected, short distances are observed more frequently than a random allocation would predict, which demonstrates that distance is indeed an important

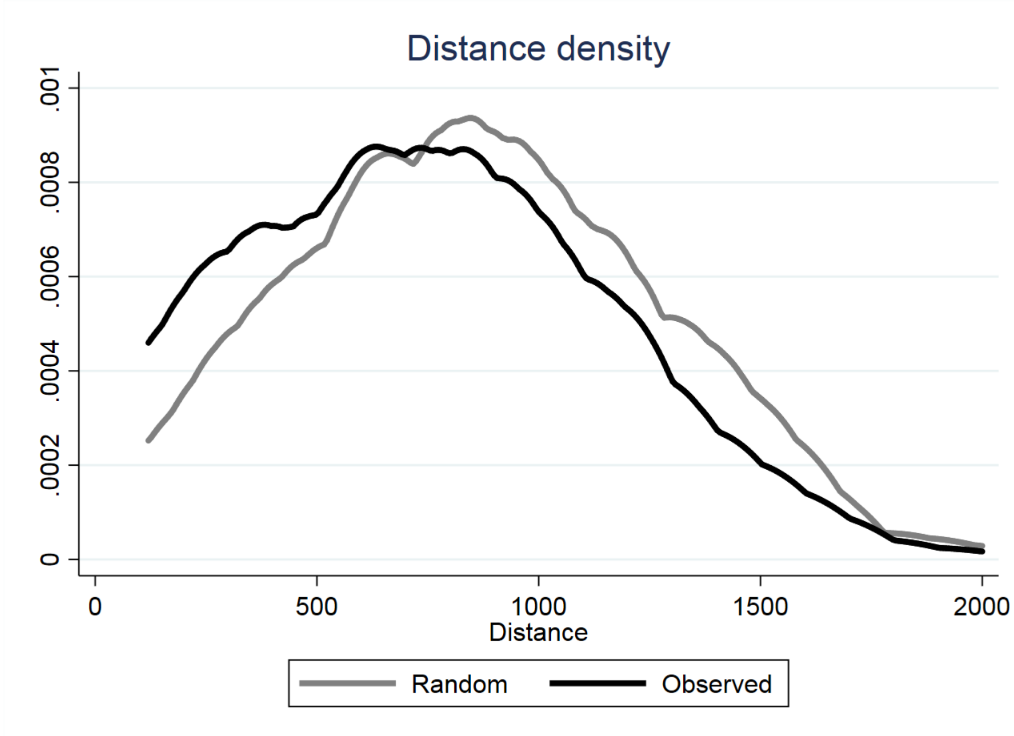


Figure 4: Histogram of distances between colleges and libraries

factor for students deciding which library to visit.

The essential distance variation of interest here is at the level of college-library, since that is the level at which we measure distance. Hence we can aggregate this equation to the level of college-library without loss of information of the dimension of interest. For convenience we aggregate the time dimension to day, and the important controls for time become term, day of week and year. The aggregated equation then becomes:

$$\ln V_{clt} = \alpha + \theta_c + \phi_l + \mu_t + \beta_1(p(d_{cl})) + \beta_2 \ln D_{cl} + u_{clt} \tag{3}$$

where d_{cl} measures the proportion of students from college c that are in division d. The term V_{clt} captures the sum of visits of students from college c in librarly l on day t. College and library fixed effects account for different sizes of colleges and libraries. This equation is closely related to an empirical gravity equation with origin and destination fixed effects, as found in studies of trade. Since our focus is on the effect of distance, it is appropriate for us to net out other factors with the use of fixed effects. In this equation we assume a functional form for the role of distance without good justification. Below we present non-parametric results that show that this is indeed a good approximation for the effect in this setting.

One concern could be that this aggregation assumes that the effect of $d_l = d_i$ is the same for all divisions. It could however be larger for one division than another if one division relies more heavily on books found in the divisional library. To address this concern, we present a robustness check that eliminates all student visits to their divisional library below.

