

# City growth and Zipf's law

**Delft University of  
Technology**

**Faculty of Technology,  
Policy and Management**

**Jaffalaan 5,**

**2628BX, Delft**

**01-Feb-15**

**Matthijs Brouns 1505157**

**Hélène van Heijningen 4076850**

---

## Preface

As part of the course ‘Agent Based Modelling of Complex Adaptive Systems – Advanced’ (SPM9555), we chose to work on a project together with the CPB, or the Netherlands Bureau for Economic Policy Analysis, where we aimed to increase understanding of the emergence of the Zipf’s law in city growth by using the method of agent based modelling. NetLogo has been chosen as the modelling environment for this specific project as it is required by the course for which this project has been conducted. The focus of this research is decision making on household level, in order to see how these decisions influence moving behaviour between cities and therefore city growth.

For more information on the model files and other documentation we would like to refer to <https://github.com/MBrouns/Zipfs-Law-and-city-development>.

---

## *Executive Summary*

Empirical research has shown that throughout history, the distribution of the sizes of the largest cities in a nation often follows a Zipfian distribution, which is a specific power law distribution. The main driver behind the emergence of this specific distribution however is still unknown. Gaining more insight in the emergence of Zipf's law can help policy makers in various domains such as special planning.

The main goal of this research was to determine to what extent decisions made at household level influence moving behaviour between cities to cause the emergence of the Zipf's law?

An agent-based model of city migrations based on household decision making was created and validated. Exploratory Model Analysis on this model shows that two factors - the distance people are willing to move and the increase in city attractiveness as its size increases - are important factors in the emergence of a Zipf's law. However, the sensitivity of the model in producing a Zipf's law indicates that this model cannot be used to fully explain the emergence of a Zipf's law as history shows that the behavior should be rather robust.

The model can be improved in several areas, most notably further improvements with regards to the migration rules of households, as well as various speed improvements.

---

## Table of contents

1. Introduction .....	4
2. System analysis .....	5
2.1 Life stages .....	5
2.3 Jobs .....	6
2.4 Housing .....	6
2.5 Other concepts .....	6
3. Narrative .....	7
4. Model logic.....	8
4.1 Household agent .....	8
4.2 Model setup .....	8
4.3 Part 1: Population model.....	8
4.4 Part 2: Moving behaviour.....	9
5. Verification.....	12
5.1 Single-agent verification .....	12
5.2 Minimal interaction level.....	13
5.3 Multi-agent level.....	14
5.4 Conclusion .....	16
6. Validation .....	17
6.1 Literature validation .....	17
6.2 Face validation through expert consultation .....	17
7. Model experimentation.....	17
8. Conclusions and Future Research .....	19
9. Reflection.....	20
9. References .....	21
Appendix A: Job attractiveness graphs .....	22
Appendix B: Model experimentation .....	25

---

## 1. Introduction

Competition between countries is becoming more and more about competition between their cities. This can be attributed to the fact that economic activity is very concentrated in large cities. Most countries around the world have a city which is by far the biggest and economically the most dominant city while the next biggest city is quite a bit smaller (BRON). It turns out that the distribution of city sizes within a country often follows the Zipfian distribution, that is, the size of any city is inversely proportional to its ranking in the list of city ranked by size (Gaujal et al., 2014).

How this Zipfian distribution emerges however has not yet been agreed on. There are many ideas on how cities form such a distribution, many of which rely on economic theories. There are of course other ways to approach this phenomenon. The Netherlands Bureau for Economic Policy Analysis (CPB) for example is interested in how the emergence of the Zipf's law could be influenced by policy making and inversely, how policy making is being influenced by this emerging behaviour. To be able to answer these questions they want to gain more understanding about the emergence of the Zipf's law. More specifically, they want to find out if decisions made at household level can explain the emergence of the Zipf's law.

This research has been conducted in order to bring light to these matters. Another interesting phenomenon concerning the Zipf's law is that for some countries the Zipfian distribution does not hold strictly, i.e. with a power coefficient of one. The Netherlands is one of those countries, with many cities of more equal size than the Zipfian distribution would predict. The CPB wonders if that is a problem or an asset? What can be the reason for this? Their first approach is focused on the decisions made at household level, which leads to the following research question posed in this research:

*To what extent do decisions made at household level influence moving behaviour between cities to cause the emergence of the Zipf's law?*

Studies conducted by Mansury and Gulyas (2007) and Gaujal et al. (2014) show that a Zipfian distribution of cities can be generated with an agent based model. Such a model tries to explain complex macro patterns, like city size, on the basis of individual decision making. These individuals, or agents, take account of their preferences and their perceived environment and base their decision making on these factors. Since CPB is interested in how individual decision making influences city growth, an agent based model is a very suitable research method. The CPB is also interested to see what it is like to work with this modelling method and how it might be of help to them in future projects.

This report will describe the conceptualization, modelling and analysis phases of this research. In chapter 2 a system analysis is conducted where the most important concepts which influence a household's moving behaviour are identified. Chapter 3 explains the narrative of the model, where the concepts identified in chapter 2 are translated into a sequence of events the households in the model will experience. Chapter 4 explains how the narrative is translated into model code by explaining the implementation of the concepts, assumptions and model relations. Next the model is verified and validated in chapter 5 and 6 after which the model experimentation is explained in chapter 7. The report finalizes with conclusions in chapter 8.

---

## 2. System analysis

This chapter will discuss the basic concepts that will be taken into account when conceptualizing the agent based model. The concepts have been identified using a literature research and by expert consultation.

There are of course many reasons for a family to move to another city. The most important reasons for moving can be categorized by reasons based on a household's life stage, job opportunities or housing (Ihrke, 2014). Next to reasons for moving, there are other concepts in play which influence people's decision-making. Together with CPB we have decided on several concepts that will be added to the model conceptualization.

### 2.1 Life stages

Research done by Bernard et al. (2014) shows that the amount of times someone moves as well as the distance between their current location and future home changes quite a bit for different life stages. This is shown in table 1.

*Table 1: Life-course transitions linked to migration (Bernard et al., 2014)*

<i>Transition</i>	<i>Age</i>	<i>Frequency</i>	<i>Distance</i>
Leaving parent's home	16–22	1+ in year	All lengths
Sexual union	20–25	0.3 per year	Short
Career position	23–30	0.2 per year	Many long
First child	23–30	0.5 in year	Short/medium
Mid-career promotion or inheritance	30–55	0.1 per year	Many long
Divorce	27–50	0.1 per year	Short
Cohabitation/second marriage	27–50	0.1 per year	All lengths
Retirement	55–68	0.3 in year	Many long
Bereavement or income collapse	70+	0.5 in year	Short/return
Frailty or chronic illness	75+	0.8 in year	Short

Note: Frequency 'in year' refers to the odds of moving home in the year of the transition event, and 'per year' to the odds of moving for that reason in any year during the age range indicated.

Not all life-course transitions need to be taken into account. For example, life stage transitions where the moving distance is short (within the same city) will not be taken into account because the model focusses on moving behaviour between cities. Furthermore, the level of detail in table 1 is not needed to produce a realistic, stable number of households during a model run, therefore only the most influential life stage transitions will be taken into account.

Examples of the most important transitions are leaving parent's home, the sexual union, the first child and retirement. Job related life stage transitions are kept out of scope for now since they will overlap with the next concept that will be discussed, namely jobs.

There are some interesting conclusions to be drawn from this graph, such as that young adults and adults up until the age of 55 move relatively often with medium up till long moving distances. After the age of 70 households do not move over long distances anymore. Between the age of 16 and 22 people move very often as well as all types of moving distances.

---

## 2.3 Jobs

Research done by Bernard et al. (2014) shows that people who move for job-related reasons are willing to travel relatively far. This makes this category very interesting. Furthermore, as many researchers indicate that a Zipf's law is probably related to economic factors such as research done by Robert Axtell and Richard Florida (2006), this category is a large factor in the model. There are three types of jobs that are often distinguished in moving models as stated by the CPB, namely primary sector jobs (agriculture, manufacturing), secondary and quaternary sector jobs (service and non-profit jobs) and finally tertiary sector jobs (for example, IT and finance jobs). These types of jobs all have different effects on city growth as for example primary sector jobs are mostly found in small cities where-as tertiary sector jobs are found in large, growing cities. These concepts are very interesting and because they might have a large influence on how cities grow, they will be taken into account in the model.

## 2.4 Housing

The housing category focusses on people moving to either a smaller or a larger house. This is tied to the housing market. In order to implement this category the housing market would have to be modelled as well. Because of the time available for this project this cannot be modelled and will therefore not be taken into account in the rest of the modelling or analysis.

