The PARSO$_{demo}$ Model

Davide Secchi & Gayanga B. Herath

Research Centre for Computational & Organisational Cognition

University of Southern Denmark

The document here presents the PARSO$_{demo}$ model (Section 1) by introducing information that can also be found on the `Info` tab in the `NetLogo` model (uploaded separately). It then gives an overlook at sensitivity analysis (Section 2) and finally shows regression analyses that were not inserted in the associated paper (Section 3).

1 MODEL INFORMATION

1.1 WHAT IS IT?

The starting point is information published in March 2018 by the Danish government *Et Danmark uden parallelsamfund* (trans. *One Denmark without parallel society*). The plan is to eliminate what they call 'parallel society' in the country (see below) by getting rid of areas officially called ghettoes by 2030. These areas are neighbourhoods where the local municipalities allow poor citizens to reside, paying a rent they can afford (social housing).
Information on how to operationalise the plan is meagre. However, before being concerned with the plan of how to erase ghettoes from the map of Danish cities, one should ask the question on whether a parallel society is something that could form, given the current socio-economic picture. The problem should be tackled by a three-step approach: (A) estimating the possibility of a 'parallel society' formation, (B) defining whether it is located in a so-called ghetto area, and (C) exploring the implications of erasing a ghetto area from the map. All these steps are tested on areas where the government indicates there are the official ghetto areas. However, such a map and the number of residents necessary to have a more realistic picture require intense computational power. Hence, before running the model full scale, we have created the current model as a test bed for dynamics to be applied later to the larger scale model.
The model currently presented here is a $6/100$ scale of an area in the city of Copenhagen, and it represents possible configurations of neighbourhoods. We have artificially created several conditions to observe whether some of the elements we want to test on the full model have an impact in this smaller scale model. And, of course, we want to analyse what this impact would be. This version of the model only covers the first step (A) on **society formation**, as described above.

1.2 HOW IT WORKS

This is a very interesting and long standing topic in sociology (see Martin Neumann's work). The concept of a 'parallel' society implies a way of organising that is different (perhaps

incompatible?) from the established ways of an already established society. Hence, it leans on (a) radically different values, (b) different social norms, (c) non-standard behaviour (standard being the one from the majority of society), and generating (d) conflict with the other 'parallel' social majority.
The first step would be to see which one between (a), (b), and (c) lead to the formation of a 'parallel' society. Clearly, a society (any society) is autonomous but not completely independent from other societies. But it is different enough, so that it can be thought of being 'separate' or 'parallel' to the other. The general idea is that those living in the ghetto constitute a 'parallel' society. Let us take the following assumptions forward in the simulation:

1.2.1 Values

We start from attributing values to agents in the entire society. We assume that some agents have values that are distant from those of the mean. These values are mostly grounded in (to stick to the Danish government's predicament) the acceptance of a *democracy*, *tolerance*, and *role of women*.

`Democracy`.
The range for democracy acceptance is such that a random-normal distribution with mean $= 1$ and standard deviation $= 0.5$ will determine those that believe in its worth (around the average and up) and those that are less positive about it, to include/consider those that are negative. The latter group include extremists of various kinds, both political extremists and actual or potential religious fundamentalists.

`Tolerance`.
This is a bit more tricky to consider. Tolerance is a value that has clear behavioural implications, at least, for simulated residents in our simulation. This is linked to the model by Schelling on segregation. In his model, ultimately agents had a general feeling of happiness directly linked to the extend to which their neighbours were similar to themseleves. That was homogeneous in the entire society. In this simulation, tolerance has to be considered heterogeneous, so that it could vary at the individual level. Tolerance is very much related to the neighbourhood one lives in. If it is high, then there is not much of an issue in defining which neighbours one has. If it is on the low, then most surrounding agents must be similar to the self. The value is set on a range that goes between 0 and 1. The only condition in the setting is that individuals with very high democratic values will also approach high tolerance. On the contrary, those with negative democratic values will be mostly intolerant. Everyone else will have random values for tolerance. When the switch `hetero-tolerance` is OFF, the model works very similarly to the segregation model in the `NetLogo` version (also available among the NetLogo Library Models). When it is turned to ON, this switch allows for the population of agents to have a max tolerance for diversity set by the proportion in the 'similar-wanted' slider.

`Gender balance`.
The simulation works around resident units. This means that each agent is a resident and this could be a single individual, or a family. The family could be formed by just two individuals, or more. 2 and 3 residential units are the most diffused in the system. Now, independent of the number of individuals in the residence unit, there is an assumption on gender balance; whether it is high/low on a scale that sees it ranging from 0 to 1, where 0.5 (or around it) is a balanced view and values going down to zero are progressively more unbalanced. Values close or down to 0

mean that there is complete separation of roles between men and women in society, with a strict masculine take. Values close to 1 have opposite meaning, showing a close to total feminist residence unit.

`The agent.`
In this simulation, the agent is a 'residence unit', not necessarily an individual. For residence unit, we mean to refer to the family that occupies a house or apartment. Hence, all information on the agent is an average of the members of the unit, unless the unit has only one individual. The 'family' is set using a random-poisson with a mean of 2. The probability of having a residence unit with more than 5 members is less than 3%.

1.2.2 The starting point: Schelling's model

To study how residents settle in a neighbourhood, this simulation takes the original segregation model by Schelling (1971), with a few adjustments.

`Similarity.`
In this simulation, the level of similarity is determined by the level of income that is attributed to each agent via a random normal distribution. The two sliders under the section 'Introduction to house prices' allow users to determine the mean and the standard deviation for income. This is such that agents receive an average amount, and this is visible by their color in the map: *yellow* agents are the middle class, *red* agents have low income, and the *blue* ones have high income. The income is used mostly when the `residential-zones` switch is turned ON. Also, notice that the income considered here is only the portion of a residential unit's money that is dedicated to housing (i.e. it is not their total income).
The slider `similar-wanted` gives the proportion of agents of similar income range (yellow, red, or blue) that each agent would like to have around it before settling. If the proportion of agents, calculated around the 'area-assessment' range, is more than the one indicated in the slider, then the agent moves around to another patch on the map that is consistent with the range expressed in the `search-area`.