4 Main results

4.1 Base results

The main table of results is Table 2. The first column is a direct application of equation 4. The first coefficient is our estimate of the distance elasticity. It is -0.29 and is significant at a 1% level of statistical significance. The next three variables refer to three library fixed effects, relative to the omitted Law Library. These coefficient show that the Social Sciences Library is overall more popular than the Law Library next door. The next three fixed effects indicate the three Oxford terms, relative to the omitted time out of term. Students spend more time in the library at the beginning and end of the academic year, F-tests indicate that the differences between terms are all statistically significant at 10%. Weekday fixed effects reveal that students are most motivated to visit the library on Mondays, an effect that monotonically decreases until Saturday. The underlying regression includes college fixed effects and $p(d_{cl})$ fixed effects giving the share of students by division for each college. We omit to show these fixed effects due to space constraints.

In the second column we repeat the exercise, but measure distance by straight line based on college and library coordinates instead of distances indicated by Google Maps. This is to address concerns related to our distance measure. All coefficients are fairly similar in terms of magnitude and statistical significance. The magnitude of the distance elasticity falls slightly to -0.275. The new and original coefficients are significantly different with a χ^2-statistic of 23.38. However, we expect the co-ordinate distances to underestimate the true distance, especially at short intervals, since the straight line distance does not take into account the street layout of the city. This is reflected in the lower magnitude of the distance coefficient in the second regression. The walking distance thus remains the preferred measure of distance.

One shortcoming of the analysis so far is the assumption that the starting point for students is their college main site. However, students might use libraries after lectures and classes in their departments or other activities elsewhere in the city. To account for

Table 2: Main Regression Results

	(1)	(2)	(3)	(4)	(5)
	$\ln(V_{cl})$	$\ln(V_{cl})$	$\ln(V_{cl})$	$\ln(V_{cl})$	$\ln(V_{cl}+1)$
$\ln D_{cl}$	-0.292***		-0.149***	-0.282***	-0.134***
	(0.00695)		(0.00719)	(0.00683)	(0.00596)
$\ln D^c_{cl}$		-0.275***			
		(0.00737)			
BOD	-0.00342	0.0208	-0.0780**	-0.00689	0.392***
	(0.0506)	(0.0505)	(0.0361)	(0.0514)	(0.00505)
RSL	-0.0440	-0.00763	0.203***	-0.0536	0.530***
	(0.347)	(0.346)	(0.0168)	(0.315)	(0.00471)
SSL	0.532***	0.486***	0.0664***	0.481***	0.411***
	(0.0500)	(0.0500)	(0.00902)	(0.0492)	(0.00481)
Michaelmas	0.784***	0.782***	0.226***	0.704***	0.729***
	(0.00847)	(0.00848)	(0.00954)	(0.00828)	(0.00449)
Hilary	0.771***	0.770***	0.168***	0.696***	0.778***
	(0.00804)	(0.00806)	(0.00866)	(0.00789)	(0.00447)
Trinity	0.821***	0.820***	0.259***	0.723***	0.745***
	(0.00821)	(0.00823)	(0.00897)	(0.00800)	(0.00471)
Mon	0.606***	0.605***	0.350***	0.446***	0.462***
	(0.00986)	(0.00990)	(0.0221)	(0.00978)	(0.00698)
Tue	0.543***	0.543***	0.325***	0.387***	0.425***
	(0.00975)	(0.00979)	(0.0221)	(0.00966)	(0.00684)
Wed	0.458***	0.457***	0.283***	0.301***	0.379***
	(0.00968)	(0.00972)	(0.0221)	(0.00965)	(0.00674)
Thu	0.414***	0.414***	0.266***	0.265***	0.349***
	(0.00968)	(0.00972)	(0.0222)	(0.00962)	(0.00669)
Fri	0.353***	0.353***	0.255***	0.211***	0.306***
	(0.00973)	(0.00976)	(0.0223)	(0.00967)	(0.00664)
Sat	-0.0277***	-0.0280***	0.0496**	-0.0998***	0.0489***
	(0.00983)	(0.00988)	(0.0224)	(0.00979)	(0.00631)
Obs.	70,675	70,675	29,263	66,105	138,880
Morning only			Yes		
After 10am only				Yes	

All columns include in additional college fixed effects, and interactions between library fixed effects and the proportion of students from each division in each college. $\ln D_{cl}$ measures the distance between colleges and libraries according to google maps, $\ln D^c_{cl}$ measures a straight line distance based on college and library coordinates. BOD, RSL and SSL are library fixed effects, Michaelmas Hilary and Trinity are term fixed effects. Robust standard errors in parentheses.*** p<0.01,** p<0.05, * p<0.1.

