

UrbanSim Capstone  
Indicators for Equity and Welfare Project  
Computer Science and Engineering  
&  
Daniel J. Evans School of Public Affairs  
University of Washington  
Final Report

Colleen Chrisinger  
Eric Choi  
Erin Covington  
Jann Röder

June 5, 2007

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Project Goals</b>	<b>2</b>
<b>3</b>	<b>Project Design</b>	<b>2</b>
3.1	Lorenz Curve and Gini Coefficient . . . . .	3
3.1.1	The Lorenz Curve . . . . .	3
3.1.2	The Gini Coefficient . . . . .	4
3.1.3	Formulas . . . . .	4
3.1.4	Integration into UrbanSim . . . . .	5
3.1.5	Limitations . . . . .	5
3.2	Residential Composition Indices . . . . .	6
3.2.1	The Dissimilarity Index . . . . .	6
3.2.2	The Interaction Index . . . . .	7
3.2.3	The Relative Concentration Index . . . . .	8
3.2.4	The Absolute Centralization Index . . . . .	8
3.2.5	The Spatial Proximity Index . . . . .	9
3.3	Welfare Index . . . . .	10
3.3.1	Concept . . . . .	10
3.3.2	Inputs . . . . .	10
3.3.3	Formula . . . . .	11
3.3.4	Assumptions . . . . .	12
3.3.5	General Procedure . . . . .	12
3.3.6	Challenges and Limitations . . . . .	12
3.4	The Reldist package . . . . .	13
3.4.1	Integration into UrbanSim . . . . .	13
3.4.2	Interpreting the plot . . . . .	14
3.4.3	Limitations . . . . .	15
<b>4</b>	<b>Project Testing</b>	<b>15</b>
<b>5</b>	<b>Issues of Professional and Ethical Responsibility</b>	<b>16</b>
<b>6</b>	<b>Project Impacts</b>	<b>17</b>

# 1 Introduction

This project was an undertaking in real-world software design in the context of the open-source UrbanSim transportation and land use modeling system.

The quarter long Capstone class combined students from the Computer Science and Engineering department with a Ph.D. Candidate in Public Policy and Management from Daniel J. Evans School of Public Affairs to form an multi-disciplinary team. During the course of our multidisciplinary team collaboration we used our team's skills and knowledge in the fields of computer science, economics and public policy to make a real contribution to the UrbanSim project and to use our knowledge within the context of another discipline.

The content of this report will discuss our project's design, individual components, testing, accomplishments and ethical considerations.

## 2 Project Goals

1. To assess the incorporation of the indicators of welfare and equity within the OPUS system.
2. To build a framework that allows flexible equity and welfare analysis.
3. To examine the Reldist module written by Mark Handcock[3] and determine the best way to integrate this code with OPUS.

## 3 Project Design

The functional components of our project are written in Python and are designed to match existing code in UrbanSim. Each component was delivered with documentation, created using L<sup>A</sup>T<sub>E</sub>X<sup>1</sup>. The documentation covers the components features, use, derivation and known limitations.

The Lorenz curve, Residential Indices, and Reldist plot were initially written as stand-alone Python files. These stand alone files were tested with known data using Python *unit tests* to ensure their accuracy. Once it was determined that the code

---

<sup>1</sup><http://www.latex-project.org/>

was correct, we integrated the code into the OPUS Core.

The integration of our stand-alone code required us to become familiar with how indicators and variables are constructed in OPUS and what their dependencies and inherited methods are. Once we determined the correct way to integrate our code, we used the Wing IDE<sup>2</sup> in the debug mode to determine how similarly structured code already in the OPUS Core behaved. We used this as a guide for integrating our own code. After the integration was complete we tested the functionality by viewing the results in various forms: Lorenz Curve, Matplotlib map, etc.

UrbanSim advocates a style of programming in which the test cases are written before any code. While we did not strictly adhere to this method, all tests were complete before the integration into OPUS. Our testing procedure is explained in greater detail in section 4.

The following sections describe in detail the accomplishments of our project.

## **3.1 Lorenz Curve and Gini Coefficient**

### **3.1.1 The Lorenz Curve**

The Lorenz curve is a graph which represents the distribution of some variable over a population[1]. It is the graphical representation of the cumulative distribution function of a probability function. One of its most common uses is to measure inequality, whether it be income, transportation or location inequality. The distribution of any variable can be measured, but the result is not always useful.

