Determinants for the energy, environmental and cost impacts of mobility patterns in the Lisbon Metro Area

Jérémy Daniel Diacre
jeremy.diacre@centrale-marseille.fr

Instituto Superior Técnico, Universidade de Lisboa, Portugal

December 2021

Abstract

Following the need to decarbonize and reduce the energy consumption in order to limit the effects of climate change, urban areas have begun to change their transportation policies in order to attain sustainable urban mobility. This notion ties together mobility, environment and sociology. The goal of this work is to find mobility or socio-demographic or economic indicators that have an impact on the energy consumption, the CO2 emissions as well as the cost.

To do so, data resulting from the Imob survey realized in the Lisbon Metropolitan Area in 2017 was analyzed. A new dataset of 50 city pairs was created in order to find relationships between selected independent variables and dependent variables being both total and per passenger.km energy consumption, CO2 emissions and cost. An exploratory data analysis was realized to determine which IV can fit the best in a following Multiple Linear Regression. It resulted that the share of public cars or of public transportation is almost always the most impactful indicator on per pkm DVs. It has also been determined that the share of people with a high education and the unemployment rate have an impact on the per pkm DVs. Concerning the totals DVs, as shown in the well-known IPAT identity, the population has a big impact on the number, but for this work with a surprising inverse relationship.

Keywords : Energy efficiency; Decarbonization; Multimodal emission factors;Sustainable urban mobility; AML

1. Introduction

The IPCC Assessment Report published in 2014, was already damning and warned everyone about the effect climate change would have on our lives and was followed by the historic Paris Agreement in 2015. The last update provided in 2021 by the Working Group I contribution to the sixth Assessment Report some information that previews an even worse situation than previously thought. The ideal limit of increase in temperature by 1,5°C above pre-industrialisation levels will possibly be exceeded as soon as 2025. Even with more CO_2 emissions limitations, the impact on the planet will be severe and irreversible (IPCC, 2021).

The concept of sustainability was first introduced in 1987 in the World Commission on Environment and Development, known as the Brundtland report. Sustainable Development is described as a way of development to “ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs” (WCED, 1987).

WBCSD stated in 2015 that "Sustainable mobility is the ability to meet society’s need to move freely, gain access, communicate, trade and establish relationships without sacrificing other essential human or ecological values, today or in the future.” (WBCSD, 2015)

Sustainable Urban Mobility Plans (abbreviated SUMP) were first introduced in 2013 by the European Commission. The objectives of such plans were clearly defined by the commission : “A Sustainable Urban Mobility Plan has as its central goal improving accessibility of urban areas and providing high-quality and sustainable mobility and transport to, through and within the urban area. It regards the needs of the 'functioning city' and its hinterland rather than a municipal administrative region.” (European Commission, 2013). Attributes that an urban transport system should possess to obtain sustainable urban mobility were stated in this report : accessibility, diversity of services, good integration of transport modes, sustainability, efficiency, adequate use of urban space, safety and security.

In 2014 Van Audenhove et al. sorted for Arthur D. Little and the IUTP 84 cities throughout the world based on their Urban Mobility Index 2.0 which assess the maturity and performance of a public transportation service through multiple criterias such as ‘Share of public transport in modal split’ or ‘Mean travel time to work’ for examples. While Lisbon’s network was finally ranked as average it was also the third to last ranked city in Europe only ahead of Athens and Rome (Arthur D. Little 2014).

In order to implement sustainable urban mobility, it will be necessary to understand how impact variables of mobility patterns such as energy consumption, CO_2 emissions or cost can be related to measurable indicators characterizing an urban area. This work aims to identify if mobility, socio - economic or demographic indicators have an influence on the aforementioned variables.