Table 3: Change in student numbers from a 10% increase in average distance between libraries and colleges

	BLL	BOD	RSL	SSL
10% increase in average distance, m	97.1	61.3	71.4	102.7
change in daily visits per college	-0.65	-1.00	-1.35	-1.00
average daily number of visiting colleges	13.8	21	20.9	19.6
change in average daily undergraduate visits	-9.0	-21.0	-28.2	-19.6

this, the sample is restricted to include only entries into libraries between 6 am and 10 am, which may be better measures of the distance from home (see Larcom et al 2017 for a development of this argument). Classes in Oxford start at 9am the earliest, and end by 10am. Then all library entries before 10 am would not be visits made after a lecture. Column 3 in Table 2 displays the sub-sample regression results for early entries only. The distance elasticity is still significant at 1% level, but its magnitude is halved from the original estimate to -0.15. Running the regression on entries after 10 am also reveals that the resulting distance elasticity from this late-entry sample is very close to the original estimate (as in Column 4). This difference could suggest that distance is less important for the more motivated early students than for the student population as a whole.

A standard problem for estimating the gravity relationship using logarithms is that 0 flows are excluded from the estimation. As Head and Mayer (2014) explain, dropping the 0 observations may lead to selection bias. The log specification on the left hand side does not allow us to include the zero flows directly, and so we transform it to the closely related specification of $\ln(C_{cl} + 1)$ in Column 5. This specification allows us to include all the zero flows. The distance effect is still statistically significant, but its magnitude declines to -0.134. One possibility for the decline is that the 0 flows occur disproportionately at larger distances, which would cause the original sample to overestimate the distance effect. Yet we see in the data that they seem to occur quite evenly across distances. Another explanation is that many of the reported zeros are really omitted libraries and alternatives such as studying in a park or a cafe. If so, we record as zero many students who have a non-zero distance for that day. This measurement error may lead to the observed attenuation bias.

To interpret the magnitude of this distance coefficient, we simulate the effect of a hypothetical 10% increase of the distance to all colleges for each of the four libraries. For a 97 m increase in distance, BLL would lose about 9 students daily, BOD would lose 21 students for a 61 m increase, RSL 28 students for a 71 m increase and SSL

20 students for a 103 m increase. Taking an average for these libraries, an increase in distance of about 83 m corresponds to a loss of about 20 students per day.

The absolute value of the found distance elasticity is significantly smaller than 1. The hypothesis test on equality yields a t-statistic of 101.9. The distance effect on student flows is thus smaller than the distance elasticity found in conventional trade gravity estimations close to -1. There are examples of studies that find distance coefficients closer to zero, such as Feyrer (2009) who estimates a distance elasticity between -0.15 and -0.46 using variation from differences in sea distance caused by the closure of the Suez Canal. The distance elasticity of student flows is also similar to the elasticity of migration flows suggested by Beine (2011) (-0.4).

We next turn to a non-parametric specification to test the validity of the functional form assumption of adding distance in a log-linear way. Here we convert distance into 200 m bins added to the regression, replacing the log distance variable. The main coefficients of interest are in Table 4. Apart from the change to the distance measure, the first two columns in this table are the same as Columns (1) and (3) in Table 2. Figure 5 shows these regression coefficients graphically. Figure 5 shows a relationship between distance and library consumption that is monotonically decreasing at all levels. The fitted OLS line gives a slope of -0.11. The dependent variable is in logs while the distance dummies are shown on a log scale in the left panel, and on a linear scale in the right panel. We add the best fit approximation of a log distance approximation in both panels. Both panels seem to suggest that a log-log specification approximates the graph well, except perhaps for the largest distance coefficients. These two last coefficients however only represent a small share of the observations. The last distance bin contains less than half of one percent of the observations used here, while the second to last includes about one percent and a half. The magnitude of the slope implies that every additional 200 m reduce the student flows by about 10.4 %[4]. To interpret the individual coefficients they should be compared to the omitted reference category of distances greater than 2000 m. Thus, at less than 200 m, there are on average 3.5 more students per college visiting a library in a day compared to distances greater than 2000 m. For a rough estimates on total visits, given that the average number of colleges represented in a library in a day is around 19, libraries can attract approximately 67 students more from shortest distances than longest distances daily.