Typically, the population is plotted on the x axis and the variable is plotted on the y axis. The Lorenz curve can then be read “the bottom x% of the population has y% of the total amount of the variable.” One special instance of the Lorenz curve is the line of perfect equality, which is defined by the function  $y = x$ . The line of perfect inequality, in contrast, would show a perfectly unequal distribution in which one person has all the income. The curve would be at  $y = 0$  for all  $x < 100\%$  and at  $y = 100\%$  for  $x = 100\%$ .

---

<sup>2</sup><http://www.wingware.com/wingide>

### 3.1.2 The Gini Coefficient

Another measure can be derived from the Lorenz curve, called the Gini coefficient. The Gini coefficient is the ratio of the area between the line of perfect equality and the Lorenz curve to the total area underneath the line of perfect equality[1]. Thus, the Gini coefficient is a scalar which also measures inequality.

Some important aspects of the Gini coefficient are:

1. It can be used to measure the inequality of any variable, but the result is not always useful.
2. It is not affected by the shape of the Lorenz curve, only by the ratio of the areas used to compute it.
3. It does not indicate how the inequality is distributed, only the total amount of inequality.

### 3.1.3 Formulas

The Lorenz curve can often be represented by a function  $L(F)$ , where  $F$  is the horizontal axis, and  $L$  is the vertical axis.

For a population of size  $n$ , with a sequence of values  $y_i$ ,  $i = 1$  to  $n$ , that are indexed in non-decreasing order ( $y_i \leq y_{i+1}$ ), the Lorenz curve is the continuous piecewise linear function connecting the points  $(F_i, L_i)$ ,  $i = 0$  to  $n$ , where  $F_0 = 0$ ,  $L_0 = 0$ , and for  $i = 1$  to  $n$ :

$$\begin{aligned}F_i &= i/n \\S_i &= \sum_{j=1}^i y_j \\L_i &= S_i/S_n\end{aligned}$$

The Gini coefficient is defined as a ratio of the areas on the Lorenz curve diagram. If the area between the line of perfect equality and Lorenz curve is  $A$ , and the area under the Lorenz curve is  $B$ , then the Gini coefficient is  $A/(A+B)$ . Since  $A+B=0.5$ , the Gini coefficient,  $G = A/(.5) = 2A = 1 - 2B$ . If the Lorenz curve is represented by the function  $Y = L(X)$ , the value of  $B$  can be found with integration and:

$$G = 1 - 2 \int_0^1 L(X) dX$$

### 3.1.4 Integration into UrbanSim

The Lorenz curve was added to the OPUS core as an image type for indicators. This is the best location for the Lorenz curve because of its flexibility. First, it is a graphical measure, and second, it can measure the distribution of any variable, regardless of which package or dataset the variable is taken from. It is up to the user to interpret the results. Because the Gini coefficient is calculated directly from the points on the Lorenz curve, it is included as part of the graphical output. Figure 1 illustrates the output of the Lorenz curve image type.

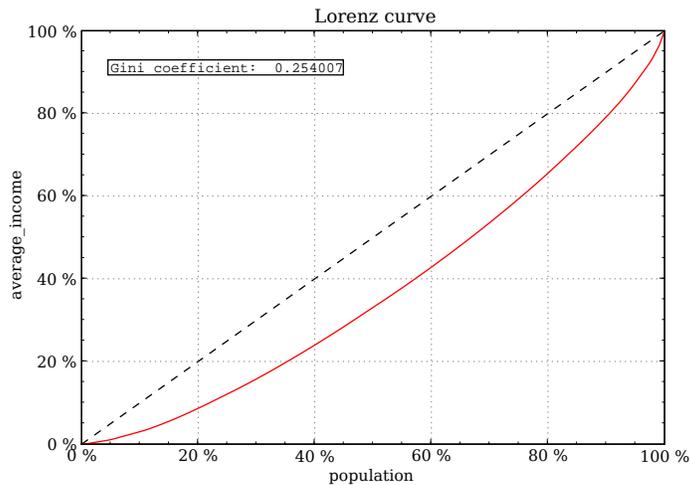


Figure 1: A sample Lorenz curve and Gini coefficient for average income in the Eugene dataset.

