Analysis of Population through Public Transportation and Yearly Income

Ekin Ilseven

Eidgenössische Technische Hochschule Zürich (ETHZ)
eilseven@student.ethz.ch

I. Introduction

Nowadays, one of the political discussions in Switzerland is about immigration. In February 2014 the majority of the Swiss people (50.3%) voted in favor of a legislative initiative which includes strict regulation of immigration to Switzerland. The surveys upon this voting showed that the main reason of the pro-voters was the insufficient infrastructure and job opportunities due to excessive immigration. The nationalist party, SVP, asserts in its party program 2011-2015 that "*(Swiss) infrastructure can not stand to the stress (of excessive immigration): jammed streets, overloaded public transport or school classes occupied mostly by foreign children are the consequences*"[1]. However further investigation indicates that an objective methodical approach to detect, to analyze and to improve this situation does not exist. The aim of this research is suggesting an objective method which enables an analysis of the population distribution according to the public transport and the standards of living. The method is based on the concepts obtained from the Social Entropy Theory (SET, *Kenneth W. Bailey, 1990*) and the main tool for calculations is the Theil index, set forth for income inequality analysis in 1967.

II. Ideal Curves

Ideal curves are the fundamentals of this research. We propose that for a certain amount of people in a particular area there must be an ideal amount of PT stops, and for a certain yearly income distribution in a particular area (L) there must be an ideal amount of people who are using PT (S). The factors P, S and L are taken from Social Entropy Theory [1].

$$P_{\text{ideal}} \leftrightarrow L_{\text{ideal}} \qquad P_{\text{ideal}} \leftrightarrow S_{\text{ideal}} \tag{1}$$

II.1 Population Density vs. Stop Density

As the first step, we define an "effective population density" (EPD_S), which indicates which proportion of the citizens in a certain area would like to actively use PT, more specifically those people who would not like to walk and prefer PT. Alternatives of PT, such as bicycle, car and etc are not included in this part, as they are luxuries that are related to the standards of living. It is assumed that EPD_S is directly proportional to stop density: the more people want to use PT, the more stops are necessary:

$$\text{EPD}_S = \pi \cdot \text{Stop Density (s)} \tag{2}$$

$$\text{Stop Density (s)} = \frac{1}{\text{Distance between stops (d)}} \tag{3}$$

π is a proportionality factor to be determined later. Care should be taken about this formulation of the stop density. The distance between stops is an averaged value. This formulation must be improved for areas with high variation in distances. The EPD_S is defined over "real population density" (RPD_S) and "passive population density" (PPD_S), which is the density

[1] Translated from german. Parentheses added for context integration. See *SVP Parteiprogramm 2011-2015, Page 54-55.*

of the people who would like to walk instead of taking PT.

$$\text{EPD}_S = \text{RPD}_S - \text{PPD}_S \qquad (4)$$

The real population density is given by the overall population density distribution and passive population density is inferred heuristically. PPD is chosen to be related to RPD by a distribution function f and the above given equations can be combined into the following one:

$$\text{RPD}_S \cdot (1 - f(d)) = \frac{\pi}{d} \qquad (5)$$

The properties of the distribution f are: (1) At a minimum distance $d_{\min}$, $f(d_{\min}) = 1$. At this distance everybody would walk. (2) At a maximum distance $d_{\max}$, $f(d_{\max}) = 0$, nobody would walk. (3) In the interval $[d_{\min}, d_{\max}]$, f decreases strictly monotone. In this research, polynomials of degree α that fulfill the properties $1 - 3$ are chosen. So the equation 5 is written as:

$$\text{RPD}_S = \frac{\pi}{d \cdot (1 - (\frac{d_{\max} - d}{d_{\max} - d_{\min}})^{\alpha})} \qquad (6)$$

or written in terms of stop density (s):

$$\text{RPD}_S = \frac{\pi \cdot s}{(1 - (\frac{d_{\max} - \frac{1}{s}}{d_{\max} - d_{\min}})^{\alpha})} \qquad (7)$$

π is determined by the total population ($\bar{P}$): $\bar{P} = \sum_{ij} RPD_S(x_{ij}) \cdot O_{ij}$, where O_{ij} is the area of the subarea (i, j) (e.g. city, canton, state). See figure 1 in appendix for the ideal population density curve depending on stop density w.r.t. different values of α.

II.1.1 Improvements to stop density curve

While this model takes only one kind of PT vehicle into account, a weighted sum of densities of various PT vehicles can be introduced in a further work, which would address the multiplicity of PT such as overlapping metro, bus and tram stops. Another point to improve could be the PPD due to bikers instead of walkers as in many european cities bikes are used frequently.