## 2.5 Other concepts

There are various other reasons for people to move or for certain cities to be more attractive than others. Together with the CPB, the most important concepts that are within the scope of this project have been identified. The decision to move is then further influenced by the following concepts:

- Having children
- Aging
- Time since moving
- Distance from current location
- City size (facilities)
- Borrowed utility

These concepts will now be described in more detail. Since moving in this project means moving between cities and families with young children often only move within their current city, the concept of having children is implemented so that when a family has young children they will move less often.

The next concept, aging, influences how often someone will move. Since literature indicates that young adults move more often than older adults, the aging of a household can be used to determine how often they move.

The time since moving also influences how often a family moves. Since people will not move right away after they have just moved or after they have been living somewhere for over 30 years, the time since moving can be used to incorporate these effects.

The distance from the current location can be used to mimic the disadvantage of moving to a city which is very nearby as well as moving to a city which is very far away from your current location. Moving to a city nearby (within commuting distance) is not needed for many families. Moving to a city very far away is in reality not advantages either because a household will have connections, friends and family nearby at their current location.

Lastly, the concepts of ‘city size’ and ‘borrowed utility’ will be explained. Studies of Gulyas (2007) and Gaujal et al. (2014) emphasize the well-known phenomenon in urban economic literature; individuals are attracted by large cities. This is because large cities have many facilities, making it more attractive to live in such a city. The concept of ‘borrowed utility’ is closely linked to this concept as people in one city can use facilities of a large city when it is located nearby, thereby increasing the attractiveness of the smaller city.

### 3. Narrative

In this chapter the concepts discussed in chapter 2 will be translated into the model narrative. The agents in the model are households. This narrative will therefore focus on what events a household experiences during a typical model run.

Figure 1 visualizes the life stage transitions a household will go through. The narrative will start off with a household of two parents with a child, shown in the left of the figure.

After the child ages it will decide to leave its parent’s home after which he or she will move to the city which is most attractive to that person. The young adult will then form a couple with a partner living in the same city. During these years the couple is relatively young and will move quite often due to finding new jobs or being promoted to a higher function.

Then the family expands by getting children. This will slow down how often the family moves to another city. After the child grows up and leaves the house the parents grow older until they retire at the age of 65. They then prefer to live on the countryside, thereby mimicking the moving behaviour of retired adults in reality. This cycle will go on and on during the model run.

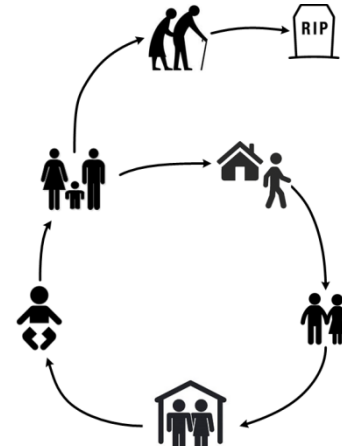


Figure 1: Household life cycle

During a household's lifespan it will move on several occasions. Whether a household moves is dependent on many factors next to the life-stage transitions, as described in the previous section. These concepts and how they influence a household are described shortly in table 2.

Table 2: Moving concepts and their influence on a household's decision making

<b>Having children</b>	Households are less inclined to move to another city when there are children.
<b>Aging</b>	A young household moves more often than an older one.
<b>Time since moving</b>	This determines when the household will be more inclined to move again. When they have just moved, they will not move immediately again, but when they have not moved for a long time, they are more inclined to stay.
<b>Distance from current location</b>	This distance determines whether moving to the city in question becomes less attractive because it is either very close and therefore moving is not needed, or it is too far away from their connections.
<b>City size</b>	The larger a city, the more attractive it is because of its increased number of facilities.
<b>Borrowed utility</b>	Utility gained from nearby cities will increase the attractiveness of the city in question.
<b>Job attractiveness</b>	Cities are more attractive based on the attractiveness of a certain job in that city.



---

## 4. Model logic

This chapter will discuss how the narrative has been translated into the actual model code. All the values for the parameters in the model can be changed before starting the model. Appendix B shows which parameters can be altered and between what values changes can be implemented. All values described in the coming paragraphs are examples of the default model setup.

The model consists of two parts. The first part is the population model, in which the life stage transitions of households are modelled. This part is used as input for the second part of the model, which models the moving behaviour of the households between cities.

### 4.1 Household agent

Agents in the agent based model represent households. A household has members, which is coded as a list of people. Every household member has the following attributes: age (0 to 100); job preference (1 to 7) and a sex (male/female).

Next to household members, a household also has a location. This location can be in one of the cities or on a patch in the countryside.

### 4.2 Model setup

The model starts by choosing locations for the number of cities the model is being run. Then the total number of households is divided over these cities and the countryside. The first city is always placed in the middle of the map, after which the other cities are distributed randomly in a circle with a certain width around this first city. This width is for example minimally 250 patches and maximally 500 patches. The cities are already coded at the start of the simulation since the focus of this research is not to study how cities emerge, but how they grow long after they have been established.

The households are distributed over the cities and the countryside. In the default settings of the model, 40% of the households are placed in the cities whereas 60% of the households are placed in the countryside at the start of the model run.

### 4.3 Part 1: Population model

Every member in a household gets older each tick of the model. They progress through different life stages in the model according to the flowchart shown in figure 2.

When a child is born a member is added to the household of 2 or more members. A child is aged 0 and the sex is chosen randomly. At birth they are already given a job preference from 1 to 7:

1. Primary sector jobs (18% of people)
2. Secondary sector jobs (32% of people)
3. Tertiary sector jobs (8% of people)
4. Quaternary sector jobs (34% of people)
5. Jobless (8% of people)

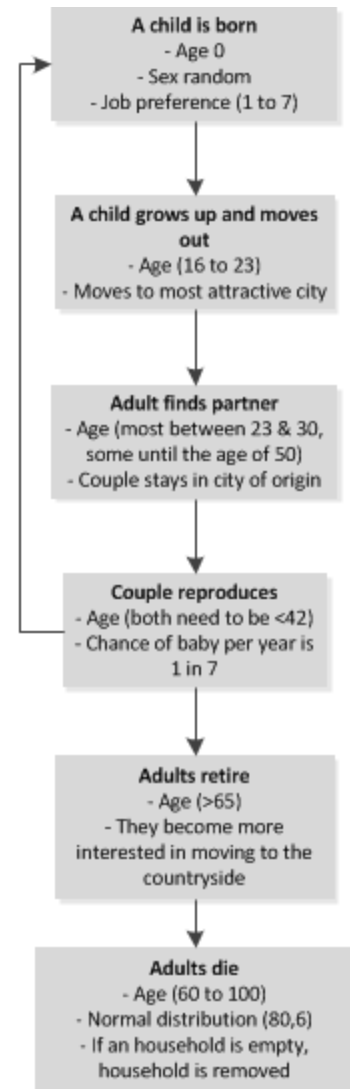


Figure 2: Life stage transitions as implemented in the model

---

These numbers are based on data from CBS (2014) so that the model is based on a realistic distribution of job types.

Each tick the child grows a year older. Between the ages of 16 and 23 the child will move out of its parents' home and move to the city that is most attractive for him/her. This will spawn a new household in the model.

Between the ages of 23 and 30 adults will then find partners that live in the same location. When a couple is matched, the households are merged into one household and they stay in the same location. It is possible for older people to be in search for a partner as well (<50 years old), because some people are not able to find a partner before they turn 30.

The couple then has a change to reproduce each year equal to 1 in 7. In order for a couple to reproduce they both have to be younger than 40. When a child is born, this child is added to the household and that child's life stage progress starts at the beginning of the flowchart.

Adults retire when they are 65 years old. They then become more interested in moving to the countryside, thereby mimicking the moving behaviour of the elderly who move out of the cities. Adults die between the age of 60 and 100. This is coded using a normal distribution with  $\mu = 80$  and  $\sigma = 6$ . When the last member of a household dies, the household is removed from the system.

#### 4.4 Part 2: Moving behaviour

Each household has a resistance to move and a willingness to move which is actually the attractiveness of a city for each household. When the attractiveness of one or a number of cities is higher than the household's resistance to move, the household will move randomly to one of the cities which attractiveness exceeds their resistance to move. This can be translated into one rule the model follows:

"A household moves to a random city for which its 'Attractiveness' is greater than the households' 'Resistance to move'.

Figure 3 shows the factors that influence the resistance to move and the attractiveness of a city.