The third column in Table 4 adds a robustness check in which we omit the main subject library for each division. Thus this specification focuses on students who are

[4]The student flows are in logs, but the distance dummies are not, so the effect of a unit change in a regressor on the dependent variable is computed as $(\exp(\beta) - 1) \times 100\%$

Table 4: Non-parametric distance coefficients

	Full sample	Before 10am	Exclude main subject library
0-200m	1.507***	0.718***	1.406***
	(0.0408)	(0.0433)	(0.043)
200m-400m	1.270***	0.547***	1.100***
	(0.0392)	(0.0417)	(0.043)
400m-600m	0.966***	0.420***	0.931***
	(0.0386)	(0.0411)	(0.042)
600m-800m	0.945***	0.408***	0.815***
	(0.0378)	(0.0403)	(0.042)
800m-1000m	0.938***	0.410***	0.797***
	(0.0368)	(0.0393)	(0.041)
1000m-1200m	0.815***	0.343***	0.699***
	(0.0358)	(0.0382)	(0.040)
1200m-1400m	0.809***	0.345***	0.660***
	(0.0359)	(0.0384)	0.041
1400m-1600m	0.688***	0.250***	0.569***
	(0.0357)	(0.0386)	0.041
1600m-1800m	0.506***	0.199***	0.419***
	(0.0333)	(0.0345)	(0.040)
1800m-2000m	0.208***	-0.0260	0.161***
	(0.0412)	(0.0402)	(0.046)
Observations	70,675	29,263	46,401

All columns include in additional college fixed effects, library fixed effects, day fixed effects, term fixed effects and interactions between library fixed effects and the proportion of students from each division in each college. Distance measures are computed from the distance between colleges and libraries according to google maps. Robust standard errors in parentheses.*** p<0.01,** p<0.05, * p<0.1.

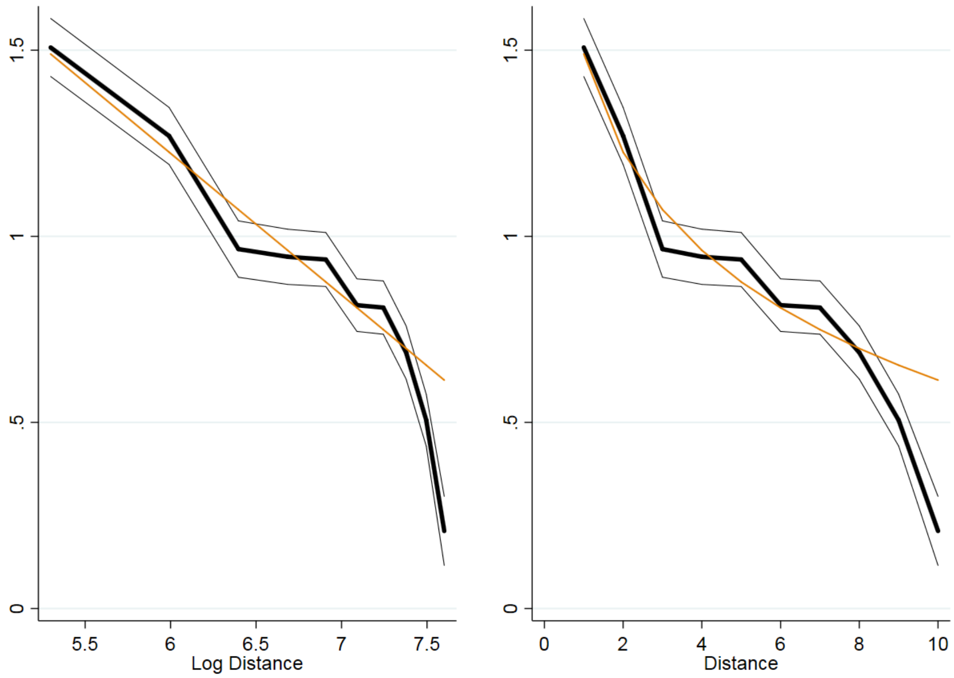


Figure 5: Distance coefficients of the bin regression, and fitted log distance

in a library that is not the main library used for their subject. These students are less likely to be influenced by lectures or the requirement for special books, and are more likely to behave like consumers looking for a quiet study space. We remove the following five subject-library combinations: Social Sciences - SSL, Medical Sciences - RSL, Mathematics etc. - RSL, Humanities - BOD and Education - BOD. These are the five outliers in Figure 1. As comparison between Columns (1) and (3) shows, this exclusion does not seem to affect the shape or significance of the distance relationship, and our main result does not appear to depend heavily on subject specific criteria.