In figure 1, the population of the example is plotted on the x axis from 0% to 100%, and the variable, `average_income`, is plotted on the y axis, also from 0% to 100%. The line of perfect equality, because of its significance as a baseline function, is always displayed as a dashed line. The red line is not a line, but actually a set of thousands of discrete points because the amount of data is finite. These discrete points are integrated numerically to find the Gini coefficient, which is displayed in the top left corner.

### 3.1.5 Limitations

There are some of the limitations to be aware of when using these indicators:

1. The Lorenz curve may understate the actual amount of inequality if richer households are able to use income more efficiently than lower income households. From another point of view, measured inequality may be the result of more or less efficient use of household incomes.
2. The Gini coefficient is influenced by the granularity of the measurements. For example, five 20% quantiles (low granularity) will yield a lower Gini coefficient than twenty 5% quantiles (high granularity) taken from the same distribution<sup>3</sup>.

## 3.2 Residential Composition Indices

The Indices of Residential Composition are an additional measure of examining the impact of policy decisions in UrbanSim. Examining the changes in how a population is distributed in a geographic area over time can provide important information, and also facilitate the detection of anomalies in data.

These indices are based on the work of sociologists Douglas Massey and Nancy Denton,[8] who have identified five dimensions on which to consider the geographic distribution of populations: evenness, exposure, concentration, centralization and clustering. Each of these dimensions has been measured in a number of ways, but can be described adequately with only one indicator, according to Massey and Denton[8].

One common use of the indices of residential composition is to show the distribution of racial and ethnic groups in an area. Massey and Denton[8] often refer to these as “Segregation Indices” because of this common use. We elected to use a different label for these terms due to the many types of groups that can be looked at with these indicators and the connotation of segregation pertaining to race alone. Examples of these other applications include considering the distribution of groups categorized by income, age, family size or type, marital status, employment status, educational attainment, or gender. The indices can also be used to examine the distribution of different types of businesses, jobs, or industries such as retail, professional, or manufacturing.

### 3.2.1 The Dissimilarity Index

The dimension of evenness considers whether members of a certain population group are spread equally across the units of a geographical area. If some units contain a

---

<sup>3</sup>[http://en.wikipedia.org/wiki/Gini\\_coefficient](http://en.wikipedia.org/wiki/Gini_coefficient)

larger share of the group than others, the distribution is thought to be uneven. The index of dissimilarity measures the extent of this unevenness. The index takes values from 0 to 1, with 0 representing complete evenness and 1 representing complete unevenness. It can also be thought of as the fraction of the group members who would need to move in order for the proportions to be equal across geographic units.

The index of dissimilarity is a weighted average of the deviation of the group proportion in each unit from the group proportion across the area. It can be represented by the function:

$$D = \sum_{i=1}^n [t_i | p_i - P | / 2TP(1 - P)]$$

where  $i$  is the identifier for each geographic unit from 1 to  $n$  within the area. The proportion of the group being considered in the geographic unit  $i$  (relative to the total in that unit) is denoted  $p_i$ , while  $P$  is the average proportion of the group in the entire area. The total population in each unit is  $t_i$ , and in the entire area is  $T$ .

### 3.2.2 The Interaction Index

The exposure dimension focuses on the likely experiences of a certain population group when living amongst another group. If the number of people in the selected group is relatively small, its members will be more likely to encounter members of other groups through the process of residing in the same neighborhood.

The interaction index for groups X and Y measures the extent to which members of group X are exposed to members of group Y. Typically, the population of group X is smaller than for group Y. The following formula indicates the procedure through which the interaction index is computed. Within each geographic unit  $i$ , let  $x_i$ ,  $y_i$ , and  $t_i$  denote the number of X, Y and total members in that unit, respectively. The fraction of group X members in each unit is multiplied by the share of group Y relative to the total in that unit. The product of these fractions is then added over all geographic units to provide an overall indicator of the probability that any member of group X lives in the same unit as a member of group Y. The index ranges from 0 to 1, with 0 indicating minimal exposure and 1 indicating complete exposure.

$${}_xP_y^* = \sum_{i=1}^n [x_i/X][y_i/t_i]$$

### 3.2.3 The Relative Concentration Index

Concentration as a dimension of residential composition describes of the use of physical space in a geographic area. Concentration is high if members of group X occupy only a few areas which are small in size. This relative confinement could, for example, be the result of official or unofficial practices such as zoning or discrimination, or perhaps exist for reasons of convenience and mutual support.