`Residential zones` When the switch 'residential-zones' is ON, then the map is divided into 5 different residential areas (visible by clicking on the `see-zone-color` button). These areas are more affordable (cheap) around the middle of the area and less so as one moves to the periphery. The levels of income—as set by the two sliders—determine how easy it will be for agents to find an appropriate residence.
This element works as an additional constraint on the tolerance-based setting. Even though that remains the ultimate assessment that agents make before settling, the definition of residential areas impose a stricter range of options because low income (red) agents cannot afford those residential areas in the periphery and high income agents do not want to settle in the centre of the map.
The mechanism has two stages. In the *first*, each agent compares the income to the price of the residential place (the patch). If this difference is positive, then it considers how much it can afford to save. Savings on residential money are indexed on education, such that more educated individuals are those who save more, and up to a maximum of 20% of their income.
The second step is defined by the "commands for social housing" (bottom left of the panel interface). If `social-housing` is turned ON, then a blue area appears on a random spot in the

map. This is set in between the two areas where residential areas have the lowest prices and it extends depending on the value indicated in the slider `social-housing-area`. The area for social housing (or 'ghetto' as per the Danish system) appears when ticks equal the number inputed in the input box `threshold-counter`.

`Entering and exiting the area`
Some residents may become `tired-of-looking` for a place to settle, after a while (the value indicated in the input box). The value indicates the number of ticks an agent has been looking for a residence without finding any. If, after that many rounds, one is still looking, then it exists the system (meaning that it moves the search elsewhere, and not in this particular residential area we are interested in). Before hitting that number, one may suggest that agents change the area they are looking to settle in. The input box 'relocate' allows users to do that, and agents that have not yet settled choose another place at random where to look for a residence.
It goes without saying that all of this only works if the switch 'enter-exit' is set to ON.

`Value adaptation`.
The assumption of the model is that values adapt very slowly. The commands at the bottom right part of the interface allow one to manipulate the extent to which values (mostly democracy and gender balance) change, based on the surroundings. When the switch `value-adaptation` is ON, then those settled residents scoring an already low level in both of these values, and living in an area where the mean values of these two values is significantly low (calculated in sd distance from the population mean), then the adaptation adjusts depending on the two sliders `democracy-update` and `gender-update`.
A general election with a shocking result would either move towards the area in which the party won (e.g., right-wing in this model; see `elections-RW`) those that are closer to that area in the democracy spectrum. At the same time, those in the opposite side of the spectrum will radicalise, with the impression to become better equipped to fight the extremism of the government. (Only the first part of the argument—i.e. radicalisation—implemented so far.)

1.3 HOW TO USE IT

Before starting the model, users should indicate the number of residents in the area, together with the size of the areas that are residential areas on the map. The first number is set by the slider `num_residents` while the second is defined by the other slider named `residential-area`. When there is incompatibility between these values, a message will pop up and advise on what to do.
The setup button has functions that are standard for `NetLogo` models, while the button `schelling` (continuous and not) has the function of the `go` button in other models. Depending on the setting described above, the simulation performs a classic Schelling segregation model or an adapted (new) version of it.

1.4 THINGS TO NOTICE AND TO TRY

Many features of this model remain similar to those from Schelling's. It is worth noting that small variations, for example, in the use of tolerance (`similar-wanted`) or of the two areas (search and assessment) show results that are very different from the original model by Schelling.

In addition to these results, one could also attempt at moving up the population mean income and reduce its standard deviation, when residential areas are set to ON. This should increase the likelihood for most people to find a residence. Vice-versa, a lower mean income lets the poor (red agents) wonder around without the possibility of settling. When social housing (a ghetto) becomes available, they tend to cram onto that spot in the map.
Also, given that they are the ones to disappear from the map if `enter-exit` is ON, when income is low on average, the area becomes a 'posh' or exclusive neighbourhood.

1.5 EXTENDING THE MODEL

As mentioned above, there are additional steps to extend this model. This is, on the one hand, the first of three models. Hence, obvious extensions relate to the disappearance of the ghetto area (called social housing in the commands) and to the adaptation that agents may have while interacting as neighbours. This second aspect is, perhaps, the most interesting because it allows to study how agents adapt their thinking/behaviour (especially on those three values we have identified above—democracy, tolerance, and gender balance) depending on the area of residence. Adaptation will then make is possible to fully explore the possibility that a parallel society may form, not straight away, but as agents (residents) live in a place. Now implemented! Still checks and improvements needed.
On the other hand, this model constitutes the groundwork for making a study on the areas of Denmark where "ghettoes" are located. In other words, the extension would be to produce a larger scale model (about 100 times larger) and on the actual territories where the government is currently planning on intervening. Such a new simulation would require a computer with vast computational power.

1.6 NETLOGO FEATURES

As social scientists, most features of the software are good enough. However, the area where much work has been dedicated is the one where residential areas have different prices. Nothing too complicated has been created there, but the way the agents relate their income to the area and then apply a version of the 'segregation' model is probably noteworthy. Also, given the modifications to the original mechanisms in Schelling's model, the code features several 'ifelse' loops that may take a while to be fully disentangled and understood. They achieve the intended goal.

1.7 RELATED MODELS

We have been referring to Schelling's model, in the version in which it is available in the NetLogo Models Library, called "segregation". There are many modifications and uses of this model, but its reference has probably been the most inspiring for this work.

1.8 CREDITS AND REFERENCES

The model is part of a paper (currently under review for the *Journal of Simulation*) and it is available on the online repository OpenABM. Full references and url will be available when the related paper is accepted.