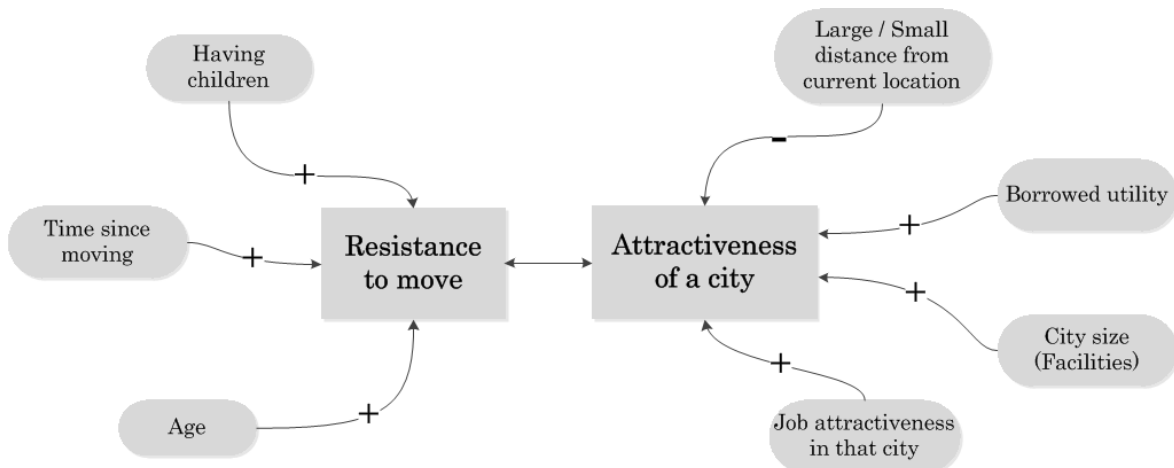


Figure 3: Model logic with regards to the moving decision making process

#### 4.4.1 Resistance to move

As you can see in figure 2, the resistance to move is influenced by three factors:

- Number of children in a household
- Time since moving
- Average age of adult household members

The effects of these factors are summarized in a graph which is shown in figure 3. The x-axis show the number of years since the last time the household moved to another location. This resistance to move (y-axis) can then become higher when a household has children and the lowest point of the graph is moved to the right when a household is relatively old, so the elderly do not move as often as young people do. The resistance to move is a value between 0 and 1.

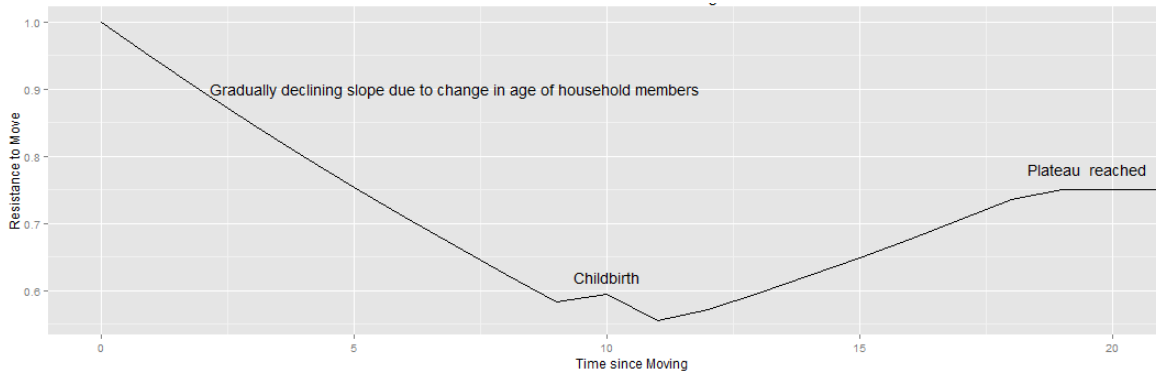
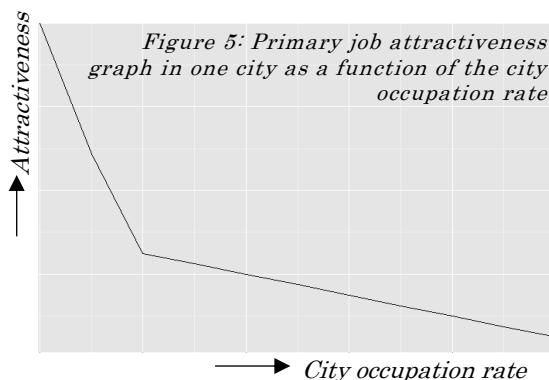


Figure 4: Measured resistance to move for a single household

#### 4.4.2 Attractiveness of a city

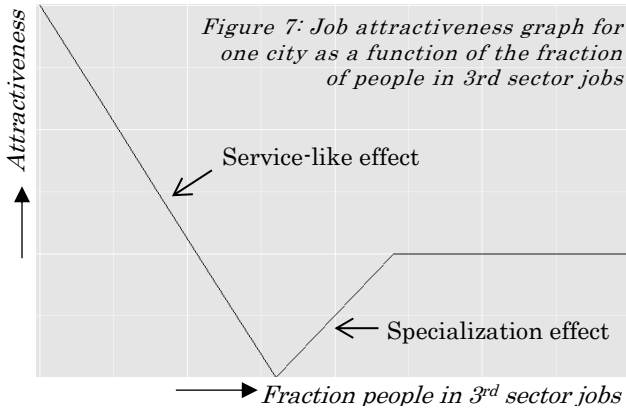
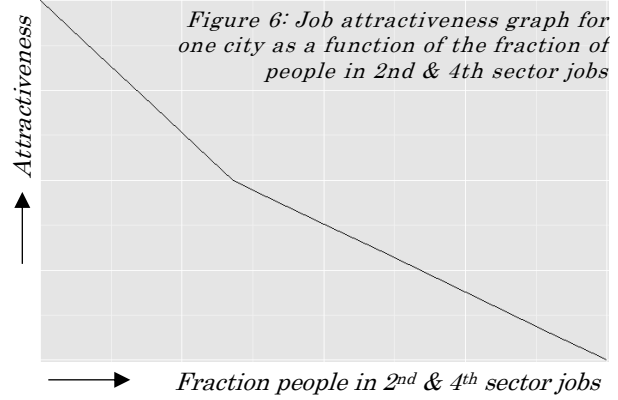
The attractiveness of a city to a certain household is dependent on the job preferences in that household and the distance from their current location to a city. The attractiveness of a city also takes on values between 0 and 1. We assume that people prefer to stay at their current location when a more attractive city is very close by their current location. We also assume that people do not move to cities that are very far from their current location. This translates into lower city attractiveness for cities that are close by or very far away from a household's current location.

The attractiveness is also influenced by the household members' job preference. Each city has a unique attractiveness score for each job based on the amount of people in the city with that job. The attractiveness of a city by job is determined for each adult household member and is then averaged into a final score for the household. The way in which this attractiveness is determined differs per job type:



Cities are more attractive to manufacturing and agriculture jobs if they are relatively small cities. The size of cities in this case is used as a proxy for land prices, which greatly influence the profitability for these types of jobs.

Jobs in the service and non-profit domain are not influenced by city size in general. Instead, jobs in these domains are expected to spread evenly over all cities. This means that cities with less service jobs than average is more attractive for service jobs.



There are also two groups of jobs which show a networking effect. Cities are more attractive for IT and finance jobs if more people are working in that domain in the city.

Finally, each city is evenly attractive for jobless people. In Appendix A more detailed explanations of the job attractiveness graphs are presented for the different job types.

#### 4.4.3 Attractiveness of the countryside

The attractiveness of the countryside is a constant value for the different job types, so that for IT and finance job type the countryside is not very attractive, but for primary and secondary jobs the countryside is very attractive.

#### 4.4.4 Distance from current location

The attractiveness of a city is also influenced by the distance from the household's current location. By default, the attractiveness of a city gets reduced by a small margin when it is either closer than a certain minimum threshold from the current location, or farther away than a maximum threshold.

#### 4.4.5 City size

Apart from distance and job preferences the attractiveness of a city is also dependent on its size. The size is used in this model as a proxy attribute for the amount and quality of services and facilities a city has to offer. By default, the attractiveness of a city increases linearly with its size. However, this can be adjusted to also grow exponentially or logarithmically.

#### 4.4.6 Borrowed utility

Finally the attractiveness of a city is not solely dependent on the city itself, but also on cities close by. Cities in the model will gain a small boost to attractiveness if they are close to other attractive cities.