4.2 Library substitution

There are two main shocks affecting the Oxford library system during the period we study. First, the reconstruction of the Law Library, which took place between September 2015 and November 2016. The library remained open, but there was significant noise disruption to readers in the library (Bodleian 2016). The second disruption is the opening of the Weston library in September of 2014. This is a main library building in central Oxford that was opened after many years of reconstruction. To see the effect of these disruptions on the use of the other libraries, we run regressions of the form:

$$\ln V_{clt} = \alpha + \theta_c + \phi_l + \mu_t + \beta_1(p(d_{cl})) + \beta_2 \ln D_{cl} + \beta_3 PostWeston + \\ \gamma_1(PostWeston \times \phi_l) + \beta_4 Construction + \gamma_2(Construction \times \phi_l) + u_{clt}. \tag{4}$$

This is the same equation as in our main regression, with the inclusion for the two treatment periods, and interaction terms for the treatment periods with each of the libraries. This specification allows us to see substitution, but it also corrects for bias of our main equation that may arise from these disruptions. We pick the Law Library as the omitted category, given that one of the two shocks affects it. Table 5 shows the main coefficients of this regression, first for the Law Library construction, then for the opening of the Weston Library, and finally both shocks in one specification. We consider the third column to be the main one here, since the omitted disruption may cause omitted variable bias in the other two. The coefficient on *Construction* is negative and significant, which suggests that fewer students used the Bodleian system during the construction period of the Law Library. The remaining three libraries were used more during the same period, as the three positive and significant interaction coefficients demonstrate. Most students moved to the Bodleian library, which is the furthest of the three, while the smallest effect is found for the Social Sciences Library, the closest one. This suggest that distance to the Law Library was not a main factor for substitution for the students going elsewhere.

The interaction coefficients with *Weston* suggest that the opening of the Weston Library attracted students mainly to the Bodleian Library, and less to the other three, with the smallest effect for the Social Sciences Library. The Bodleian Library is the closest one of the four to the Weston library, again suggesting that distance between libraries was not an important factor to explain substitution. Reasons why the closest library had a positive inflow of students include: (1) amenities such as the Cafe in the Weston could have made the nearby Bodleian Library more attractive; (2) positive agglomeration effects from students enjoying the possibility of running into each other or (3) students who visited the new Weston building out of curiosity may have entered the BOD in a similar touristic spirit, perhaps even for the first time.

To summarize this section, we show these results, but we do not see any clear lesson on the distance coefficient in it. We do not see that students move to the nearest library when their favourite library is under construction. We also do not see that a new library draws students away from its nearest neighbour. This could be explained by unobserved changes to amenities.

Table 5: Spacial substitution

	(1) Construction	(2) Weston	(3) Both
Construction	-0.186***		-0.180***
	(0.018)		(0.025)
Weston		-0.136***	-0.022
		(0.017)	(0.024)
BOD	0.063	0.051	0.016
	(0.054)	(0.056)	(0.055)
RSL	0.356	0.332	0.332
	(0.387)	(0.378)	(0.385)
SSL	1.066***	1.077***	1.077***
	(0.041)	(0.042)	(0.042)
BOD × Constr	0.342***		0.304***
	(0.019)		(0.019)
RSL × Constr	0.237***		0.224***
	(0.017)		(0.020)
SSL × Constr	0.179***		0.198***
	(0.017)		(0.0192)
BOD × Weston		0.212***	0.075***
		(0.015)	(0.018)
RSL × Weston		0.122***	0.026
		(0.016)	(0.019)
SSL × Weston		0.045***	-0.038**
		(0.015)	(0.019)
Observations	70,675	70,675	70,675

All columns include in additional college fixed effects, library fixed effects, day fixed effects, term fixed effects and interactions between library fixed effects and the proportion of students from each division in each college. Distance measures are computed from the distance between colleges and libraries according to google maps. Robust standard errors in parentheses.*** $p<0.01$,** $p<0.05$, * $p<0.1$.