The relative concentration index can take values in the range of -1 to 1. It is computed with the following formula.

$$RCO = \frac{\left\{ \left[ \sum_{i=1}^n (x_i a_i / X) \right] / \left[ \sum_{i=1}^n (y_i a_i / Y) \right] - 1 \right\}}{\left\{ \left[ \sum_{i=1}^{n_1} (t_i a_i / T_1) \right] / \left[ \sum_{i=n_2}^n (t_i a_i / T_2) \right] - 1 \right\}}$$

In the same way as for the other measures,  $x_i$ ,  $y_i$ , and  $t_i$  denote the number of X, Y and total members in geographic unit  $i$ , respectively. Also, this measure requires that geographic units be ordered from smallest to largest size in area. The area in each unit is denoted  $a_i$ . The number  $n_1$  is the rank in the ordering of the unit for which the cumulative total population of all the smaller units is equal to the total number of group X members in the entire geographic area.

The number  $n_2$  is the rank of the unit for which the cumulative total population of all the smaller units is equal to the number of group X members in all of the units larger than  $n_2$ , including unit  $n_2$  itself. The total population in units 1 to  $n_1$  is denoted  $T_1$ , and the total population in units  $n_2$  to  $n$  is denoted  $T_2$ .

The RCO compares the ratio of X and Y group concentrations to the ratio that would result if X were maximally concentrated and Y minimally so. A value of -1 means that group Y's concentration is as high as possible compared to group X, while a value of 1 means X's concentration is as high as possible relative to Y's concentration. A score of 0 suggests that the two groups are equally concentrated.

### 3.2.4 The Absolute Centralization Index

The centralization dimension measures the extent to which a group lives close to or far from the central place of a geographic area. The central place might be the downtown area or central business district (CBD), for example.

The absolute centralization index is computed as follows. The units are ordered by area.  $A_i$  denotes the cumulative proportion of the entire land area that is located in unit  $i$  and all the units smaller than  $i$ . The cumulative fraction of group X members in those units is denoted  $X_i$ . The next smallest unit's area and fraction are denoted  $A_{i-1}$  and  $X_{i-1}$ , respectively. The summed products are then compared according to the formula:

$$ACE = \left( \sum_{i=1}^n X_{i-1} A_i \right) - \left( \sum_{i=1}^n X_i A_{i-1} \right)$$

In practice, the first term above is implemented in UrbanSim as the sum of the first  $n - 1$  population fractions multiplied by the cumulative area of the next largest units, leaving out the cumulative population fraction of the last and largest unit. If  $n=4$ , it would be  $X_1 A_2 + X_2 A_3 + X_3 A_4$ . The second term is the opposite: the first  $n - 1$  area units are multiplied by the next larger population unit, such as  $X_2 A_1 + X_3 A_2 + X_4 A_3$ .

The range is from -1 to 1. An ACE value of 0 suggests that the members of group X are spread evenly across the area, while a positive score would indicate that members live close to the central place and a negative score indicates living further from the central place. One interpretation of the ACE is the fraction of group members that would be required to move in order to achieve a uniform geographic distribution.

### 3.2.5 The Spatial Proximity Index

The clustering dimension measures whether the units containing large fractions of group X are themselves located close together within the area. If so, the population is thought to be highly clustered.