2 SENSITIVITY ANALYSIS

The analysis is performed with the small (toy) model of residential areas. The `NetLogo` model was uploaded on the HPC (Abacus 2.0), the Danish supercomputer, for calculations. The version used is `ghetto-coding2.2.nlogo` and 11 computational experiments ran on the infrastructure; most of them took up to 24 HPC hours to complete, generating files of up to 3.34 Gb.
Parameters in the simulation were let vary as much as possible and given the meaningful range, when compared to the full-scale simulation (the one representing the area around Mjølnerparken in København). The full model has $283,071$ patches with $98,654$ that can be used as residential areas (ca. $p.34\%$). The small model has $1,681$ patches in total, with 440 that can be used as residential areas. This is obtained by putting the selector for areas at 1. More detailed descriptions on the features of the model can be found in the 'Info' tab of the model file in `NetLogo`.
The different sets of parameters used in the model can be described as follows. The first column names the parameter, the second gives the value interval for that parameter, then 'SA' (sensitivity analysis) indicates the number through which the value changes and finally, the fourth and last column describes the parameter.
In the following pages, I present results of a sensitivity analysis. I opted for multiple regressions on the outcome variable, letting all parameters take different values one at a time and *ceteris paribus* because it is relatively easy as opposed to other methods (Broeke et al., 2016). The assessment on each parameters will be based on results of the regression, especially on the ΔR^2, only secondarily on the β coefficient estimations.
From the table below, it is possible to determine the number of possible combinations of parameters that is $5 \times 3 \times 6 \times 4 \times 4 \times 2 \times 2 \times 3 \times 2 \times 2 \times 2 \times 2 = 276480$ experiments[1]. Given the size of the exercise, every combination of parameters is performed only once in the simulation environment. Each configuration of parameters evolves over 208 steps, where each step represents one week time and this number equals 52 weeks $\times 4$ years. Results are reported below, after Table 1, accompanied by some descriptive text. See below for the scheme with the conditions and the computational experiments (Table 2).

[1] This is $\times 2$ again, if one considers the split with exit, hence hitting the 552960 experiments.

Table 1: Parameter Notations and Values

Parameter	Values	SA	Description
residents	$[400, 800]$	100	The initial number of agents that are looking to settle and become residents of the area.
residential area, A_i	$[1, 3]$	1	The proportion of the area in the environment that can be occupied by residents The value indicates how many other patches (positions) around a random 100 patches will become available as residential areas.
similar-wanted, S	$[0.3, 0.8]$	0.1	A constant (if 'heterogeneity' is turned off; see below) indicating how many agents similar to oneself one wants around before settling. Similarity is calculated in terms of income.
search area	$[1, 9]$	3	The area around the agent (radius) in which the search for a residence is performed.
assessment area	$[1, 9]$	3	The area around the agent (radius) in which the tolerance/similarity assessment is performed.
heterogeneity	binary	on/off	Diversifies the tolerance levels for each agent, using a random uniform distribution with max = 'similar-wanted'. When agents have a democracy attitude that is high-to-very-high ($\geq$ 1.5) then tolerance for diversity is distributed $\sim \mathcal{N}(0.85, 0.05)$; when it is very low ($\leq 0.1$) then it is a random number between 0 and 0.25.
residential zones	binary	on/off	When 'on', it attributes prices to residences, on 5 levels, from less to more affordable.
income (mean)	$[2, 4]$	1	Mean of a random normal distribution that attributes income to agents. Income serves two purposes, one is that of defining the agent's "status" (low, mid, or high), the other is to make residences affordable or not.
income (standard dev.)	$[1, 2]$	1	Standard deviation of a random normal distribution that attributes income to agents.
social housing (sh)	binary	on/off	It selects a random area (patch) and defines it as social housing, exclusively available to low income agents. The area starts at a point in the simulation that is defined by a threshold.
sh area	$[2, 4]$	2	Expands the area around sh by a radius that is equal to the value.
enter/exit	binary	on/off	Agents choose another area for their search if they have not settled after 20 steps (weeks). After another 20 or 30 steps (weeks), agents who have not settles move out of the area—i.e. they *exit* the simulation. When this happens a random number of new agents *enter*.
democracy attitudes	$\sim \mathcal{N}(1, 0.5)$	–	This is the belief on the value of democracy that each agent is attributed at random, through the function specified. Low levels indicate a dislike for democracy.
gender balance	$\sim \mathcal{N}(0.5, 0.25)$	–	This is the attitudes towards the role of women in society. The function is constrained to values between 0 and 1. Levels close to zero indicate a belief in a male-dominated society while values close to 1 indicate radical feminism. The mid range is a balanced view for both genders.

The outcome variable

The choice here is the number of areas in which there is at least one agent (resident) that settles and has the three values on the low end at the same time. This would be *attitudes toward democracy* < 0.5, *gender balance* $\rightarrow 0$, and *tolerance* > 0.7. This does not mean that these areas of residents around the *divergent* ones are necessarily low on the three values. So, the outcome variable for the sensitivity test is for a potential that these areas may form.

Table 2: Experimental scheme and log with files

#	Exp	res	res-area	sim-wtd	search-a	assmnt-a	het	zones	in(m)	in(sd)	sh	sh-area	ent/ex	t-o-l
1	1.0basics (ref.)	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on/off	off	2.5	1	off	–	off	–
2	1.2_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on/off	off	2.5	1	off	–	on	[30, 40]
3	1.1_soc_housing	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on/off	off	2.5	1	on	[2,4]	off	–
4	1.1.1_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on/off	off	2.5	1	on	[2,4]	on	[30, 40]
5	2.0_res_zones	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on/off	on	[2, 4]	[1, 2]	off	–	off	–
6	2.0.0_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	off	on	[2, 4]	[1, 2]	off	–	on	[30, 40]
7	2.0.1_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on	on	[2, 4]	[1, 2]	off	–	on	[30, 40]
8	2.1.0_soc_housing	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	off	on	[2, 4]	[1, 2]	on	[2,4]	off	–
9	2.1.1_soc_housing	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on	on	[2, 4]	[1, 2]	on	[2,4]	off	–
10	2.1.0_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	off	on	[2, 4]	[1, 2]	on	[2,4]	on	40
11	2.1.1_enter	[400, 800]	[1, 3]	[0.3, 0.8]	[1, 9]	[1, 9]	on	on	[2, 4]	[1, 2]	on	[2,4]	on	40

Note. The conditions are taken from Table 1 above. The experiment 1.0basics is the reference point for all the others.