2. Data and methods

In 2018 the Portuguese National Institute of Statistics (INE) released the Mobility Survey in the Metropolitan Areas of Porto and Lisboa 2017 referred more commonly as Imob (Imob, 2018). The goal of Imob was to identify and quantify the trips made by the inhabitants in metropolitan areas (Porto and Lisbon). This way the travel undertaken by the population living in each region can be characterized and used for further policies. A little less than 30000 households spread across AML's eighteen municipalities were surveyed collecting around 120000 trips made by more than 60000 inhabitants of the metropolis. In order to describe precisely the transportation in the AML each household was assigned a weighing coefficient (coded as "pesofin") based on their living area that makes it correspond to the real number of people with potentially similar mobility patterns. As multiple modes of transport are being used by the inhabitants of the metropolis and are detailed in the data, it is also necessary to take into account the different occupancies of these different categories of vehicles.

Our main goal is to determine what impacts the most energy consumptions, CO2 emissions and cost of transportation. To do that we need the data corresponding to these variables. Based on a previous work (Belga, 2021), energy consumption, CO2 emissions and cost were associated with each trip.

Once the values per passenger.km of energy consumption, emissions and costs have been determined for each trip, it is possible to group trips from the same origin and destinations together in order to quantify energy and environmental impacts of each OD pair. As referred previously, each origin and destination correspond to a TAZ (traffic analysis zone). In our study, each TAZ corresponds to a municipality of the AML. Again, our goal is to evaluate the energy, emission and cost performance of the main mobility dynamics ("corridors") between the 18 municipalities and within each one. These values can be turned into an OD matrix. The figure 1 presents the OD matrix for energy consumption (MJ/pass.km). Each cell represents the energy performance of that OD pair or transport "corridor".

Municipios	Alcochete	Almada	Amadora	Barreiro	Cascais	Lisboa	Loures	Mafra	Moita	Montijo	Odivelas	Oeiras	Palmela	Seixal	Sesimbra	Setúbal	Sintra	Vila Franca De Xira	Total
Alcochete	2.28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Almada	-	1.32	-	-	-	1.07	-	-	-	-	-	-	-	1.25	-	-	-	-	3.64
Amadora	-	-	2.28	-	-	1.22	-	-	-	-	1.59	1.65	-	-	-	-	1.83	-	8.57
Barreiro	-	-	-	1.89	-	-	-	-	1.84	-	-	-	-	-	-	-	-	-	3.73
Cascais	-	-	-	-	2.12	1.56	-	-	-	-	-	2.08	-	-	-	-	1.80	-	7.56
Lisboa	-	1.06	1.24	-	1.31	1.41	1.13	-	-	-	0.88	1.52	-	1.37	-	-	1.51	1.51	12.94
Loures	-	-	-	-	-	1.12	1.82	-	-	-	1.82	-	-	-	-	-	-	1.92	6.69
Mafra	-	-	-	-	-	-	-	2.63	-	-	-	-	-	-	-	-	-	-	2.63
Moita	-	-	-	1.93	-	-	-	-	2.16	-	-	-	-	-	-	-	-	-	4.09
Montijo	-	-	-	-	-	-	-	-	-	2.30	-	-	-	-	-	-	-	-	2.30
Odivelas	-	-	-	-	-	0.83	1.06	-	-	-	2.04	-	-	-	-	-	-	-	3.94
Oeiras	-	-	2.24	-	1.80	1.74	-	-	-	-	-	2.46	-	-	-	-	1.50	-	9.74
Palmela	-	-	-	-	-	-	-	-	-	-	-	-	2.29	-	-	1.59	-	-	3.88
Seixal	-	1.18	-	-	-	1.48	-	-	-	-	-	-	-	1.73	-	-	-	-	4.40
Sesimbra	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.16	-	-	-	2.16
Setúbal	-	-	-	-	-	-	-	-	-	-	-	-	1.70	-	-	2.03	-	-	3.73
Sintra	-	-	1.76	-	1.90	1.53	-	-	-	-	-	1.32	-	-	-	-	2.25	-	8.75
Vila Franca De Xira	-	-	-	-	-	1.32	1.80	-	-	-	-	-	-	-	-	-	-	2.04	5.15
Total	-	3.55	7.51	3.82	7.13	13.30	5.81	2.63	4.00	2.30	6.34	9.03	3.99	4.35	2.16	3.62	8.88	5.46	93.88

Figure 1 - OD matrix of MJ/pkm in AML (from Belga, 2021)

In this work EDA will be used to preemptively filter the variables that won't have any impacts or tackle variables that are too similar to one another. Furthermore, the variables that will be the most useful later on can also be detected in EDA. To do so several actions will be taken. First of all, plotting the data is an elemental part of EDA and will be realized. Then a correlation table will be built using the corresponding function on the Excel software that calculates the Pearson correlation coefficient between variables. Finally p-values between the variables will be calculated.