Massey and Denton cite White's index of spatial proximity as a good measure for clustering. It requires some kind of data on the distances between each geographic unit in the area. The distance between units  $i$  and  $j$  is denoted  $d_{ij}$ . This distance is then transformed by the exponential function to create  $c_{ij} = \exp(-d_{ij})$ . An intermediary calculation is required for each group. For group X, it is as follows:

$$P_{xx} = \frac{\sum_{i=1}^n \sum_{j=1}^n x_i x_j c_{ij}}{X^2}$$

Analogously, one computes  $P_{yy}$  and  $P_{tt}$  for group Y and the total population  $T$  of the area. The index of spatial proximity is a weighted average of the proximities for

groups X and Y, calculated as follows. The total population is computed as the sum of the two group populations so that groups can still be compared even if they do not sum to the actual total in the population.

$$SP = (XP_{xx} + YP_{yy})/TP_{tt}$$

The index takes the value 1 if the groups are not differentially clustered, and is greater than 1 if the units with larger numbers of group X or Y members are closer to each other than to those units with larger numbers of the other group. It is only less than 1 in the event that members of opposite groups are located closer to each other than to members of their own group.

### 3.3 Welfare Index

#### 3.3.1 Concept

This welfare indicator is based on the idea of compensating variation (CV). The purpose of compensating variation is to measure the amount of money with which an individual would need to be compensated in order to return to the same satisfaction level that he or she had before a policy was enacted. This amount can be positive if the policy is harmful, or negative if the change is beneficial. As an aggregate social measure, CV is added up for all individuals in order to compare alternative policies. The policy with the lowest total compensating variation can be thought of as the least harmful or most beneficial, relative to the original utility levels of members of society.

#### 3.3.2 Inputs

During the estimation process, UrbanSim calculates the utility that each household would enjoy under various alternative policy scenarios. These calculations fall within the general notion of discrete choice modeling and the specific method of random utility maximization (RUM). Random utility maximization assumes that the utility,  $v_{i,k}$ , for each household,  $i$ , and under each alternative,  $k$ , is a linear function of the characteristics of the household and the characteristics of the policy alternative, where such characteristics are denoted by  $x$  in the example equation that follows.

$$v_{i,k} = \beta_a x_{i,a} + \beta_b x_{i,b} + \beta_{c,k} x_{c,k}$$

The  $\beta$  coefficients are estimated by maximum likelihood, a method that finds the values which are most likely to be correct given the observed data. The utility  $v_{i,k}$  is sometimes called a representative utility because each household's true utility level  $U_{i,k}$  is not fully captured by the above specification. An error term  $\epsilon_{i,k}$  measures the discrepancy. This disturbance is assumed to be distributed in a certain manner, usually according to the Gumbel (a.k.a. type I extreme value) distribution. Gumbel is a subset of generalized extreme value (GEV).

$$U_{i,k} = v_{i,k} + \epsilon_{i,k}$$

One of the primary inputs into the welfare calculation is the representative utility for each household and alternative,  $v_{i,k}$ . Another important input is the marginal utility of money. This is the  $\beta$  coefficient on the income term in the utility function. As explained in the section on assumptions, the term must be linear in income in order for the results to be valid. A ratio would not be sufficient.

### 3.3.3 Formula

Formula to be Calculated: Following Joel Franklin's dissertation[2] as well as Small & Rosen[10], each household's expected compensating variation when moving from policy 0 to policy 1 is:

$$E(CV^{0,1}) = \frac{1}{\lambda} [\log \sum_{k \in A} e^{v_{i,k}^1} - \log \sum_{k \in A} e^{v_{i,k}^0}]$$

Only one pair of scenarios can be considered at a time. The utility amounts are converted into dollar amounts through multiplication by the inverse of  $\lambda$ , which is the marginal utility of money. In Joel's framework, households have a choice of travel modes (alternatives  $k$  within the set  $A$ ) under each policy scenario, which is the reason for the summation across mode alternatives. In a more general analysis, there may be other dimensions of choice on the part of households, and thus more summations. We should define the policy scenarios and household- or business-level alternatives carefully to determine which ones we want to compare. An example of a policy scenario would be the construction of a new freeway. That doesn't necessarily change the  $\beta$  coefficients, but it might change the characteristics of the policy alternatives in the utility formula above, and lead one choice to become more attractive than another.

### 3.3.4 Assumptions

The previously described method can only be applied to discrete choice models that are specified in a certain way. He suggests we first create a utility that checks that the model specification conforms to these restrictions and sends a warning to the user when one of the restrictions is violated. The restrictions are[5]:

A.1 Additive disturbances; i.e.,  $U_j = V(y - p_j, q_j) + \epsilon_j$ ,

A.2 GEV (generalized extreme value) disturbances, and

A.3 Constant Marginal Utility of Income; i.e.,  $V(y - p_j, q_j) = a(y - p_j) + f(q_j)$ .