Analysis of Experiment #1

Given that the area forms even if there is only one agent (with these characteristics) that settles, the new variable in `R` is built by looking at the standard deviation of the democracy values for the areas (see the file `R code ghetto.rtf` for details).
Results shown in Table 3 below come from experiment `#1` (data file `s1.0.csv`), the data sources are always declared in the title of the tables. The models reflect the range in which each parameter varies. At the bottom of each table R^2 appears for the models as well as the differential between models. The idea is that smaller values of variation—intended here as < 0.10 although not strictly interpreted and depending on the relative explanatory power of the parameter—make it such that the range of variation of the parameter affects the outcome variable more or less strongly. A few comments on each table follow below.
Table 3 shows the results for heterogeneity (on/off) and the variations for the res. area parameter. By considering the first two models (mod1.0 and mod1.1), it is apparent that the on/of switch for heterogeneity has a very strong impact on the outcome variable. This means that increasing the variability in the way agents are tolerant also makes it more likely that "deviant" agents would settle. The same trend is repeated when the *residential area* parameter is kept constant and *heterogeneity* is turned on. Overall, this is a switch that should be taken into consideration to study how more or less homogeneous tolerance levels affect the settling of "deviant" individuals. Residential areas variations are explored below.

Table 3: OLS regression results for $s1.0$ (#1) heterogeneity and res.area variations

	mod1.0 h=off	mod1.1 h=on	mod1.0.1 h=off, Ai=1	mod1.1.1 h=on, Ai=1	mod1.0.2 h=off, Ai=2	mod1.1.2 h=on, Ai=2	mod1.0.3 h=off, Ai=3	mod1.1.3 h=on, Ai=3
(Intercept)	6.038^{***}	4.948^{***}	8.553^{***}	6.261^{***}	5.634^{***}	6.035^{***}	4.472^{***}	5.447^{***}
	(0.442)	(0.378)	(0.549)	(0.488)	(0.698)	(0.579)	(0.791)	(0.690)
num.res	0.003^{***}	0.002^{***}	0.002^{*}	-0.000	0.003^{***}	0.003^{***}	0.003^{**}	0.004^{***}
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
res.area	0.091	0.483^{***}						
	(0.085)	(0.073)						
similar.wanted	-2.083^{***}	0.176	-4.882^{***}	-0.132	-0.868	0.486	-0.500	0.175
	(0.408)	(0.349)	(0.549)	(0.488)	(0.698)	(0.579)	(0.791)	(0.690)
search.area	0.076^{***}	0.164^{***}	0.179^{***}	0.284^{***}	0.021	0.136^{***}	0.027	0.071
	(0.023)	(0.020)	(0.031)	(0.028)	(0.039)	(0.033)	(0.045)	(0.039)
area.assessment	-0.160^{***}	-0.725^{***}	-0.345^{***}	-0.726^{***}	-0.146^{***}	-0.783^{***}	0.011	-0.665^{***}
	(0.023)	(0.020)	(0.031)	(0.028)	(0.039)	(0.033)	(0.045)	(0.039)
R-squared	0.075	0.511	0.338	0.628	0.062	0.564	0.023	0.400
ΔR^2 (h)		0.436		0.290		0.502		0.377
ΔR^2 (A_i)		–			0.276	0.064	0.039	0.164
AIC	6892.342	6443.580	2059.916	1947.429	2291.039	2111.144	2411.274	2279.487
N	1440	1440	480	480	480	480	480	480

Table 4 explores variations on the *number of residents* in the simulation. If one considers variations of 100 residents, it is apparent that the effects on the outcome variable is minimal since ΔR^2 is almost always below 0.10. When larger distances are considered, then there is a difference in explanatory power that seems to provide more meaningful results.
Given these findings, the two extreme values will be taken as reference points for the full scale simulation; hence $[400, 800]$.

Table 4: OLS regression results for $s1.0$ (#1), number of residents variations

	mod2.0.0 res=400	mod2.0.1 res=500	mod2.0.2 res=600	mod2.0.3 res=700	mod2.0.4 res=800
(Intercept)	6.622^{***}	7.472^{***}	7.567^{***}	8.743^{***}	8.409^{***}
	(0.497)	(0.510)	(0.549)	(0.510)	(0.542)
res.area	0.096	0.112	0.172	0.461^{***}	0.594^{***}
	(0.126)	(0.129)	(0.139)	(0.129)	(0.138)
similar.wanted	-1.030	-0.652	-0.113	-1.747^{**}	-1.226
	(0.603)	(0.619)	(0.666)	(0.619)	(0.658)
search.area	0.137^{***}	0.172^{***}	0.114^{**}	0.108^{**}	0.069
	(0.034)	(0.035)	(0.038)	(0.035)	(0.037)
area.assessment	-0.292^{***}	-0.393^{***}	-0.445^{***}	-0.533^{***}	-0.547^{***}
	(0.034)	(0.035)	(0.038)	(0.035)	(0.037)
hetero.tolerance: true/false	-1.549^{***}	-1.573^{***}	-1.437^{***}	-1.767^{***}	-1.681^{***}
	(0.206)	(0.211)	(0.228)	(0.211)	(0.225)
R-squared	0.209	0.268	0.251	0.370	0.344
ΔR^2	–	0.059	-0.017	0.119	-0.026
ΔR^2 (2nd order)	–	–	0.042	0.102	0.107
ΔR^2 (3rd order)	–	–	–	0.161	0.076
ΔR^2 (4th order)	–	–	–	–	0.133
AIC	2684.404	2714.388	2800.026	2714.311	2785.581
N	576	576	576	576	576

Next come results on the size of *residential areas* in the simulation model. These are presented in Table 5 (mod2.1.0, mod2.1.1, and mod2.1.2) and it is apparent that the explanatory gaps between models are particularly strong, especially between the first and second and the first and third, even though there is a decrease in explanatory power of these models.

From the results of the first and second order differences, we can retain all of the values; therefore, the full simulation will present variations of this variable for $1, 2, 3$.
The other parameter shown in Table 5 is *similar-wanted* (mod2.2.0 through mod2.2.5). Variations in the 0.1 domain do not seem to have particularly strong effects on explaining the variability of the outcome variable, always with $\Delta R^2 < 0.10$. A second-order variation (i.e. considering 0.2 intervals as opposed to 0.1) gives some variation only with *s-w* between 0.4 and 0.6, and 0.6 and 0.8. On average, deltas are higher.
Variations from lower values of the parameter to the two top values 0.7 and 0.8 seem to offer quite a significant appreciation in the way the ΔR^2 moves. However, the difference between these two values do not seem to be particularly meaningful. It may be good to test one of the low values, a mid-range, and a high value: $0.3, 0.5, 0.8$.