---

## 5. Verification

Before employing the model to examine the effect of household decision making on the emergence of a Zipf's law, it is necessary to ensure the model is verified and valid. Model verification is used to determine whether the built model performs in a manner in which it was intended to perform. The model was verified on three distinct levels: Single-agent level, in which the behaviour of a single household is analysed; minimal interaction level, in which the interaction between a small set of agents is observed; and multi-agent verification which analyses the emergent behaviour of multiple agents.

### 5.1 Single-agent verification

There is only a single type of agent in this urban migration model, namely households. Households live in a certain location and consist of household members.

As described in section x, households follow a certain life stage progression chain. This life stage progression has been verified by attempting to falsify the outcomes. To do this, we have first determined the possible composition s of households. Then, a large number (100,000) of households were generated and the model was run for 100 years. In addition, a test was written in NetLogo to verify whether any of the households were in a state which should not be possible.

The possible household compositions are:

- Single member household with  $16 < \text{age} < 30$  OR  $\text{age} > 53$
- Two member household with  $\text{age} > 23$
- Households with children with two  $\text{ages} > 23$  and children with  $0 < \text{age} < 23$

No households were found outside of the expected composition ranges and therefore it is assumed the household progression is implemented correctly.

The second main behaviour of the households is determining the attractiveness of cities and migrating. Since determining the attractiveness of cities largely depends on the households living in them, this aspect cannot be tested at the single-agent level, but will be tested on the minimal interaction level.

Aspects which can be verified on a single-agent level are the calculation of the resistance to move, of which a plot is shown in figure 4 and the actual choice of moving to a different city with preset city attractiveness. Both were found to show no errors.

#### 5.1.1 Agent Robustness

The implementation of agents in this model can be considered to be quite robust as no normal model runs can cause the agent to break. However, some notions need to be considered when generating custom household agents or when expanding on the model. First, the model assumes that the list of household members is always sorted descending by age. This assumption is made since it drastically improves the performance of several checks in the model. Furthermore, households with children that have fewer than two parents can cause the model to behave unexpected in several fields such as the calculation of city attractiveness and the overall life stage progression.

---

## 5.2 Minimal interaction level

The main behaviour on a minimal interaction level is determining the attractiveness of cities based on the people living in them. Several scenarios were set up and in each of these scenario's the attractiveness for a representative set of households was determined.

In the first scenario the attractiveness of cities for manufacturing and agriculture jobs was tested. Three cities were set up containing respectively 5% of total inhabitants, 20% of total inhabitants and 50% of total inhabitants. The attractiveness of two households consisting of solely agricultural members and solely manufacturing members were then checked against the assumptions and found to be correct. The results are shown in table 3.

*Table 3: Results of three scenarios on the attractiveness of cities for primary sector jobs*

Inhabitants	Jobs	Attractiveness
5%	Agriculture	0.68
5%	Manufacturing	0.68
20%	Agriculture	0.4
20%	Manufacturing	0.4
50%	Agriculture	0.25
50%	Manufacturing	0.25

The second and third scenarios were set up to verify the attractiveness calculation for service and non-profit jobs. Three cities for each job were set up containing respectively a low number of people in that job (1%), a medium number of people (30%), and a high number of people (80%). The results are shown in table 4.

*Table 4: Results of three scenarios on the attractiveness of cities for secondary and quaternary sector jobs*

People in Job	Job	Attractiveness
1%	Service	0.84
1%	Non-Profit	0.85
30%	Service	0.53
30%	Non-Profit	0.56
80%	Service	0.14
80%	Non-Profit	0.15

The final two scenarios were set up to test the attractiveness for IT and Finance jobs. Four cities for each job were set up containing respectively a low number of people in that job (1%), an average number of people (4%), a moderately high number of people (6%), and a high number of people (8%). The results are shown in table 5.

*Table 5: Results of four scenarios on the attractiveness of cities for tertiary sector jobs*

People in Job	Job	Attractiveness
1%	IT	0.87
1%	Finance	0.87
4%	IT	0.4
4%	Finance	0.4
6%	IT	0.6
6%	Finance	0.6
8%	IT	0.6
8%	Finance	0.6

### 5.3 Multi-agent level

On the multi-agent level the overall emergent behaviour of the model is tested under various circumstances. First, an extreme value analysis is performed on the two global model parameters which set the total number of households and the total number of cities in the model.

Varying the the total cities in the model from 1 to 25 gives results as shown in figure 8. As can be expected, the model does not behave with a very low number of cities. This is mostly because the option space of agents is then too small for interesting behaviour to emerge. From 10 cities and up the model shows expected behaviour. When further increasing the number of cities (>25) some care should be taken that the total number of households will also be increased. If this is not done, the average households per city will be too low which causes unexpected behaviour as described in the following section.

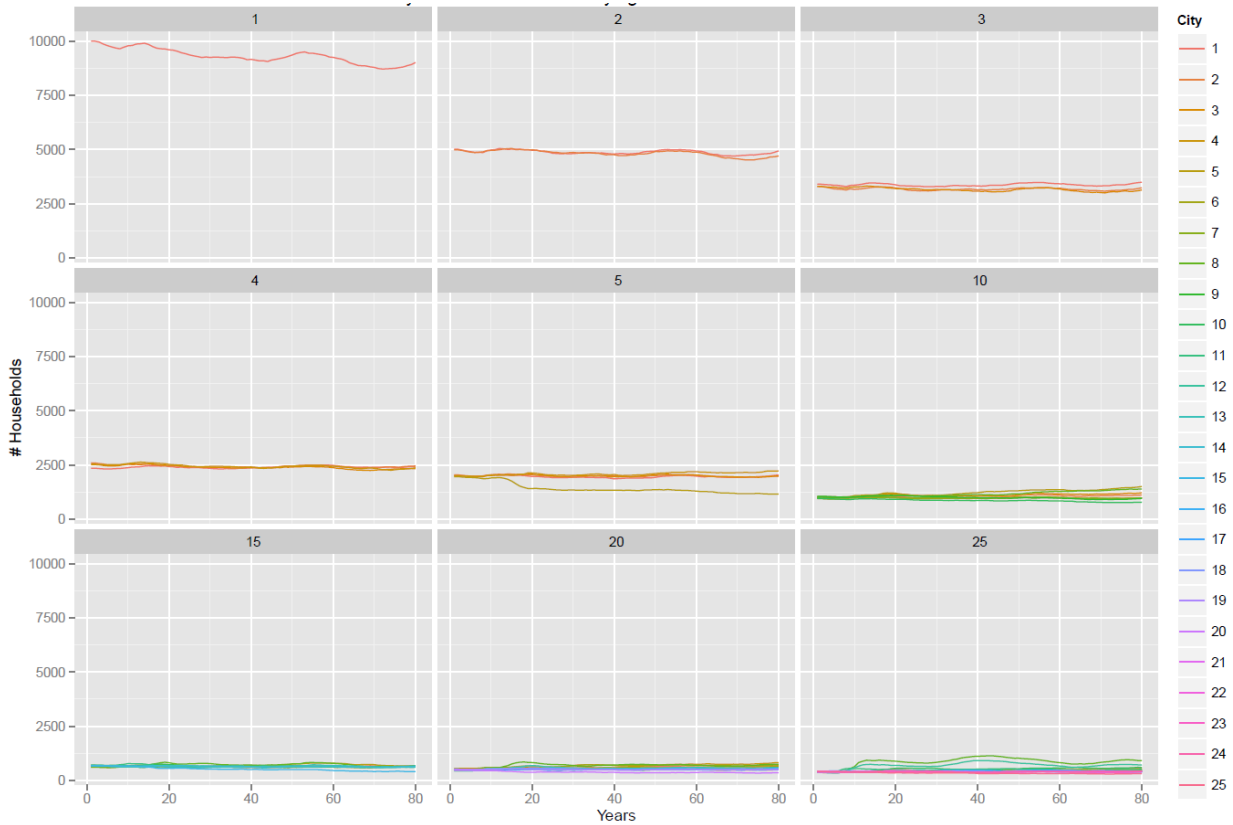


Figure 8: City Sizes over time for a varying number of cities in model

A similar test was performed by varying the total number of households in the model from 1,000 to 100,000 (figure 9). This analysis shows that the model is less suitable for use with a very low number of households (< 10,000). This has several causes: First of all, with such a low number of households it is more difficult for agents in the model to find partners, leading to a decline in total households over time. Furthermore, the city attractiveness effects are smaller which leads to a lower than intended rate of migration. When running the model with a very large number of agents, it can happen that there are large oscillations of city size in the start of the run. This behaviour is caused by two aspects. Firstly, there are several large oscillations in the population system underlying the model in the transitory state (see figure 10). Another

reason is that normally the city attractiveness only gets calculated on a yearly basis to improve performance. In order to increase the fidelity of large-scale models the "updateCityAttractivenessFreq" slider can be used to set the model to update this attractiveness on a more frequent basis. When this slider is set to a suitably small value, the model results will closer reflect the behaviour as it is with lower amounts of households.