The IPAT identity, as seen in the literature review, is an equation that quantifies the environmental impact of mankind as a function of three factors : Population, Affluence and Technology. The expression is the following :

$$I = P \times A \times T \tag{1}$$

In this work we will use total energy consumption, total CO2 emissions and total cost as the impact in the IPAT identity because this latter isn't really suited for variables per passenger.kilometer. In order to put into light an equation corresponding to IPAT with our dataset, we will need to use a Multiple Linear Regression.

The multiple linear regression is a mathematical regression method used to describe the variation of a dependent variable based on the variations of independent variables. The theoretical model is the following :

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon \tag{2}$$

Multiple linear regression is based on various assumptions that need to be tested. The regression is usually done in a software. In this work the software Rstudio will be used. The first regression that is done uses variables selected beforehand thanks to the exploratory data analysis. If the data doesn't pass the tests or isn't satisfying enough meaning that the p-value of one of the variables is too high we eliminate this variable to test another one. It is possible though to keep a variable that won't be significant in order for this model to respond better to tests.

The dataset is made with the one retrieved with Imob and complemented with Belga's data presents itself under the form of a listing of 90000 trips with details on origin, destination, trip length and more, as well as the added values for energy consumption and CO2 emissions. Ultimately the data that will be used for the correlation concerns Origin-Destination pairs rather than trips. That's why instead of studying trips separately, they will be gathered together for each OD pair they correspond to. This way we can obtain in particular the total number of trips, total number of weighed passengers, total quantity of energy consumed and total quantity of CO2 emitted. As each trip also has its transport mode indicated, it is also possible to calculate the modal share for each OD pair by simply dividing the number of weighed passengers for one transport mode by the total number of ponderated passengers for an OD pair.

In order to be able to examine what determinants can impact the dependent variables which are energy consumption, CO2 emission and cost we need to craft carefully a set of relevant and pertinent independent variables. The variables are divided into three sections : mobility, socio-demographic and -economic indicators and respondents. The set of variables used in this work contains 24 different of them but only the more relevant will be detailed here.

Mobility variables concern variables that indicate how certain aspects of mobility are shaped in the area studied. In this study the variables belonging to the mobility section are mostly linked to the modal share. Socio-demographic and economic variables combine factors that define people in a specific population on a social or economic scale. The data for the following variables are retrieved from Contemporary Portugal Database called Pordata that presents official and verified statistics about Portugal and Europe (Pordata, 2021) as well as from the government office for national statistics of Portugal INE (INE, 2021). The respondents section gathers information about the respondents of the Imob survey based on their answers.

The first and easiest way to explore the dataset is to plot variables together in order to visually analyze if they are more or less correlated, if their relationship is linear or non-linear and if their correlation corresponds to the common sense, the assumptions made beforehand as well as fellow studies. To give an example, the following figure 5 represents gCO2/pkm by % of public transportation for the Intra trips.