### 3.3.5 General Procedure

Because the Welfare Indices were not implemented, we felt that it was important to outline our work in the hopes that this would project would be completed at some time.

General Steps:

1. If assumptions are not satisfied, generate a warning message and stop. Otherwise, continue.
2. Obtain representative utility and marginal utility of money amounts for each household under each scenario and alternative. Save these to the record for each household. This requires running UrbanSim for two different scenarios and then being able to retrieve data from both of them. The same number of alternatives should be considered for each scenario in order to make fair comparisons.
3. Sum across alternatives and take logs according to the formula for expected compensating variation.
4. Once CVs have been calculated for each household, sum across all households. Provide this information to the user and specify which scenarios were compared

### 3.3.6 Challenges and Limitations

Implementation of the preceding steps necessitates the development of new features within UrbanSim, particularly for purposes of comparing the results of two separate computations under different scenarios. Also, Paul Waddell and Joel Franklin have described the importance of comparing results of similar procedures for each scenario. Since the choice alternatives are sampled rather than fully enumerated, it is likely

that the choices available to any given household in each model run will differ. The variation in particular choices used for estimation arguably is not a problem by itself, according to Paul as well as David Layton. However, the number of alternatives sampled certainly needs to be the same across the scenarios. The reason for this is that the measure is based on the sum across alternatives, and so is sensitive to the number of alternatives considered. Finally, a variable for income needs to be included in any choice model, in order to obtain the marginal utility of money. The variable can be income minus housing or other costs, but should not be a fraction such as housing cost divided by income. The marginal utility of money should be the same across the two model runs.

### 3.4 The Reldist package

The Reldist package<sup>4</sup> is an extension for the open source statistics software R<sup>5</sup>. It provides functions for the comparison of distributions and was written by Mark S. Handcock of the University of Washington. Detailed information about the functions implemented in the Reldist package can be found in his book[3].

We integrated the Reldist package using a R wrapper for Python called RPy<sup>6</sup>. UrbanSim is currently in the process of converting from numarray (a re-implementation of an older Python array module called Numeric) to Numpy. This conversion required us to use the current version from the subversion repository, since the latest released version doesn't have complete support for Numpy<sup>7</sup>, the numeric package that is used in UrbanSim. The current implementation uses the function called "reldist". It can be used to compare two distributions. For example: in the context of UrbanSim this could be the distributions of the average income on a zone level for the years 1980 and 1982.

#### 3.4.1 Integration into UrbanSim

The integration of the Reldist package into UrbanSim is very similar to the integration of the Lorenz curve plot. It can be used like any other plot type. If it is given one year as input, the given year's distribution is compared to the perfectly equal distribution. It must be noted that this is not a significant comparison for

---

<sup>4</sup><http://cran.r-project.org/src/contrib/Descriptions/reldist.html>

<sup>5</sup><http://www.r-project.org>

<sup>6</sup><http://rpy.sourceforge.net>

<sup>7</sup><http://numpy.scipy.org>

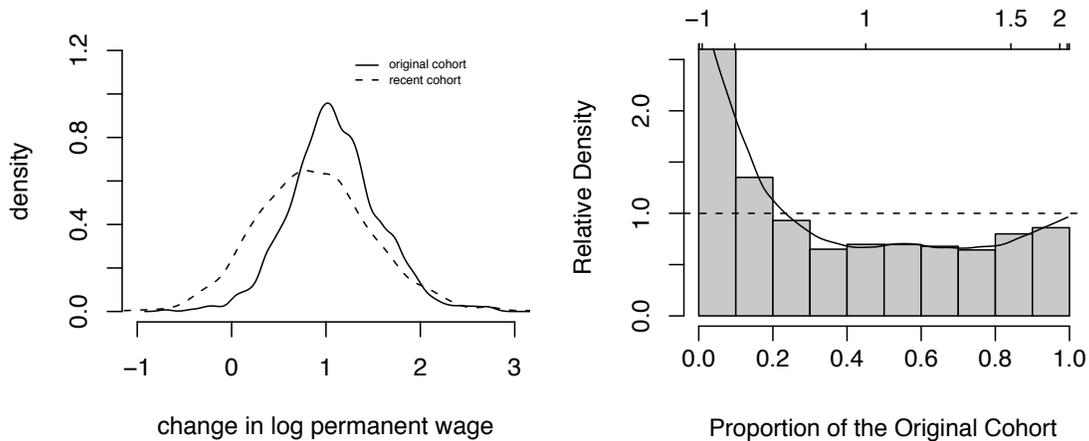


Figure 1: The two distributions (left) and the corresponding Reldist plot (right)