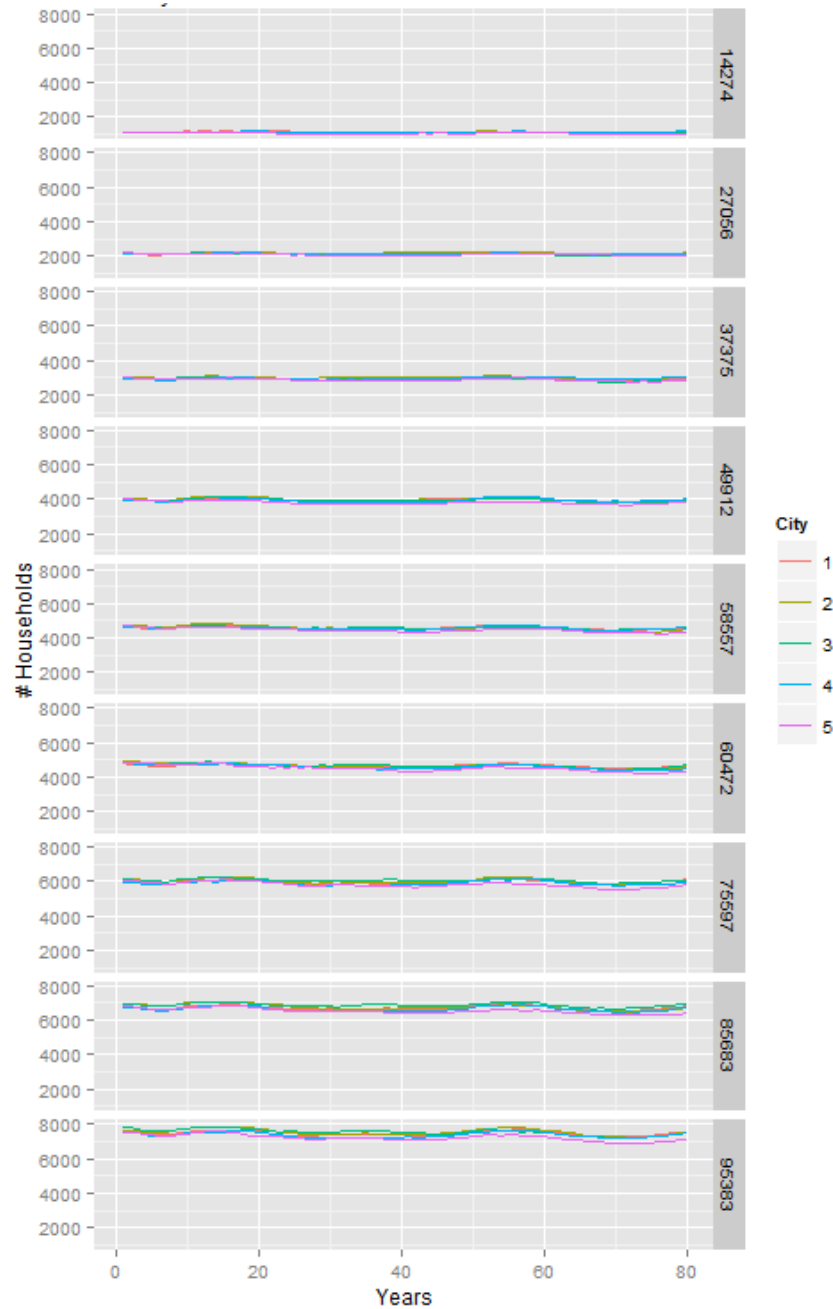


Figure 9: City Sizes over time for different total household sizes



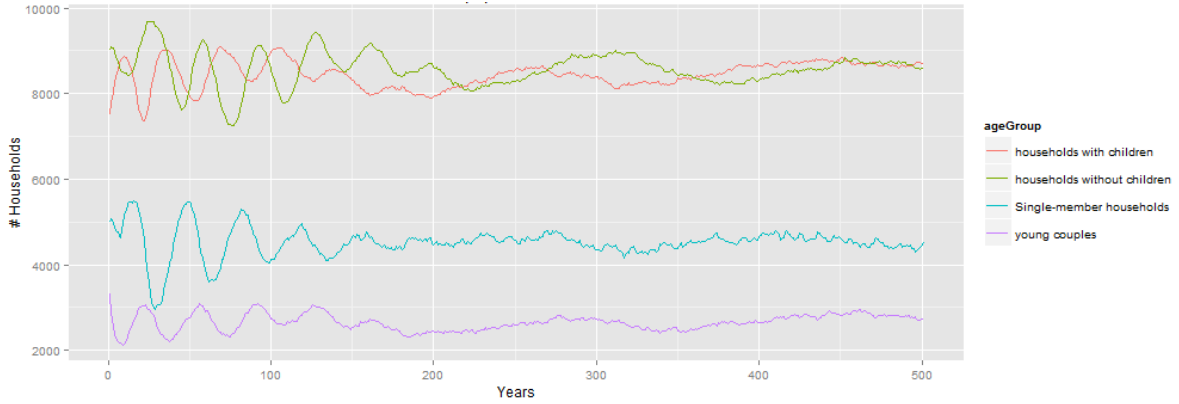


Figure 10: Behaviour of population model over time

Apart from extensively testing these two model input parameters a full parameter sweep with 300 model runs on all other possible input parameters has been performed. In none of these model runs did the model give any errors and the results are all within the expected bounds (see figure 11).

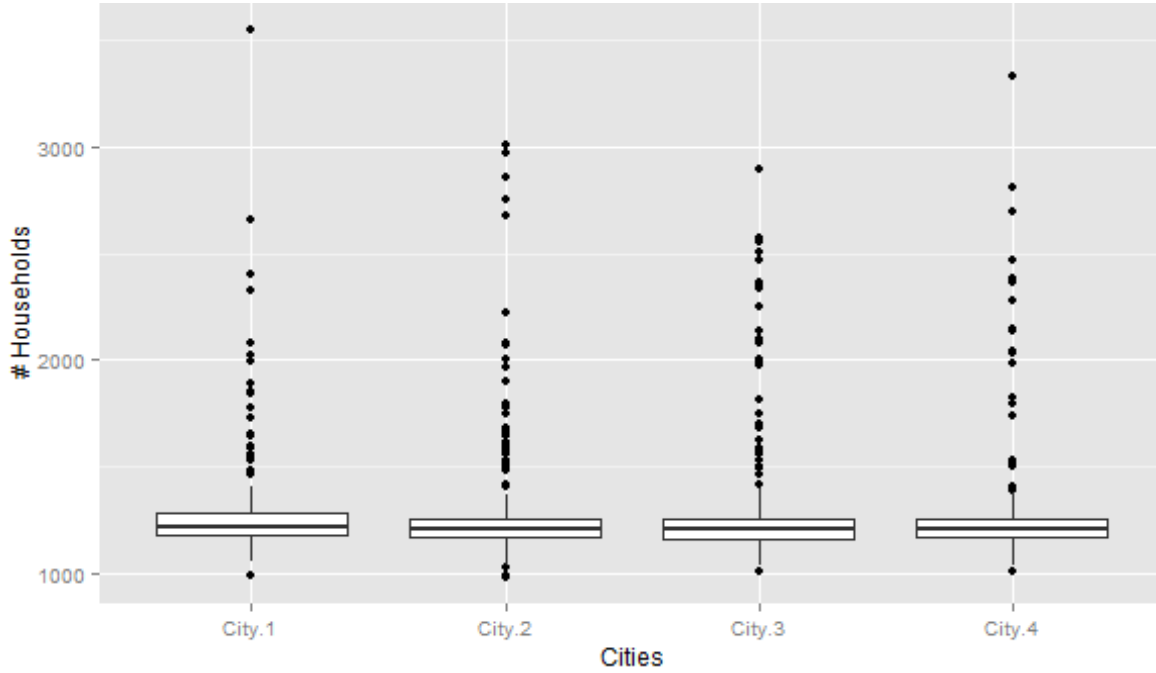


Figure 11: A parameter sweep for the number of households per city

## 5.4 Conclusion

This model has been verified by 6 tests on a single-agent, minimal interaction and multi-agent level. All of these tests show that the model is implemented in line with the intended implementation as described earlier. However, some care must be exercised in setting up the model for experimentation to make sure the input parameters are within certain bounds to ensure the model behaves as expected.