II.2 Local Population vs. Standards of Living

First of all, it is assumed that people need to travel (*Homo peregrinus*). So they will buy a transportation vehicle when possible. For quantitative analysis an effective population distribution (EPD_L) -similar to the EPD_S in equation 4- is necessary once again. The real population distribution (RPD_L) is the yearly income distribution of people and the PPD_L is subtracted from RPD_L depending on two factors that are closely related to the standards of living: (i) affordability (A), (ii) comfort (C):

$$\begin{aligned} \text{EPD(L)} &= \text{RPD(L)} - \text{PPD(L)} \\ &= \text{RPD(L)}(1 - f(A, C, L)) \end{aligned} \qquad (8)$$

The PTS's affordability and comfort will be compared with those of bicycles and cars, so we can name all of these as A_p, C_p A_b, C_b A_c, C_c respectively. While weighing the RPD, Fermi distribution combined with the Heaviside function will be used, which are ideal for the mathematization of above mentioned parameters:

$$f_i(A, C, x) = \frac{2 \cdot \theta(x > A_i)}{e^{-\frac{(A_i - x)}{C_i}} + 1} \qquad (9)$$

In this distribution x is the yearly income, $C_i \in [0, \infty]$ and $A_i \in [0, x_{\max}]$. Heaviside function is justified by the fact that there can be people using the corresponding vehicle only if they have sufficient yearly income. C equals 0 means that the corresponding vehicle is completely uncomfortable, so it is less likely that a person would buy it. As C grows, the tendency of people to buying that vehicles grows, which leads to less demand for PT. Putting equation 8 and 9 together, we arrive at:

$$\begin{aligned} \text{EPD(x)} = \text{RPD(x)} \cdot \theta(x > A_p) \cdot \\ (1 - (\frac{2 \cdot \theta(x > A_b)}{e^{-\frac{(A_b - x)}{C_b}} + 1} + \frac{2 \cdot \theta(x > A_c)}{e^{-\frac{(A_c - x)}{C_c}} + 1})) \end{aligned} \qquad (10)$$

Here it is assumed that a person who does not want to buy a bicycle or a car, but wants to travel, must use PT, so $C_p \to \infty$. See figure 2 in appendix for a hypothetical yearly income distribution and the EPD_L. Furthermore $\text{RPD}(x)$ must satisfy:

$$RPD_S = \int_0^{x_{\max}} RPD(x)dx \qquad (11)$$

At this point it must be emphasized that the EPD of (S) and that of (L) are not the same. The standards of living is a necessity for the use of PT (e.g. someone with insufficient yearly income can not afford PT). So the EPD_L is the distribution of people who can and will be using PT. It is not a density. In order to transform the EPD_L into RPD_L, we integrate the EPD_L over the all non-zero income interval $[A_p, x_{\max}]$ and then divide by the spatial area of the corresponding subarea.

$$L \to EPD_L \to \frac{d}{dS}\int_{A_p}^{x_{\max}} EPD_L dx = RPD_L$$
$$S \to RPD_S \qquad (12)$$

Now we have reached at the first aim. We have two ideal curves that relate the PT spatial structure and standards of living with the ideal amount of population density for an area (1). These ideal values can now be compared with computational or real data.

III. Negentropy, Theil index and its Application

III.1 Negentropy and Theil-index

$$H_{\text{tot}} = \sum_{ij} H(x_{ij}) \qquad (13)$$

$$H(x_{ij}) = \sum_{k\in\{S,L\}} P(x_{ij}) \ln\left(\frac{P(x_{ij})}{P_k^{\text{ideal}}(x_{ij})}\right) \qquad (14)$$

In this research we look at a country that is divided into small subareas. For simplification purposes the initial tests will be ran on a square country with square subareas (see Fig. 3 in appendix). Each subarea is given a stop density (s_{ij}) and a yearly income distribution ($L_{ij}(x)$). s_{ij} and $L_{ij}(x)$ being held constant, the country has a population density P_{ij}. For each subarea, the above given definition of negentropy (at the same time Theil index) is used to calculate how ideal the PTS is. The effective deviations from locally ideal population densities are weighed by the local population densities and summed up (Equation 14 is the contribution of a subarea (ij) to the total negentropy equation 13). A negentropy level of 0 means that the population is distributed ideally ($\ln(\frac{P(x_{ij})}{P_k^{\text{ideal}}(x_{ij})}) = \ln(1) = 0$). This definition of negentropy enables local improvement suggestions in the structure of the city, such as in population distribution or PTS distribution [2]. The subareas with H_{ij} greater than 0 implies that the subarea is overpopulated and those with H_{ij} less than 0 are underpopulated with respect to the s_{ij} and L_{ij}. Equation 14 can be analyzed in its summands (S, L) and suggestions can be more precise. In order to be able to interpret this entropy level easily, the following index $H^n \in [0, 1]$ is suggested:

$$H^n = \frac{H_{tot} - H_{min}}{H_{max} - H_{min}} \qquad (15)$$

where:

$$H_{min} = \bar{P}\ln\left(\frac{\bar{P}^2}{(\sum_j \sqrt{P_j^S P_j^L})^2}\right)$$
$$H_{max} = \bar{P}\ln\left(\frac{\bar{P}}{\min(\sqrt{P_i^S P_i^L})}\right) \qquad (16)$$

See appendix for detailed derivation. One can see that the negentropy is bounded below by $H_{min} \geq 0$. As long as the condition $\sum_j \sqrt{P_j^S P_j^L} = \bar{P}$ is not fulfilled the country can not have the ideal population distribution. A change in $H_{tot} \to H_{min}$ will improve the country's PTS, but the change in $H_{min} \to 0$ is necessary for the ideal distribution.

III.2 An Example and Analysis

The equation 17 shows the stop density in a subarea and the equation 18 the local yearly income distribution of a fictitious country with $\bar{P}$ many citizens. The stop density is taken to be constant throughout the country, while the yearly income distribution is a gaussian curve, which has the same width in every subarea, but an increasing mean value in the southeast direction.

$$s_{ij} = \rho \tag{17}$$

$$L_{ij}(x) = N_{ij} e^{-\frac{(x-\beta\sqrt{i^2+j^2})^2}{2\sigma^2}} \tag{18}$$

For this case $N_{ij} = N \;\; \forall i, j$. The population distribution is linearly increasing in the southeast direction (Eq. 19).

$$P_{ij} = \frac{\bar{P}}{O_{ij}} \cdot \frac{(i+j)}{\sum_{ij}^{N}(i+j)} \tag{19}$$

Now the H_{ij}^S, H_{ij}^L, H_{ij}, H_{tot} can be calculated by ideal population densities and equations 13 and 14. See appendix for detailed calculations and the graphs of results. This example gives an $H_{tot} = 1.5132 \times 10^4$ and $H^n = 0.0054$. According to these values, the population distribution seems to be doing well with the corresponding PTS, but it must be kept in mind that the H_{min} is far away from 0 level. So even the best population distribution will not lead to the ideal distribution. This shows us that the L_{ij} and s_{ij} must be altered in such a way that $H_{min} \to 0$, while $H_{tot} \to H_{min}$.

IV. Conclusion

In this research, an objective method of PTS quality testing is proposed. The mutables P, S and L are utilized by some micro- and macrosociological distributions and the quality of the PTS is measured by an H-index, which is based on negentropy and Theil's index. The given population distributions were successfully compared with the ideal values by using a form of the Theil index. The flexibility of this method must be underlined. We see that there are many ways to interpret results: *(i)* The role of stop density or income distribution can be inspected separately. *(ii)* We have an index that is easy to interpret as it is between 0 and 1. *(iii)* The H_{min} tells in which way the stop density or income distribution should be improved. *(iv)* Local improvements in population distribution can be suggested. Next to these few methodical ones, further interpretations can be found. The allocation of immigrants in the yearly income distribution can be used to justify or disprove the suggestions mentioned in the introduction. Looking at the current situation of Switzerland by using this method can expand the solution possibilities to the overload in PT (such as providing incentives for relocating citizens from overpopulated cantons to the underpopulated ones). It must be pointed out that this model is based on very simple few assumptions, which makes it too general to be realistic. For example, a problem of this model is the high end tail of the yearly income distribution and EPD based on it are not realistic. However with further improvements, it could lead to an efficient way of inspecting PTS. As the next step, the theory should be tested with real data.

References

[1] Kenneth D. Bailey "Social Entropy Theory", *SUNY Press*, ISBN 0-7914-0056-5 (1990)

[2] Conceicao, Pedro and Ferreira, Pedro "The Young Person's Guide to the Theil Index: Suggesting Intuitive Interpretations and Exploring Analytical Applications" *UTIP Working Paper No. 14* (February 29, 2000)