all indicators in UrbanSim. If it is given two years, the distributions of the given variable in the two years is compared. For an example of this see the next section.

### 3.4.2 Interpreting the plot

Interpreting the produced plot is not straightforward. The example in Figure 1 was taken from a paper by Mark S. Handcock[4]. One can see that in areas where the recent cohort curve in the distribution is above the original cohort curve, the curve in the Reldist plot is above the 1.0 line. In areas where it is the other way around, the curve in the Reldist plot is below the 1.0 line. The significance of the plot is to demonstrate how the distribution has shifted. In this case the plot would show that there are more people with lower wages, fewer people with medium wages and about the same amount of people with high wages.

However if we produce a plot from the UrbanSim data (Eugene dataset) the plot looks a bit different. See Figure 2.

The two plots show the relative distribution of the average income in the years 1981 and 1982. The only difference is the smoothing parameter to the Reldist function. On the left plot it was set to 0.35, on the right plot it was left out which causes the Reldist function to use the value that minimizes the GCV. The plot with smoothing factor set to 0.35 (left) is much closer to the actual data than the one on the right. With a little bit of imagination one can see how you can get the right plot from the left one by doing a lot of smoothing. It has yet to be determined why the data

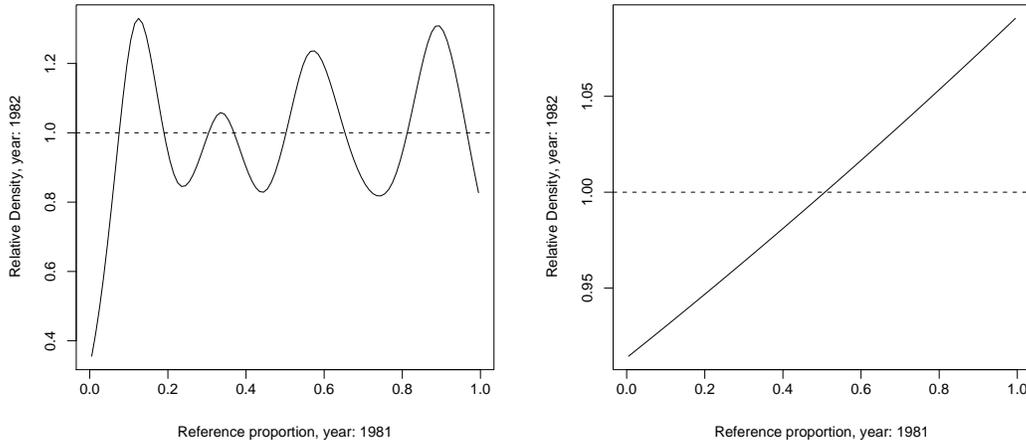


Figure 2: Reldist plots of UrbanSim data. Average income on zone level in the Eugene dataset

from UrbanSim seems so noisy. An explanation might be that there are hardly any changes, and the Reldist package is simply exaggerating very small changes. In any case the Reldist plot seems to be a good way to detect anomalies in the data.

### 3.4.3 Limitations

There are currently two limitations in the Reldist integration:

1. It is not possible to compare distributions that come from different scenarios. Only different years of the same scenario can be compared.
2. It is not possible to create plots on the gridcell level, because R crashes if you attempt to do this. The reasons for this are probably lie inside the Reldist package, e.g. it can't handle datasets that big.

## 4 Project Testing

As described in the 'Project Design' section, we used Python *unit tests* for testing our code. Each Python file was written as stand-alone code containing specific unit tests that provide known input to use in the computation, and compare the output

of the computation with the known correct value.