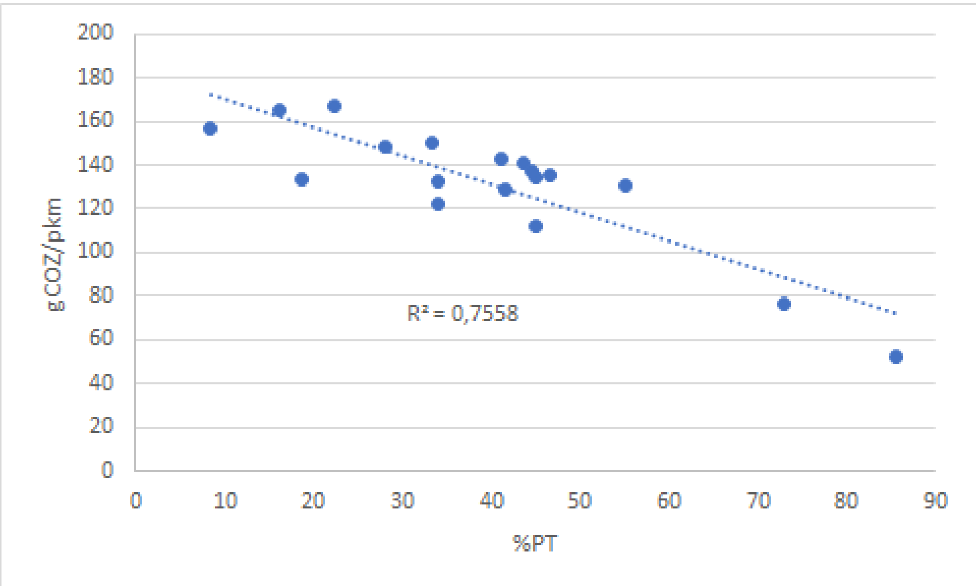

Figure 2 - gCO2/pkm by %PT plot for Intra trips

3. Results and discussion

Based on the results of the EDA , we can select several variables that will constitute the basis for our models. These variables differ between DVs per passenger.km or totals.

Table 1 - Chosen variables for DVs per pkm

Type of IV	Independent variable	r MJ/pkm	r gCO2/pkm	r cost/pkm
Mobility	% private cars	0,49474662	0,942112449	0,863295553
Socio-economic/demographic	Purchase power	-0,3809313	-0,596988719	-0,56258876
Respondents	% high education	-0,2947293	-0,443939901	-0,4706231

The three variables chosen for the base model are presented in the table 9. The variable from the mobility type is the share of private cars. The choice of this indicator makes sense as it is the one presenting one of the best correlation coefficients for each DV. Concerning the socio-economic or demographic determinant, the choice of the purchase power indicator is logical as it presents a valid correlation coefficient for each variable and has the best one of the independent variables of its type. Finally for the variables from the respondents the choice of the percentage of respondents with a high education has been made as it presents a good correlation coefficient for emissions and cost and it's the only variable of its category with a coefficient above 0,3 on absolute value. It does not however pass this threshold for energy consumption but only just barely therefore it has still been chosen. For total DVs the chosen variables will be : %PT, population, % high education.

. The first regressions are done with the variables chosen in the part before that are supposedly the best ones. After some trials and errors, the socio-economic variable that gives the better results is unemployment despite its Pearson coefficient being lower than 0,3 in absolute therefore this one will be used instead of purchase power.

The regressions gave good results for the chosen variables, except for MJ/pkm and total CO2 emissions were some reservations can be made. Here are the results for one pkm DV and one total DV :

- for cost/pkm :

Table 2 - Regression results for cost/pkm

IVs	p-value
%PT	2,02E-14
%high ed	0,00974
Unemployment	0,05949
Tests	
adjusted r-squared	0,7965
DW test	stat : 2,456588 / p-value = 0,104
Highest Condition Index	20,321857

Table 3 - Estimates and elasticities for cost/pkm

Ivs	estimate	elasticity
%PT	-1,45624	-3209,0036
%high ed	-0,33728	-320,69161
Unemployment	-3,73212	-738,96777

All of the signs are negative here meaning all IVs have inverse relationships with cost/pkm. However the share of public transportation seems more important regarding the other variables than for the two other DVs. Furthermore the importance of %high ed and unemployment are closer for cost/pkm than for the other.

