Mode-Choice Regressions

Lucas Javaudin

April 4, 2022

1 Data and Variables

INSEE 2017 census data (survey carried out between 2015 and 2019).[1]

We restrict the dataset to individuals living and working in the Rhône department.

The endogenous variable is the mode of transportation chosen by the individual for her commute trip, as reported in the census. The exogenous variables are the age of the individual, the number of car per employed individuals in the household, the sex, the professional occupation and the travel time for the morning commute.

Table 1: Share of each mode of transportation reported

Mode of transportation	Share
Car	60.69 %
Public transit	25.07 %
Walking	8.83 %
Cycling	3.95 %
Motorcycle	1.47 %

Table 2: Description of the numeric socio-demographic variables

Name	Description	Mean
Age	Age rounded to the nearest five-year age group	38.49
Cars per individual	Number of cars owned per number of employed in the household	0.84

Table 3: Description of the categorical socio-demographic variables

Name	Description	Most frequent category
Sex	Sex of the individual	man (50.67 %)
Occupation	Occupation of the individual, using INSEE nomenclature	employee (24.93 %)

Computing travel times The census data report the city where the individuals live and work (the district for individuals living or working in Lyon). The departure and arrival point of the trip is set to the administrative center of the cities (or districts).

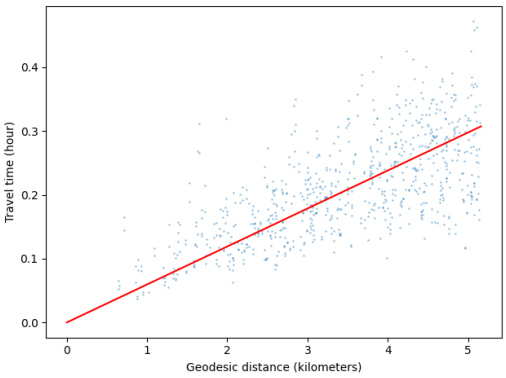
[1] `https://www.insee.fr/fr/statistiques/4507890?sommaire=4508161`

For the modes of transportation walking, cycling and motorcycle, travel times are computed using the open source routing engine GraphHopper (based on OpenStreetMap data), assuming no congestion.

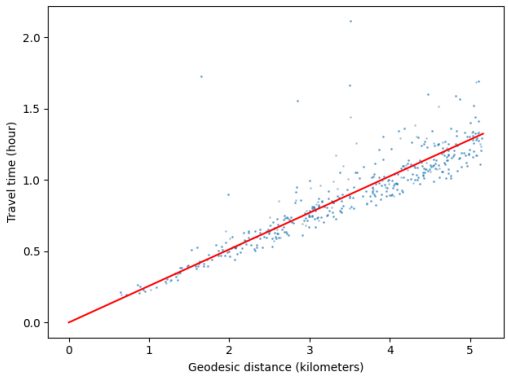
For trips by car, travel times are computed from HERE API, with congestion, assuming a departure at 8 AM on a Tuesday.

For trips by public transit, travel times are computed from HERE API, using predicted timetables, assuming a departure at 8 AM on a Tuesday. For some individuals, no public transit trip is available (10.87 % of total weight).

For intracity trips (i.e. trips where the living city is the same as the working city), the departure point is the same as the arrival point. Instead of setting travel times to zero for intracity trips, we use the following method. The geodesic distance (distance as the crow flies) of the trip is assumed to be equal to the radius of the city (assuming the city is a circle).[2] For each mode of transportation, the speed of the trip is assumed to be equal to the average speed of all intercity trips with a geodesic distance below 5 kilometers (computed by regressing the travel time of the trips on the geodesic distance of the trips, without constant).[3] Figures 1a, 1b and 2 show the regression of travel time on geodesic distance for intercity trips, for car, walking and public transit respectively.



(a) Regression of car travel time on geodesic distance for intercity trips



(b) Regression of walking travel time on geodesic distance for intercity trips

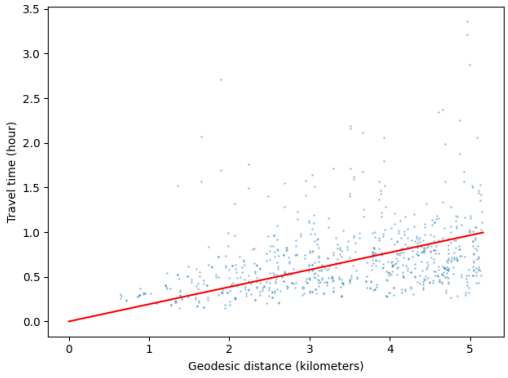


Figure 2: Regression of public transit travel time on geodesic distance for intercity trips

[2]The geodesic distance of an intracity trip for a city with an area x is assumed to be $d = \sqrt{x/\pi}$

[3]The geodesic speed is 16.8 km/h for car, 5.2 km/h for public transit, 3.9 km/h for walking, 12.2 km/h for cycling and 30.1 km/h for motorcycle. The speed (in real distance) is 25.7 km/h for car, 5 km/h for walking, 16.8 km/h for cycling and 46.6 km/h for motorcycle.

Individuals are removed if the travel time with their reported mode of transportation is larger than 90 minutes (3.28 % of total weight).