---

## 6. Validation

In this chapter the validity of the model is discussed. In order to determine whether the model is valid, it must be checked whether the model outcomes correspond with reality, i.e. does a Zipf's law emerge, and if the model makes correct predictions. The goal of this research was to understand how individual decisions at household level influence moving behaviour in cities to cause the emergence of the Zipf's law.

In order to answer the questions posed above, two types of validity tests have been conducted, namely literature validation and face validation through expert consultation.

### 6.1 Literature validation

The model parameters in the population model are based on literature research and this part of the model is therefore considered valid. Since the part of the model focused on the decision making on whether to move or not is also based on concepts found in literature and concepts identified by experts, the factors included in this part of the model are also deemed valid. What exactly has been validated through expert consultation will be discussed in the next paragraph.

### 6.2 Face validation through expert consultation

The model consists of two parts, the life stage progression of households and the moving behaviour of households. The first part has been validated by CPB experts by showing them the behaviour of the population model over time.

The second part of the model, the moving behaviour of households, has been validated by discussing the underlying assumptions, included concepts and model relations. The CPB deemed our model to be valid for the purpose of this research.

## 7. Model experimentation

So far this document has focussed on the conceptualising, building and testing of a model for determining the effect of household-level decision making on the emergence of a Zipf's Law in city size distribution. The following section will go in depth on the experimental design used to analyze the model's behaviour and the results which came from it.

The model should be considered to be an infinite system. Such a system will not return to its initial state after a predetermined amount of time. Since the model is an infinite system a warm-up time will be required before data can be gathered from the model. The system is expected to reach steady-state after a full life-cycle of the initial households. Since the life-expectancy of a household (not a person) is about 40-60 year, the model warm-up time should be higher than 60 years. Apart from reaching steady state in migrational behaviour, the model also exhibits some transitory behaviour in the initial period with regards to the total population (see figure 10). This population system reaches steady state at around 250 years. However, in experimentation it was found that this transitory behaviour does not affect the outcomes heavily and therefore this factor was omitted in determining the warm-up period as to reduce total model run time. Therefore, for this experimentation a warm-up period of 80 years is chosen.

As became clear in section x, there are large uncertainty bands around many of the model parameters. For this study, a multi-variate analysis was set up in order to determine which factor or combination of factors has the largest effect on an emergence of a Zipf's Law.

Due to the large input space a full-factorial experimental design is unfeasible within the set time-constraints. Within these time-constraints it was possible to perform about 300 model runs. In order to still get decent results, latin hypercube sampling was used as a sampling method to ensure a uniform distribution over the input space.

After running the model the individual runs were evaluated on how well they approach a Zipf's Law. This was done by calculating the root-mean-square error of a run compared to a true zipfian distribution with 1 as a Zipf's coefficient.

Analyzing the effects of the change in input parameters on the model behaviour was done by using a decision tree. This decision tree was inferred using an C4.5 Decision Tree inference algorithm supplied by the RPart R Package. In order to avoid overfitting on the data, the tree was pruned optimizing on a lowest cross-validation error. The results of this decision tree are shown in figure 12.

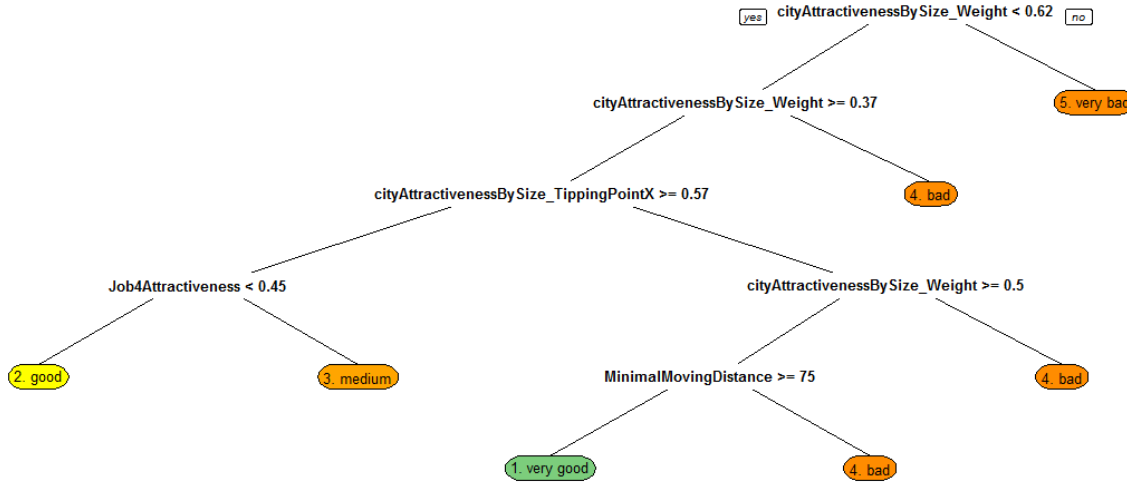


Figure 12: The effects of changes in input parameters visualized in a decision tree

This decision tree shows that a Zipf's law is most likely to emerge when the weight of the city attractiveness by size modifier is between 0.5 and 0.63, and when the people are less inclined to move to cities within 75 patches from them.

This result has been verified by running 20 new independent replications with these settings. From these runs we get an average RMSE of 435 while the original runs had an average RMSE of 1412. This result shows a massive improvement in the emergence of a Zipf's law which leads us to conclude that this moving distance is indeed a large factor in the emergence of a Zipf's law in the model.

---

## 8. Conclusions and Future Research

The main goal of this research was to determine to what extent decisions made at household level influence moving behaviour between cities to cause the emergence of the Zipf's law.

In order to determine the size and growth of cities over time, a population model was build based on a household's life stage transitions and several concepts were included on which households based their decision making regarding their decision on whether to move or not.

The model shows that there are two important factors causing the emergence of a Zipf's law, namely the effect in which a city's attractiveness increases as its size increases, and the distance households are willing to migrate. Adjusting these values results in about 30% of the runs showing a good to very good Zipfian distribution and reduces the RMSE compared to a true Zipfian distribution from 1400 to 400. However, considering the robustness of the Zipf's law throughout history and the fact that the model needs rather specific values to produce a Zipf's law, it seems that the model is lacking an important effect in city migration behavior.

Therefore we conclude that the model cannot truly explain why a Zipf's law emerges.

Several improvements to the model can be made. A large improvement could be the changing of the main decision rule by which households migrate. Currently, a household moves to a random city which attractiveness is higher than the household's resistance. The model would be more realistic if the added attractiveness of a target city compared to the current city's attractiveness is taken into account. Another major improvement could be the addition of a rudimentary housing market.

Interesting topics for future research in this field could be focused on why the Zipf's law doesn't emerge as heavily in The Netherlands as it does in other cities. Our research currently attributes this to the distance between the cities being relatively short, however this might not be the only cause. Another interesting topic is what happens to cities across Europe when the borders fade more over time and how digital communication affects city growth.

---

## 9. Reflection

Apart from the main research question a large relevance of this project for the CPB lies in determining the applicability of agent-based modeling and Netlogo for researching these types of policy problems.

Overall we feel that the application of agent-based modeling for these types of problems is very suitable since it is relatively easy and fast to create a model featuring complex decision making mechanics and interaction rules. However, a main disadvantage over more high-level modeling paradigms such as system dynamics is that the model runs are much slower.

As shown in the verification section, the effects in this model only emerge when having a large number of agents in the system. Due to this large number of agents it was infeasible for the model to calculate the complex variables such as city attractiveness whenever something changed. A large part of model development time was spent on increasing the model speed in order for it to run reasonably well. Netlogo attempts to simplify the modeler's job by abstracting many of the fundamentals underneath the language such as data structures. However, many of the data structures which were chosen by the Netlogo developers don't seem to be chosen with performance in mind. For example, agents are stored in an array which means the only way to select an agent is in linear time  $O(n)$  instead of near constant time if a hash table was used. Some other performance issues that were encountered are that the patches-in-radius command as well as set subtraction seems to be quadratic time  $O(n^2)$  which makes it nearly unusable for large numbers of patches as well as the distance function being implemented using a power function to calculate the square of a number which uses a Taylor series instead of a regular multiplication which is ten to a hundred times faster.

Furthermore, lacking functionalities in the model testing such as unit tests and assertions make developing large-scale models over time very error-prone.

Finally, although RNetlogo should offer a very useful interface to run simulation runs from the R programming language, its usefulness is diminished greatly by the large memory leaks when finalizing a model run, making it nearly impossible to run a large number of runs without human interaction.