- for total energy consumption :

Table 4 - Regression results for total energy consumption

IVs	p-value
%PC	0,00175
%high ed	0,03892
Population	1,23E-06
Tests	
r-squared	0,6338
DW test	stat : 1,047975 / p-value = 0
Highest Condition Index	9,576689

Table 5 - Estimates and elasticity for total energy consumption

IV	estimate	elasticity
%PC	-2,11E-02	-5,79E-02
%high ed	8,22E-06	1,64E-05
Population	-1,51E-02	-2,26E+02

The elasticity results showcase three different levels of magnitude for the Independent variables. By far the most important one is the population which as said previously is expected when working with totals as the IPAT identity can show. Then the share of public cars is way less significant than population and finally the importance of the impact of the share of people with a high education is really low.

Now that we have the results of the regression on the logarithm of total DVs we can come back to the IPAT identity. The three regressions of each total DV present similar results : big impact of the population with an inverse relationship with the DV, small impact of the share of private cars with a direct relationship with the DV and a even lower impact for the share of people with high education with an inverse relationship with the DV. The population having a big impact on the dependent variable correlates with the creation of the IPAT identity that was done to show that population also have a big impact on the human impact on the environment.

For cost/pkm and MJ/pkm we can see that the share of public transportation has a big impact on them with an inverse relationship. This means that the more the public transportation in the modal share the less energy is consumed and the less the cost per passenger and per km. Concerning the cost it is logical as the cost/pkm of pubic transportation is lower than the cost/pkm of a car regardless the mean of public transportation. The same can be said for the energy consumption as for the kind of trips studied the rail is the mean of public transportation the most used and it is also one of the most energy efficient. Concerning the emissions, it makes also sense that emissions decrease when public transportation increases in the same vein as energy however its impact is lower.

For the 3 DV per pkm the share of higher educated people has an inverse relationship with the DV, meaning that when people are more educated the energy consumption, CO2 emissions and cost per pkm are lower.

4. Conclusion and future works

The data used was retrieved from the 2017 Imob survey relaized in the Lisbon Metropolitan Area. This recorded with details on transport mode, length and duration, social and economic situation of respondents around 120000 trips (cut down to 90000) from a little less than 30000 households. The energy consumption, CO2 emissions and cost were calculated for each trip by Belga in 2021. Socio-economic and socio-demographic indicators were added to the dataset in order to create one that could answer the goal of this work after analysis. The trips were gathered in so-called city pairs that presented the sum for total variables and average for the others.

It was first determined the more interesting indicators thanks to an exploratory data analysis. The three indicators for the per pkm variables were : share of public transportation in the modal split, share of people with a high education and purchase power of inhabitants. For the total variables the share of public cars was rather used as well as the population instead of unemployment.

These variables were then used in a multiple linear regression. The first thing noteworthy was that even if it presented good results in the EDA, purchase power didn’t fit well in the model of regression. The unemployment rate was used instead beacause of the better results it gave even if this variable didn’t show up in the EDA.

Except for MJ/pkm and total CO2 emissions for which some reserves one can have, all the models found show a clear relationship between DVs and IVs. For most of them the most impactful factor is the mobility one, it being the share of cars or of public transportation, with an inverse relationship with the DV. Depending on the variables the impactfulness of the other variables can be more or less important but was always significant.

As mentioned in the work one of the future work is to tackle the Intra trips by studying them at a parish level to have a better understanding of the mobility dynamics.

We could also integrate other variables such as the opinion of inhabitants on transportation, a variable present in the Imob data but that required a little bit more work like a factorial analysis.

The process shown in this work can also be done with data from other urban areas to check if the results are similar or not and then conclude about what this means.

References

European Commission. (2013). *A Concept for Sustainable Urban Mobility Plans*.

INE website - https://ine.pt/

Instituto Nacional de Estatística (2018) “Inquérito à mobilidade (Imob) - Sessão de apresentação dos resultados definitivos”

IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

Pordata website – https://www.pordata.pt/

van Audenhove, F.-J., Korniichuk, O., Dauby, L., & Pourbaix, J. (2014). *The Future of Urban Mobility 2.0: Imperatives to Shape Extended Mobility Ecosystems of Tomorrow*.

World Commission on Environment and Development. (1987). *Our Common Future*.