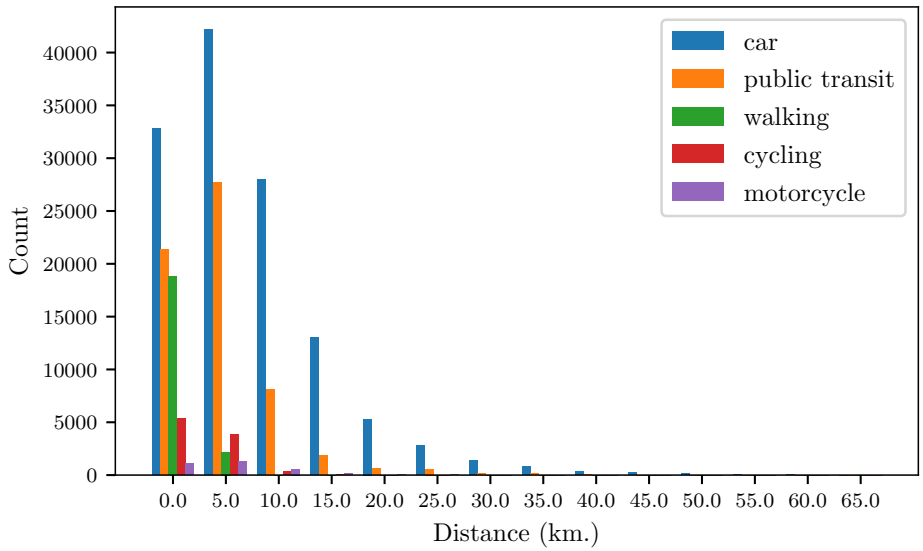


Figure 3: Mode shares as a function of geodesic distance between origin and destination

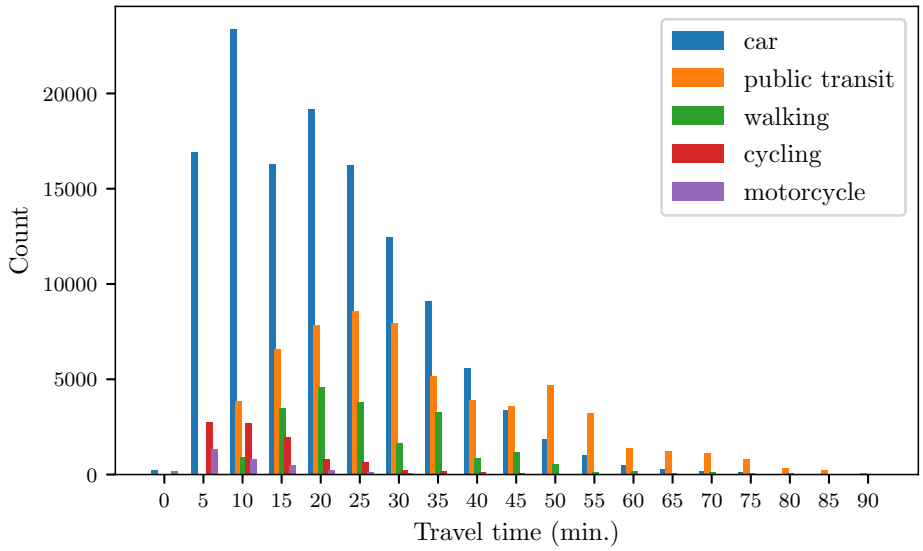


Figure 4: Mode shares as a function of travel time between origin and destination

2 Results

There are 221 571 individuals in the final dataset. The multinomial logit regression was estimated using STATA.

Table 4: Correlation of travel times for the modes of transportation

	car	public transit	walking	cycling	motorcycle
car	1	0.6095	0.8916	0.8949	0.9208
public transit	0.6095	1	0.6370	0.6325	0.6427
walking	0.8916	0.6370	1	0.9968	0.9864
cycling	0.8949	0.6325	0.9968	1	0.9866
motorcycle	0.9208	0.6427	0.9664	0.9866	1

Table 5: Multinomial logit model of mode choice

	(1) car	(2) public_transit	(3) walking	(4) cycling	(5) motorcycle
constant		2.7709***	2.8659***	1.1340***	-0.7284***
		(0.0395)	(0.0488)	(0.0509)	(0.0773)
age		-0.0150***	-0.0026***	-0.0139***	-0.0019
		(0.0008)	(0.0009)	(0.0010)	(0.0015)
woman		0.5349***	0.4361***	-0.3882***	-1.6909***
		(0.0194)	(0.0248)	(0.0242)	(0.0527)
car_per_indiv	1.2138***				
	(0.0161)				
car_per_indiv>0	1.5604***				
	(0.0245)				
occupation: farmer		-3.9054***	-1.0434***	-2.3653***	-0.8798**
		(0.4140)	(0.2012)	(0.5073)	(0.4400)
occupation: artisan		-1.7023***	-1.2153***	-0.7848***	-0.2261***
		(0.0525)	(0.0566)	(0.0651)	(0.0841)
occupation: executive		0.1522***	0.2031***	1.1710***	0.2986***
		(0.0255)	(0.0327)	(0.0337)	(0.0575)
occupation: intermediate		-0.2283***	-0.1447***	0.4259***	-0.0060
		(0.0242)	(0.0311)	(0.0349)	(0.0584)
occupation: blue-collar		-0.7579***	-0.9691***	-0.4808***	-0.0259
		(0.0318)	(0.0413)	(0.0467)	(0.0616)
travel_time	-1.6281***	-1.1746***	-2.1032***	-2.8474***	-3.2075***
	(0.0530)	(0.0480)	(0.0492)	(0.0581)	(0.0968)
travel_time × age			-0.0026**		
			(0.0010)		
travel_time × woman			-0.1134***		
			(0.0266)		
travel_time × occupation: farmer			1.1027***		
			(0.3621)		
travel_time × occupation: artisan			-0.0763		
			(0.0918)		
travel_time × occupation: executive			-0.3671***		
			(0.0354)		
travel_time × occupation: intermediate			-0.1986***		
			(0.0330)		
travel_time × occupation: blue-collar			0.2623***		
			(0.0403)		

Reference category is male employee
Travel time is expressed in hours
Standard errors are reported in parentheses
*** p<0.01, ** p<0.05, * p<0.1