In conclusion we can state that Netlogo is suitable for rapid prototyping and developing small models. However, as models get larger it prohibits the developer from making efficient choices by abstracting many of these choices. Therefore, it becomes less suitable as model size increases.

---

## 9. References

- Axtell, R. and Florida, R. 2006. *Emergent cities: micro-foundations of Zipf's Law*.
- Bernard, A., Bell, M. and Charles-Edwards, E. 2014. Life-Course Transitions and the Age Profile of Internal Migration. *Population and Development Review*, 40 (2), 213-239.
- CBS. 2014. Beroepsbevolking; geslacht en leeftijd. Retrieved from <http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71738ned&D1=1-26,32-34&D2=a&D3=a&D4=0&D5=l&VW=T>
- Gaujal, B., Gulyas, L., Mansury, Y. and Thierry, E. 2014. Validating an agent-based model of the Zipf's Law: a discrete Markov-chain approach. *Journal of Economic Dynamics & Control*, 41, 38-49.
- Ihrke, D. 2014. *Reason for Moving: 2012 to 2013*. United States Census Bureau.
- Mansury, Y. and Gulyas, L. 2007. The emergence of Zipf's Law in a system of cities: an agent-based simulation approach. *Journal of Economic Dynamics & Control*, 31, 2438-2460.

---

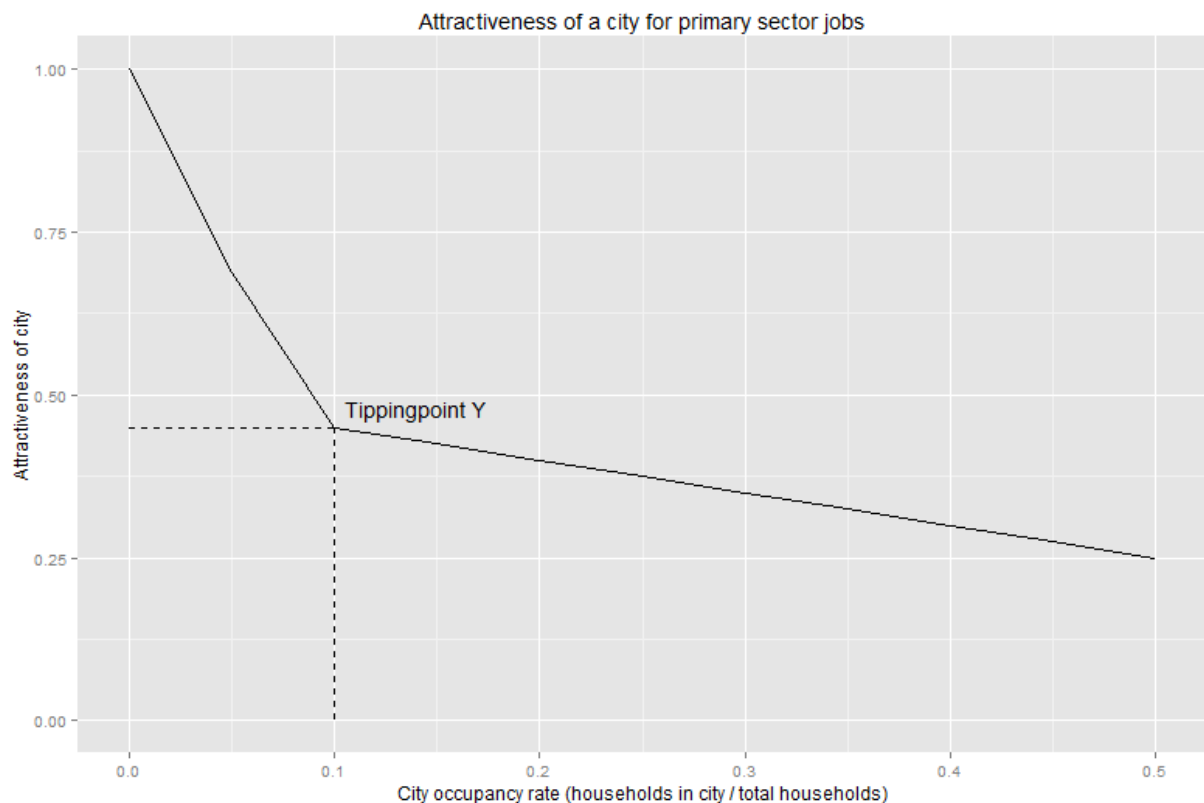
## Appendix A: Job attractiveness graphs

This appendix shows under which circumstances certain jobs are attractive. Different job types rely on different formulas to calculate the attractiveness of a city for this job type, which results in different ranges in x-axis. Why they differ will be explained for each job type.

### Primary sector jobs

Job 1 and 2 are primary and secondary sector jobs, which attractiveness for a city is determined by the fraction of households in a city. The graph showing the attractiveness of a city for that job type for different fractions is shown in figure 4. The y-axis is the city attractiveness for that job type and the x-axis is the fraction of households in that city. These jobs are coded in such a way that cities with low density are more attractive for these job types.

The x-axis ranges from 0 to 0.5 because it depicts the fraction of households in city X from the total number of households in the country. In the base case of the model there are 5 cities, therefore if all households were distributed equally, each city would have this fraction be equal to 0.2 and a city with a fraction equal to 0.5 would be an extremely large city. Therefore a range from 0 to 0.5 takes into account all possibilities.

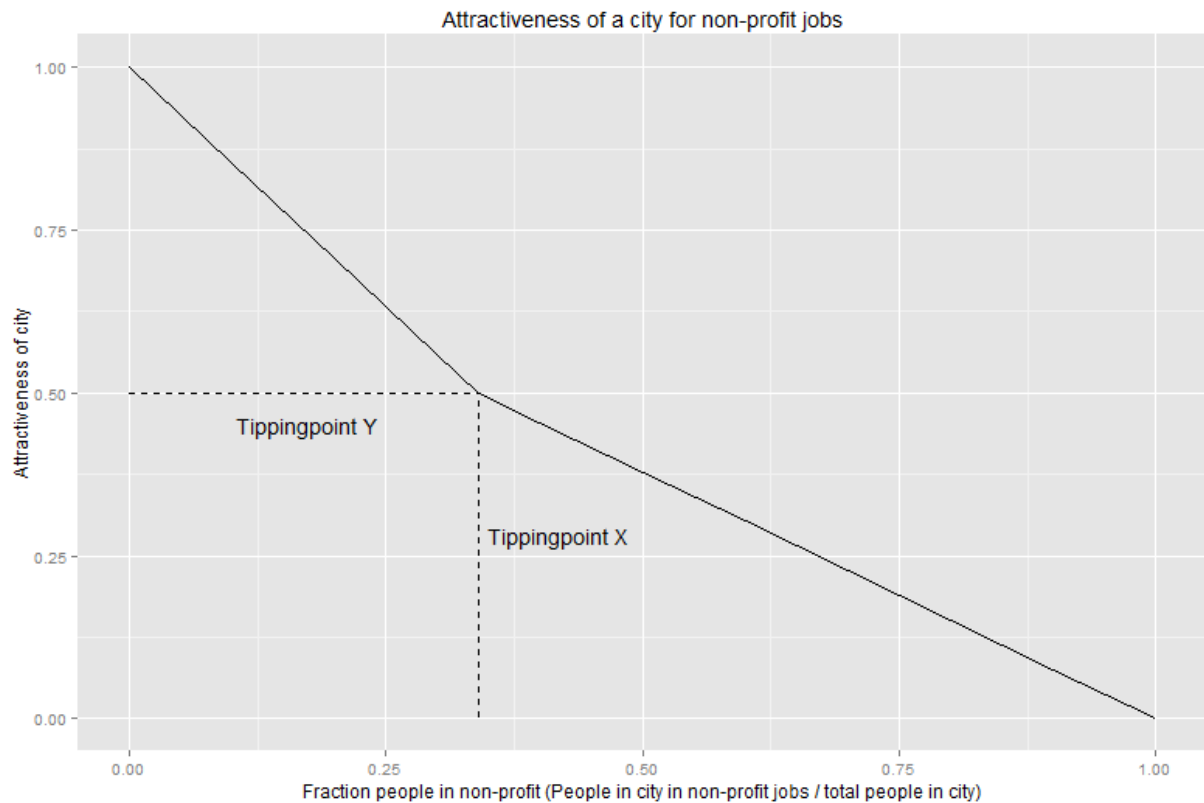


### Service and non-profit jobs

Job 3 and 6 are service and non-profit jobs, which attractiveness for a city is determined by the fraction people in service / non-profit. These jobs are coded in such a way that all cities need a certain percentage of people performing these jobs compared to the total number of households

in that city. The graphs in which these effects are taken into account is shown in figure 5 and 6 respectively. The x-axis shows the fraction of people in service and the y-axis shows the city attractiveness for that job type.