4.3 Further discussion

In this study we treat the distance to the Bodleian Libraries as exogenous from the students' perspective. This would be violated if some students chose their college based on its location relative to the library. Indeed, undergraduate students apply to a college, and graduate students state a college preference on their application. Most students do get selected into their preferred colleges: in 2015, 18.1% of undergraduate applicants did not name a preference for any particular college and 73.5% of those who did got their preferred choice (University of Oxford 2017b). Although this suggests that there is scope for selection bias, the distance from a college to Bodleian libraries seems unlikely to be a major decision factor in the choice of colleges. To help prospective students select a college, Oxford University Student Union (2016) suggests considering college library opening times among other things, but there is no mention of relations of colleges with Bodleian libraries. In addition, Oxford students' academic achievements are fairly uniform across colleges. The University of Oxford publishes an academic ranking of its colleges based on a Norrington score, which assesses the performance of undergraduate students in their final exams. The scores are very close to each other, the top and bottom colleges are separated by only 10 percentage points in the 2015/2016 ranking (University of Oxford 2017c). This closeness of academic ability does not suggest that selection of students by college is an important factor in Oxford.

Another worry that one might have is that the effect of distance is not driven by the choice of students, but by the centrality of libraries and colleges. It might be that at the time of building, colleges and libraries were chosen by architects and planners to be located next to each other. In this case we would also observe higher student flows at shorter distances, but it would only be an artefact of the locations of colleges and libraries. Figure 4 plots the histogram of distances in the data. We would expect most of the mass to be at short distances if colleges and libraries were purposefully built close together. The histogram, however, shows great variation in the probabilities of distances between colleges and libraries below 1000 m. After 1000 m the histogram gradually declines, but even there the slopes of the estimated distance coefficients and the histogram are very different. This provides some reassurance that colleges and libraries are spread around the city and the distance effect is not driven by their location.

5 Conclusion

This paper investigates consumer behaviour across space by analysing students' library choices. Drawing upon the structure of the University of Oxford that includes several libraries and colleges, we create a dataset that includes both the starting and end points of students, overcoming a key challenge for research on consumer behaviour. Using data on students' use of libraries within a city also provides the opportunity to analyse economic flows on a small enough scale to avoid aggregation bias.

We study the student flows using an empirical framework related to the gravity model, traditionally applied to aggregate trade flows in the economics literature. This provides new evidence on the economic gravity relationship at small scale. The results suggest that the empirical regularity of a -1 distance elasticity found in trade gravity is not replicated at small scale. Another important finding of this study is some evidence that the non-parametric relationship between distance and economic flows seems to be well approximated by an exponential decay function.

The main results of this paper also provide an estimate of local consumer demand. The findings suggest that distance is a significant factor in students' choice of libraries and that the distance elasticity of daily student flows is about -0.29. A one percent increase in the distance to libraries reduces the number of visiting students by 0.0029 percent. This implies that on average, an 83 m increase in distance corresponds to a loss of about 20 students for a library per day.

References

- Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). 'The economics of density: Evidence from the Berlin Wall'. Econometrica, 83(6), 2127-2189.

- Bakker, Jan, Christopher Robert Parsons, and Ferdinand Rauch. "Migration and Urbanisation in Post-Apartheid South Africa." (2018), *World Bank Economic Review*, forthcoming.

- Beine, M., Docquier, F. and Calar Özden (2011), 'Diasporas', *Journal of Development Economics* 95(1), 30-41.

- Bodleian Libraries (2016), 'Law library refurbishment'. http://www.bodleian.ox.ac.uk/law/about/law-library-refurbishment, accessed 28.3.2017.

- Bodleian Libraries (2017a), '10 tips'. http://www.bodleian.ox.ac.uk/using/introduction/using, accessed 22.4.2017.

- Bodleian Libraries (2017b), 'Libraries'. http://www.bodleian.ox.ac.uk/subjects-andlibraries/libraries, accessed 28.3.2017.

- Bodleian Libraries (2017c), 'Renovations'. http://blogs.bodleian.ox.ac.uk/lawbod/ renovations/, accessed 28.3.2017.

- Bodleian Libraries (2017d), 'Strategy'. http://www.bodleian.ox.ac.uk/about/strategy, accessed 22.4.2017.