Table 5: OLS regression results for $s1.0$ (#1), res. areas and similar-wanted

	mod2.2.0 s-w=0.3	mod2.2.1 s-w=0.4	mod2.2.2 s-w=0.5	mod2.2.3 s-w=0.6	mod2.2.4 s-w=0.7	mod2.2.5 s-w=0.8	mod2.1.0 r.a.=1	mod2.1.1 r.a.=2	mod2.1.2 r.a.=3
(Intercept)	5.906^{***}	5.901^{***}	5.664^{***}	6.222^{***}	5.596^{***}	5.327^{***}	8.434^{***}	6.575^{***}	5.594^{***}
	(0.722)	(0.703)	(0.647)	(0.647)	(0.597)	(0.602)	(0.398)	(0.497)	(0.567)
num.res	0.003^{**}	0.003^{**}	0.003^{***}	0.002^{**}	0.003^{***}	0.002^{*}	0.001	0.003^{***}	0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
res.area	0.075	0.038	0.319^{*}	0.153	0.466^{***}	0.672^{***}			
	(0.159)	(0.155)	(0.143)	(0.143)	(0.132)	(0.133)			
search.area	0.092^{*}	0.118^{**}	0.118^{**}	0.137^{***}	0.122^{***}	0.131^{***}	0.232^{***}	0.078^{**}	0.049
	(0.043)	(0.042)	(0.038)	(0.038)	(0.035)	(0.036)	(0.022)	(0.028)	(0.031)
area.assessment	-0.314^{***}	-0.340^{***}	-0.429^{***}	-0.429^{***}	-0.485^{***}	-0.657^{***}	-0.535^{***}	-0.464^{***}	-0.327^{***}
	(0.043)	(0.042)	(0.038)	(0.038)	(0.035)	(0.036)	(0.022)	(0.028)	(0.031)
hetero.tolerance: true/false	-2.037^{***}	-1.771^{***}	-1.813^{***}	-1.746^{***}	-1.775^{***}	-0.467^{*}	-2.054^{***}	-1.481^{***}	-1.269^{***}
	(0.260)	(0.253)	(0.233)	(0.233)	(0.215)	(0.216)	(0.134)	(0.167)	(0.191)
similar.wanted							-2.507^{***}	-0.191	-0.162
							(0.392)	(0.490)	(0.559)
R-squared	0.213	0.217	0.310	0.299	0.382	0.449	0.506	0.295	0.159
ΔR^2	–	0.004	0.093	-0.011	0.083	0.067	–	-0.211	-0.236
ΔR^2 (2nd order)	–	–	0.097	0.082	0.072	0.150	–	–	-0.367
ΔR^2 (3rd order)	–	–	–	0.086	0.165	0.139	–	–	–
ΔR^2 (4th order)	–	–	–	–	0.169	0.232	–	–	–
ΔR^2 (4th order)	–	–	–	–	–	0.236	–	–	–
AIC	2374.565	2347.961	2268.471	2269.085	2191.846	2198.973	4133.377	4560.602	4814.533
N	480	480	480	480	480	480	960	960	960

The following Table 6 shows the differences in the *search area* (s.a.) radius and in the *assessment area* (a.a.) radius. They both have taken a four-value variation, including $1, 3, 6, 9$. The search area variation seems to affect the R^2 when the distance is large,hence the two only values considered are 1 and 9.
The s.a. seem to follow a slightly difference logic, with both the last two values being relevant, hence variations taken forward will be $1, 6$ and 9.

Analysis of Experiment #2

Experiment 2 is based on one only variation from #1, that is that the condition *enter/exit* is turned ‘on’. When this is the case, then the ‘t-o-l’ — the time when agents exit the simulation — may take two values (30 or 40). So, we proceed in two steps (presented in Table 7, one looks at differences between the ‘on’ and the ‘off’ mode, while the other considers the ‘on’ mode and the ways in which the two values of the parameter bring difference in the explanatory power of the outcome variable.
As the results show, there is an extremely relevant variation in the explanatory power of the model when the exit/enter is turned to off (false). This leads to conclude that (a) the t-o-l difference can

Table 6: OLS regression results for $s1.0$ (#1), search and assessment areas

	mod2.3.0 s.a.=1	mod2.3.1 s.a.=3	mod2.3.2 s.a.=6	mod2.3.3 s.a.=9	mod2.4.0 a.a.=1	mod2.4.1 a.a.=3	mod2.4.2 a.a.=6	mod2.4.3 a.a.=9
(Intercept)	4.391^{***}	8.049^{***}	7.407^{***}	7.604^{***}	3.054^{***}	3.907^{***}	3.910^{***}	5.902^{***}
	(0.553)	(0.621)	(0.626)	(0.606)	(0.565)	(0.510)	(0.429)	(0.357)
num.res	0.004^{***}	0.001^{*}	0.002^{**}	0.002^{**}	0.006^{***}	0.002^{**}	0.002^{***}	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
res.area	0.925^{***}	0.075	0.129	0.019	-0.983^{***}	1.081^{***}	0.812^{***}	0.237^{***}
	(0.109)	(0.122)	(0.123)	(0.119)	(0.111)	(0.100)	(0.084)	(0.070)
similar.wanted	-1.367^{**}	-1.024	-0.514	-0.910	1.452^{**}	0.169	-2.090^{***}	-3.345^{***}
	(0.520)	(0.584)	(0.588)	(0.570)	(0.531)	(0.479)	(0.403)	(0.336)
area.assessment	-0.463^{***}	-0.475^{***}	-0.430^{***}	-0.401^{***}				
	(0.029)	(0.033)	(0.033)	(0.032)				
hetero.tolerance: true/false	-1.972^{***}	-1.700^{***}	-1.494^{***}	-1.239^{***}	0.306	-0.075	-2.822^{***}	-3.814^{***}
	(0.178)	(0.199)	(0.201)	(0.195)	(0.181)	(0.164)	(0.138)	(0.115)
search.area					0.022	0.157^{***}	0.198^{***}	0.102^{***}
					(0.030)	(0.027)	(0.023)	(0.019)
R-squared	0.412	0.288	0.246	0.226	0.199	0.183	0.472	0.636
ΔR^2	–	-0.124	-0.042	-0.020	–	-0.016	0.289	0.164
ΔR^2 (2nd order)	–	–	-0.166	-0.042	–	–	0.273	0.453
ΔR^2 (3rd order)	–	–	–	-0.186	–	–	–	0.437
AIC	3301.232	3468.754	3479.456	3433.370	3330.883	3183.922	2935.719	2672.189
N	720	720	720	720	720	720	720	720

be discarded completely and (b) turning enter/exit off would result in making most of the parameters from statistically significant to non-significant, in the base case, at least.