The x-axis has a range from 0 to 1, which is different from the graph in figure 4, because the fraction used to make this graph is different. Here the fraction is not the number of households in city X compared to the total number of households, but the number of people in service or non-profit jobs compared to the total number of households in that city. When this fraction is 0, the attractiveness for this job should be very high as a percentage of the total households in that city is always required to be filled for this job. When this fraction is equal to 1, the city attractiveness for this job type should be equal to 0 so that no more people take on such jobs than is necessary. Therefore the range of the x-axis is from 0 to 1.



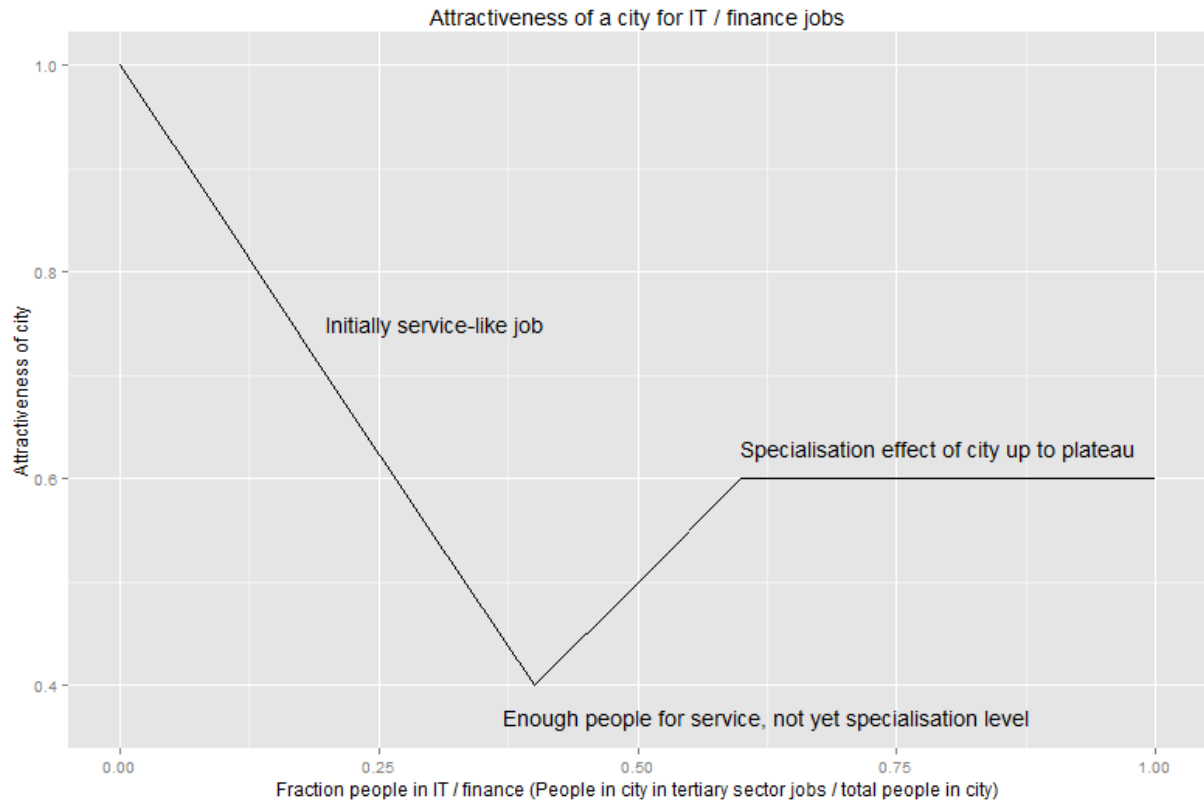
### Finance and IT jobs

Finally, job 4 and 5 are finance and IT jobs, which attractiveness grows the more finance and IT jobs are taken in a certain city. This effect is shown in figure 7 where the x-axis presents the fraction of people in finance/IT and the y-axis shows the attractiveness of a city for these job types. The attractiveness is bound at 0.6 when the fraction is high so that these job types don't cause exponential growth.



---

The x-axis for this graph shows values from 0 to 0.1. This is because the x-axis represents the fraction of people in finance or IT jobs compared to the number of households in a city. Because the number of IT and finance jobs in total is very low, only 4% of all people, this value in a single city will range from 0 to very high. Because this maximum value is entirely dependent on the type of people that live in a city (more IT/finance people in a city will exponentially bring even more IT and finance people), we have capped the attractiveness at 0.6 as explained before. The x-axis is then capped at 0.1 since the y value is stable anyway.



---

## Appendix B: Model experimentation

In the model experimentation the parameter values in the model are changed to determine which parameters are most influential in creating a Zipf's law. The table below shows all the parameters that were changed in this phase and within which values they were changed. For each parameter a small explanation is given.

Variable name	Min value	Max value	Explanation
cityAttractivenessBySize_Weight	0.5	0.6	Weight of the effect of city size on city attractiveness
cityAttractivenessBySize_StartY	0.4	0.6	Start Y-value for the 'effect of city size' graph
cityAttractivenessBySize_TippingPointX	0.57	0.8	X-value tipping point for the 'effect of city size' graph
cityAttractivenessBySize_TippingPointY	0.6	0.8	Y-value tipping point for the 'effect of city size' graph
cityAttractivenessBySize_EndY	0.8	1	End Y-value for the 'effect of city size' graph
borrowedUtilityMaxDistance	50	250	The maximum distance between two cities for which the city in question receives attractiveness from borrowed utility from a nearby city
borrowedUtilityWeight	0	0.25	Weight of the effect of borrowed utility of city attractiveness
populationGrowth	1	1.01	Population growth percentage per year
job4_Modifier	8	12	Multiplier effect in the job attractiveness graph of job type 4
job5_Max	0.5	0.8	Plateau value of the job attractiveness graph of job type 5
MinDistCityAttractiveness	0.01	0.5	Effect of minimum moving distance on city attractiveness
Job6Attractiveness	0.4	0.6	Job attractiveness for type 6 in the countryside
job7_Value	0.4	0.6	Attractiveness of job type 'jobless' in cities

---

job6_TippingPointX	0.2	0.4	X-value for the tipping point in the attractiveness graph of job type 6
rtm_TippingPointY	0.25	0.75	Y-value for the tipping point in the resistance to move graph
job2_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 2
Job3Attractiveness	0.4	0.6	Job attractiveness for type 3 in the countryside
job5_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 5
minDistBetweenCities	30	150	Minimum distance between cities in model setup
maxDistBetweenCities	200	500	Maximum distance between cities in model setup
rtm_TippingPointX	5	15	X-value for the tipping point in the resistance to move graph
rtm_PlateauPointX	15	25	X-value for the plateau point in the resistance to move graph
rtm_PlateauPointY	0.5	1	Y-value for the plateau point in the resistance to move graph
rtm_AgeModifier	0.01	0.3	Effect of age on the resistance to move
rtm_ResistancePerChild	0.01	0.3	Effect of having a child on the resistance to move
MinimalMovingDistance	75	200	Minimal distance from which people will want to move
MaximumMovingDistance	200	400	Maximum distance until people will not want to move
MaxDistCityAttractiveness	0.01	0.5	Effect of maximum moving distance on city attractiveness
Job1Attractiveness	0.4	0.6	Job attractiveness for type 1 in the countryside
Job2Attractiveness	0.4	0.6	Job attractiveness for type 2 in the countryside

---

---

Job4Attractiveness	0.4	0.6	Job attractiveness for type 4 in the countryside
Job5Attractiveness	0.4	0.6	Job attractiveness for type 5 in the countryside
Job7Attractiveness	0.4	0.6	Job attractiveness for type 7 in the countryside
job1_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 1
job3_TippingPointX	0.2	0.4	X-value for the tipping point in the attractiveness graph of job type 3
job3_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 3
job4_TippingPointX	0.01	0.1	X-value for the tipping point in the attractiveness graph of job type 4
job4_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 4
job4_Max	0.5	0.8	Plateau value of the job attractiveness graph of job type 4
job5_TippingPointX	0.01	0.1	X-value for the tipping point in the attractiveness graph of job type 5
job5_Modifier	8	12	Multiplier effect in the job attractiveness graph of job type 5
job6_TippingPointY	0.4	0.6	Y-value for the tipping point in the attractiveness graph of job type 6