- Brooks, C., Kaufmann, P. and Lichtenstein, D. (2008), 'Trip chaining behavior in multidestination shopping trips: A field experiment and laboratory replication', Journal of Retailing 84(1), 29-38.

- Chaney, Thomas. 'The gravity equation in international trade: An explanation.' *Journal of Political Economy* 126.1 (2018): 150-177.

- Cliquet, G. (2013), Geomarketing, Wiley, chapter, *Retail Location Models*, pp. 137-164.

- Couture, Victor (2016). 'Valuing the Consumption Benefits of Urban Density', University of California Berkeley, Working paper.

- Davis, Donald, Jonathan Dingel, Joan Monras and Eduardo Morales, 'How Segregated is Urban Consumption?', *The Journal of Political Economy*, 2018.

- Dion, D. and Cliquet, G. (2013), Geomarketing, Wiley, chapter, *Consumer Spatial Behaviour*, pp. 29 - 56.

- Disdier, Anna and Keith Head (2008), 'The puzzling persistence of the distance effect on bilateral trade.', *Review of Economics and Statistics* 90(1), 37 - 48.
- Eaton, Jonathan, and Samuel Kortum. 'Technology, geography, and trade.' Econometrica 70.5 (2002): 1741-1779.
- Eizenberg, Alon, Saul Lach, and Merav Yiftach. 'Retail prices in a city: an empirical analysis.' Maurice Falk Institute for Economic Research in Israel. Discussion paper series. 7 (2015).
- Etzo, Ivan. "Internal migration: a review of the literature." (2008).
- Feyrer, J. (2009), 'Distance, trade, and income - the 1967 to 1975 closing of the suez canal as a natural experiment', Working Paper 15557, National Bureau of Economic Research. http://www.nber.org/papers/w15557, accessed 14.4.2017.
- Glaeser, E. L., J. Kolko, and A. Saiz (2001): 'Consumer city', *Journal of Economic Geography*, 1(1), 27-50.
- Head, K. and Mayer, T. (2014), 'Gravity equations: Workhorse,toolkit, and cookbook', pp. 131-195.
- Hillberry, R. and Hummels, D. (2008), 'Trade responses to geographic frictions: A decomposition using micro-data', *European Economic Review* 52(3), 527-550.
- Jenke, T., Cronenberg, G., Burgdörfer, J., Chizhova, L., Geltenbort, P., Ivanov, A., Lauer, T., Lins, T., Rotter, S., Saul, H., Schmidt, U. and Abele, H. (2014), 'Gravity resonance spectroscopy constrains dark energy and dark matter scenarios', *Physical Review Letters* 112(15).
- Larcom, Shaun, et al. "The benefits of forced experimentation: striking evidence from the London underground network." *The Quarterly Journal of Economics* 132.4 (2017): 2019-2055.
- Lendle, A., Olarreaga, M., Schropp, S., and Vezina, P. L. (2016). 'There goes gravity: eBay and the death of distance.' *The Economic Journal*, 126(591), 406-441.
- Oxford University Student Union (2016), The 'Alternative prospectus: Colleges'. http://apply.ousu.org/colleges/, accessed 22.4.2017.
- Rauch, Ferdinand. 'The geometry of the distance coefficient in gravity equations in international trade.' *Review of International Economics* 24.5 (2016): 1167-1177.
- Rauch, Ferdinand. "Cities as spatial clusters." *Journal of Economic Geography* 14.4 (2014): 759-773.

- Redding, S. J. (2010), 'The Empirics of New Economic Geography', Journal of Regional Science 50(1), 297-311.
- Robinson, Paul L., et al. 'Does distance decay modelling of supermarket accessibility predict fruit and vegetable intake by individuals in a large metropolitan area?.' Journal of health care for the poor and underserved 24.1 0 (2013): 172.
- University of Oxford (2017a), 'Accomodation'. https:// www.ox.ac.uk/students/ life/ accommodation?wssl=1, accessed on 25.3.2017.
- University of Oxford (2017b), 'College success rates'. https://www.ox.ac.uk/about/ factsand-figures/admissions-statistics/undergraduate/additional-info/college- successrates, accessed 26.4.2017.
- University of Oxford (2017c), 'Undergraduate degree classifications'. https://www.ox.ac.uk/ about/facts-and-figures/undergraduate-degreeclassifications?wssl=1, accessed 7.4.2017.