Table 7: OLS regression results for $s1.2$ (#2) for enter/exit and t-o-l

	mod1.2.0 ent/ex=off	mod1.2.1 ent/ex=on	mod1.2.2 t-o-l=30	mod1.2.3 t-o-l=40
(Intercept)	6.294^{***}	4.456^{***}	4.200^{***}	4.711^{***}
	(0.313)	(0.262)	(0.372)	(0.368)
num.res	0.002^{***}	0.001^{***}	0.001^{**}	0.001^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287^{***}	0.162^{**}	0.214^{**}	0.109
	(0.060)	(0.050)	(0.071)	(0.070)
similar.wanted	-0.954^{***}	0.136	0.368	-0.095
	(0.286)	(0.238)	(0.339)	(0.335)
search.area	0.120^{***}	-0.020	-0.025	-0.014
	(0.016)	(0.013)	(0.019)	(0.019)
area.assessment	-0.442^{***}	-0.020	0.010	-0.050^{**}
	(0.016)	(0.013)	(0.019)	(0.019)
hetero.tolerance: true/false	-1.601^{***}	0.107	0.146	0.068
	(0.098)	(0.081)	(0.116)	(0.115)
R-squared	0.288	0.006	0.008	0.007
ΔR^2	–	-0.282	–	-0.001
ΔR^2 (w/1.0)	–	–	-0.200	-0.201
AIC	13724.082	29356.411	14713.779	14648.343
N	2880	5760	2880	2880

Analysis of Experiment #3

This experiment introduces an area of *social housing* in the environment, available to low income residents. This is the only aspect that is different from Exp #1; for this reason, it could be easily compared to Exp #2 as well. Table 8 does exactly that, although taking only the general true/false (on/off) effect from Exp #2.

There is no difference between the two sizes of the areas 2 or 4, and the existence of the *social housing* area makes the models non statistically significant overall, explaining a very small portion of variance.

Table 8: OLS regression results for $s1.1$ (#3) for social housing and s.h. area

	mod1.0.0 mod1.0 exp1	mod1.2.0 mod1.2 exp2	mod1.1.0 s.h.=on	mod1.1.1 s.h.area=2	mod1.1.2 s.h.area=4
(Intercept)	6.294^{***}	4.456^{***}	4.713^{***}	4.654^{***}	4.772^{***}
	(0.313)	(0.262)	(0.262)	(0.370)	(0.372)
num.res	0.002^{***}	0.001^{***}	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287^{***}	0.162^{**}	0.164^{**}	0.207^{**}	0.120
	(0.060)	(0.050)	(0.050)	(0.071)	(0.071)
similar.wanted	-0.954^{***}	0.136	0.142	0.036	0.248
	(0.286)	(0.238)	(0.239)	(0.337)	(0.339)
search.area	0.120^{***}	-0.020	-0.002	0.014	-0.017
	(0.016)	(0.013)	(0.013)	(0.019)	(0.019)
area.assessment	-0.442^{***}	-0.020	0.007	0.013	0.001
	(0.016)	(0.013)	(0.013)	(0.019)	(0.019)
hetero.tolerance: true/false	-1.601^{***}	0.107	0.090	0.081	0.099
	(0.098)	(0.081)	(0.082)	(0.115)	(0.116)
R-squared	0.288	0.006	0.003	0.004	0.002
ΔR^2 (from exp.1)	–	-0.282	-0.285	-0.284	-0.286
ΔR^2 (from exp.2)	–	–	-0.003	-0.002	-0.004
ΔR^2 (between)	–	–	–	–	-0.002
AIC	13724.082	29356.411	29377.207	14682.262	14708.070
N	2880	5760	5760	2880	2880

Given the limited impact of this variable, we keep just one condition for the area, 4. *Social housing* can be left 'on' because it shows that there is no or very limited impact on the outcome variable.

Analysis of Experiment #4

The analysis of results from Table 9 confirms the findings from the experiments above. The *enter/exit* parameter is meaningful although its effect is low, and it remains low even when *social housing* is on. No meaningful difference is found for the two values of *t-o-l*. The two sh-area differences are not tested, since results do not seem to be particularly revealing anyway.[2]

Analysis of Experiment #5

Table 10 compares results from Exp.#5 to the baseline experiment (#1) and to the variations of mean and standard deviation of the income. Overall, the making of *residential zones* affects the outcome variable so the switch should be turned on. On the contrary, the difference between the different levels of the mean do not seem particularly relevant, hence it is dropped and left at 2.5 as in the first four experiments. The variation of standard deviation is instead relevant and both levels are kept, with st. dev $[1, 2]$.

Analysis of Experiment #6

This experiment confirms some of the results introduced above. For example, larger standard deviation for income seem to affect results more strongly when heterogeneity is off, and mean

[2] A quick run with two regressions shows that the R^2 between the two conditions is very similar and around 0.34 for most conditions, reaching 0.3907 for *t-o-l*$= 30$ and *s.h. area*$= 4$. In the end, this is only a slight difference from the other conditions, just ≈ 0.05. Hence, it is not included in Table 9.