After determining that the code functioned as expected we integrated it into OPUS and used the data provided from the Eugene cache as input to our methods. In order to test this code we stepped through the execution with the Wing IDE and examined the behavior at various breakpoints. The visual output was also checked for correctness by examining the values returned and ensuring that the visual output was consistent with these values.

As a final step in ensuring the correctness of our code, the Lorenz curve was subjected to a code review.

## 5 Issues of Professional and Ethical Responsibility

Our project's goal was to make a meaningful contribution to the UrbanSim code base. In that way we needed to be aware of the issues of professional and ethical responsibilities that accompany this project. UrbanSim is a tool currently being used in making policy decisions. It is possible that our contribution to UrbanSim could be used in shaping a policy and as such we have a responsibility to ensure that our code is correct, freely available for inspection, and that our resources and assumptions are well documented.

The documentation provided with each of our indicators specifies the mathematical formulas used in the computation and the resources we used to justify these formulas. While this mathematical documentation is essential to legitimize our algorithm, it is also our professional and ethical responsibility to document other less technical details. For this reason, in our documentation we explicitly state known limitations and what our assumptions about the kind of input that will be used are.

Modeled after other aspects of UrbanSim, the data to be evaluated is chosen exclusively by the user. The output of the indicators are displayed visually with the scale and units clearly marked. The significance of this is that we are not interpreting the results for the user, but merely displaying them.

## 6 Project Impacts

The impact of UrbanSim in general is that it provides a transparent way to view the effects of policy change. The flexibility built into UrbanSim provides stakeholders to look at those effects based on a set of values that they can choose. For example it may be very important in a community to ensure that bicycle routes are not negatively impacted, or that open space remain at a specific percent of the total area. The user can then generate indicators that track what impact a policy would have on the percentage of open space, or the net loss/gain in bicycle routes.

Our project's contribution to UrbanSim provides the user with another way to visualize the indicators that are outputs of the model. For example the Lorenz curve can be modeled to look at how the distribution of wealth shifts from year to year under different scenarios. This can help analyze trends and compare different scenarios.

The idea of debating the implementation of various policies is a contemporary issue. Currently in Seattle, there is a debate regarding the Alaskan Way Viaduct. Various plans have been proposed including: rebuilding the viaduct, creating a tunnel or removing it completely. It is essential, for meaningful debate, to be able to predict the long-term consequences of each of these actions. In this way tools like UrbanSim can help to generate indicators based on the different scenarios. These indicators can then be compared. This assists policy makers and citizens in supporting the plan that best fits with their values.

By adding to the tools that allow a community to compare scenarios in ways that reflect their values we have made an impact in these decision making processes.

## References

- [1] Cowell, Frank A. *Measuring Inequality*. Oxford: P. Allan, (1977).
- [2] Franklin, J. *The distributional effects of transportation policies: the case of a bridge toll for Seattle*. Doctoral dissertation. University of Washington. (2006).
- [3] Handcock, Mark S. and Morris, Martina. *Relative Distribution Methods in the Social Sciences* (1999), Springer-Verlag, ISBN 0387987789
- [4] Handcock, Mark S. and Aldrich, Eric M. Working Paper no. 27, "Applying Relative Distribution Methods in R" (December 2002)

- [5] Herriges, J. A. and Kling, C. L. Nonlinear income effects in random utility models. *The Review of Economics and Statistics*, 81(1):62-72. (1999).
- [6] Jara-Diaz, S. and Martinez, F. On the specification of indirect utility and willingness to pay for discrete residential location models. *Journal of Regional Science*, 39(4): 675-688. (1999).
- [7] Lambert, P. J. *The Distribution and Redistribution of Income*. Manchester University Press. (1989).
- [8] Massey, D. and Denton, N. The dimensions of residential segregation. *Social Forces*, 67(2): 281-315. (1988).
- [9] Morey, E. R., Sharma, V. R., and Karlstrm, A. A simple method of incorporating income effects into logit and nested-logit models: Theory and application. *American Journal of Agricultural Economics*, 85(1):248-253. (2003).
- [10] Small, K. A. and Rosen, H. S. Applied welfare economics with discrete choice models. *Econometrica*, 49(1). (1981).
- [11] Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.(2003).