Table 9: OLS regression results for $s1.1.1$ (#4) for enter/exit and t-o-l

	mod1.0.0 mod1.0 exp1	mod1.2.0 mod1.2 exp2	mod1.1.0 mod1.1 exp3	mod1.4.0 ent/ex=on	mod1.4.1 t-o-l = 30	mod1.4.2 t-o-l = 40
(Intercept)	6.294***	4.456***	4.713***	8.958***	8.801***	9.116***
	(0.313)	(0.262)	(0.262)	(0.163)	(0.228)	(0.232)
num.res	0.002***	0.001***	0.000	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287***	0.162**	0.164**	0.048	0.092*	0.004
	(0.060)	(0.050)	(0.050)	(0.031)	(0.043)	(0.044)
similar.wanted	−0.954***	0.136	0.142	−1.818***	−2.056***	−1.579***
	(0.286)	(0.238)	(0.239)	(0.148)	(0.208)	(0.212)
search.area	0.120***	−0.020	−0.002	0.014	0.028*	−0.000
	(0.016)	(0.013)	(0.013)	(0.008)	(0.012)	(0.012)
area.assessment	−0.442***	−0.020	0.007	−0.534***	−0.538***	−0.530***
	(0.016)	(0.013)	(0.013)	(0.008)	(0.012)	(0.012)
hetero.tolerance: true/false	−1.601***	0.107	0.090	−2.222***	−2.322***	−2.122***
	(0.098)	(0.081)	(0.082)	(0.051)	(0.071)	(0.072)
R-squared	0.288	0.006	0.003	0.351	0.367	0.336
ΔR^2 (from exp.1)	–	−0.282	−0.285	0.079	0.085	0.054
ΔR^2 (from exp.2)	–	–	−0.003	0.345	0.361	0.330
ΔR^2 (from exp.3)	–	–	–	0.348	0.363	0.333
ΔR^2 (between)	–	–		–	–	−0.031
AIC	13724.082	29356.411	29377.207	55740.021	27755.316	27974.384
N	2880	5760	5760	11520	5760	5760

Table 10: OLS regression results for $s2.0$ (#5) for res. zones, income (m) and income (stdev)

	mod1.0.0 mod1.0 exp1	mod1.5.0 r.z.=on	mod1.5.1 st.dev=1	mod1.5.2 st.dev=2	mod1.5.3 mean=2	mod1.5.4 mean=3	mod1.5.5 mean=4
(Intercept)	6.294***	−0.466***	−1.021***	0.176	0.530**	2.679***	4.961***
	(0.313)	(0.112)	(0.171)	(0.186)	(0.188)	(0.222)	(0.251)
num.res	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287***	0.847***	0.909***	0.811***	0.781***	0.904***	0.815***
	(0.060)	(0.014)	(0.024)	(0.026)	(0.027)	(0.032)	(0.036)
similar.wanted	−0.954***	−0.997***	−1.059***	−0.890***	−0.823***	−1.289***	−0.873***
	(0.286)	(0.069)	(0.114)	(0.124)	(0.129)	(0.153)	(0.173)
search.area	0.120***	0.254***	0.262***	0.246***	0.242***	0.290***	0.215***
	(0.016)	(0.004)	(0.006)	(0.007)	(0.007)	(0.009)	(0.010)
area.assessment	−0.442***	−0.357***	−0.380***	−0.336***	−0.277***	−0.380***	−0.410***
	(0.016)	(0.004)	(0.006)	(0.007)	(0.007)	(0.009)	(0.010)
hetero.tolerance: true/false	−1.601***	−1.383***	−1.321***	−1.345***	−1.002***	−1.434***	−1.581***
	(0.098)	(0.023)	(0.039)	(0.042)	(0.044)	(0.052)	(0.058)
income.mean		1.033***	1.257***	0.779***			
		(0.016)	(0.027)	(0.030)			
income.stdev		0.015			0.463***	0.069	−0.520***
		(0.029)			(0.053)	(0.063)	(0.071)
R-squared	0.288	0.467	0.524	0.410	0.427	0.466	0.398
ΔR^2 (from exp.1)	–	0.179	0.236	0.322	0.139	0.178	0.110
ΔR^2 (1st)	–	–	–	−0.114	–	0.039	−0.068
ΔR^2 (2nd)	–	–	–	–	–	–	−0.029
AIC	13724.082	115292.545	37601.546	39149.839	21106.069	22950.655	24317.117
N	2880	27705	9236	9234	5542	5540	5541

differences (still on income) do not affect results meaningfully from Experiment 1; however, they seem to do from Experiment 5 although they all have very similar effects (in explaining less). Another result that is confirmed from the above entails the parameter *tired of looking* (or t-o-l in Table 12). Variation is amounts to 0.000, hence we can confirm this difference is discarded.

Overall, we move to accept the effects of parameter *enter/exit* but not t-o-l. Same as above, income mean differences are not useful to consider while standard deviation income differences are.

Table 11: OLS regression results for $s2.0.0$ (#6) for res. zones, income (m & stdev), and enter/exit

	mod1.0.0 mod1.0 exp1	mod1.5.0ht_on exp5 Ht=on	mod2.6.0 base exp6	mod2.6.1 t-o-l=30	mod2.6.2 t-o-l=40	mod2.6.3 st.dev=1	mod2.6.4 st.dev=2	mod2.6.5 mean=2	mod2.6.6 mean=3	mod2.6.7 mean=4
(Intercept)	6.294^{***}	-0.599^{***}	1.117^{***}	0.667^{**}	1.568^{***}	-0.637^{*}	1.630^{***}	2.025^{***}	4.462^{***}	6.290^{***}
	(0.313)	(0.141)	(0.179)	(0.254)	(0.253)	(0.292)	(0.325)	(0.351)	(0.362)	(0.392)
num.res	0.002^{***}	0.001^{***}	0.003^{***}	0.003^{***}	0.002^{***}	0.002^{***}	0.003^{***}	0.002^{***}	0.003^{***}	0.003^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287^{***}	0.744^{***}	0.824^{***}	0.812^{***}	0.836^{***}	0.978^{***}	0.670^{***}	0.726^{***}	0.984^{***}	0.762^{***}
	(0.060)	(0.018)	(0.026)	(0.037)	(0.036)	(0.034)	(0.038)	(0.042)	(0.043)	(0.047)
similar.wanted	-0.954^{***}	0.028	-2.608^{***}	-2.516^{***}	-2.699^{***}	-2.799^{***}	-2.416^{***}	-2.525^{***}	-3.121^{***}	-2.176^{***}
	(0.286)	(0.087)	(0.120)	(0.170)	(0.169)	(0.159)	(0.177)	(0.195)	(0.201)	(0.218)
search.area	0.120^{***}	0.260^{***}	0.024^{***}	0.029^{**}	0.019^{*}	0.060^{***}	-0.012	-0.013	0.061^{***}	0.023
	(0.016)	(0.005)	(0.007)	(0.010)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)
area.assessment	-0.442^{***}	-0.508^{***}	-0.392^{***}	-0.392^{***}	-0.391^{***}	-0.414^{***}	-0.369^{***}	-0.338^{***}	-0.434^{***}	-0.403^{***}
	(0.016)	(0.005)	(0.007)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)
het. (true/false)	-1.601^{***}									
	(0.098)									
income.mean		0.880^{***}	1.319^{***}	1.335^{***}	1.304^{***}	1.612^{***}	1.026^{***}			
		(0.021)	(0.025)	(0.035)	(0.035)	(0.033)	(0.037)			
income.stdev		0.003	0.130^{**}	0.171^{**}	0.088			0.778^{***}	0.006	-0.395^{***}
		(0.036)	(0.041)	(0.058)	(0.057)			(0.066)	(0.068)	(0.074)
tired.of.looking						0.027^{***}	0.019^{**}	0.024^{***}	0.029^{***}	0.017^{*}
						(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
R-squared	0.288	0.554	0.387	0.388	0.388	0.484	0.309	0.281	0.362	0.272
ΔR^2 (w/exp1)	–	0.266	0.099	0.100	0.100	0.196	0.021	-0.007	0.074	-0.016
ΔR^2 (w/exp5)	–	–	-0.167	-0.166	-0.166	-0.070	-0.245	-0.273	-0.192	-0.282
ΔR^2 (1st)	–	–	–	–	0.000	–	-0.174	–	0.081	-0.090
AIC	13724.082	54617.225	55519.449	27789.281	27709.105	26936.608	28266.127	17981.882	18228.465	18884.920
N	2880	13846	12413	6207	6206	6207	6206	4140	4137	4136

Analysis of Experiment #7

Experiment #7 is to test the effect of *enter/exit* from the simulation, when *heterogeneity* is set to 'on'. The effect on the outcome variable is overall very strong compared to the baseline model (Exp #1). As previously, *t-o-l* change does not increase explanatory power of the regression model. Contrary to the other regression analyses, here standard deviation of the income does not significantly affect results while the mean (especially *mean income* $= 3$) is relevant. Results are visible from Table 12

We can confirm results as above and do not register any change in the test model.

The other Table 13 is to compare Experiment #6 with Experiment #7. The largest decreases in R^2 are with an income standard deviation of 2 and with mean of 4.

Experiment #8, #9, #10, and #11

At this point, we have already enough information to update our knowledge of which parameter values to carry over in the general testing.

The existing factorial design tested to determine the impact of the parameters—i.e. sensitivity analysis—is shown in Table 14 as well as the experiment in which they have been tested and the update to be carried in the main model testing.

From the data introduced in Table 14, a number for the repetitions one may perform per configuration of parameters must be found. We may take one from the above regressions to estimate effect size for a power analysis calculation (Secchi and Seri, 2017). By taking mod1.0basics (from Exp. #1), we have that *SSR*$= 7979.633$ and *SST*$= 27660.83$ hence one may

Table 12: OLS regression results for $s2.0.1$ (#7) for same as #6 but hetero-tol.=ON

	mod1.0.0 mod1.0 exp1	mod1.5.0ht_on exp5 Ht=on	mod2.7.0 base exp7	mod2.7.1 t-o-l=30	mod2.7.2 t-o-l=40	mod2.7.3 st.dev=1	mod2.7.4 st.dev=2	mod2.7.5 mean=2	mod2.7.6 mean=3	mod2.7.7 mean=4
(Intercept)	6.294^{***}	-0.599^{***}	2.884^{***}	2.672^{***}	3.096^{***}	1.058^{***}	2.098^{***}	2.009^{***}	4.932^{***}	7.285^{***}
	(0.313)	(0.141)	(0.153)	(0.216)	(0.216)	(0.266)	(0.276)	(0.287)	(0.329)	(0.337)
num.res	0.002^{***}	0.001^{***}	0.000^{**}	0.001^{***}	0.000	0.000	0.001^{***}	0.001^{***}	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
res.area	0.287^{***}	0.744^{***}	0.480^{***}	0.422^{***}	0.538^{***}	0.621^{***}	0.339^{***}	0.506^{***}	0.554^{***}	0.380^{***}
	(0.060)	(0.018)	(0.023)	(0.033)	(0.033)	(0.032)	(0.033)	(0.035)	(0.040)	(0.041)
similar.wanted	-0.954^{***}	0.028	-0.147	-0.169	-0.125	-0.271	-0.023	-0.167	0.042	-0.316
	(0.286)	(0.087)	(0.109)	(0.153)	(0.154)	(0.150)	(0.156)	(0.166)	(0.190)	(0.195)
search.area	0.120^{***}	0.260^{***}	-0.045^{***}	-0.046^{***}	-0.043^{***}	-0.037^{***}	-0.052^{***}	-0.074^{***}	-0.039^{***}	-0.021
	(0.016)	(0.005)	(0.006)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.011)
area.assessment	-0.442^{***}	-0.508^{***}	-0.717^{***}	-0.703^{***}	-0.730^{***}	-0.712^{***}	-0.721^{***}	-0.528^{***}	-0.753^{***}	-0.869^{***}
	(0.016)	(0.005)	(0.006)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.011)
het. (true/false)	-1.601^{***}									
	(0.098)									
income.mean		0.880^{***}	1.054^{***}	1.053^{***}	1.055^{***}	1.253^{***}	0.855^{***}			
		(0.021)	(0.023)	(0.032)	(0.032)	(0.031)	(0.032)			
income.stdev		0.003	-0.002	-0.049	0.045			0.462^{***}	-0.133^{*}	-0.335^{***}
		(0.036)	(0.037)	(0.052)	(0.052)			(0.056)	(0.064)	(0.066)
tired.of.looking						0.033^{***}	0.042^{***}	0.034^{***}	0.044^{***}	0.034^{***}
						(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
R-squared	0.288	0.554	0.511	0.504	0.522	0.540	0.496	0.408	0.502	0.554
ΔR^2 (w/exp1)	–	0.266	0.223	0.216	0.234	0.252	0.208	0.120	0.214	0.266
ΔR^2 (w/exp5)	–	–	-0.043	-0.050	-0.032	-0.014	-0.058	-0.146	-0.052	-0.000
ΔR^2 (1st)	–	–	–	–	0.018	–	-0.044	–	0.094	-0.052
AIC	13724.082	54617.225	71270.986	35575.231	35593.053	35233.495	35812.787	22403.877	23824.742	24082.370
N	2880	13846	15761	7881	7880	7881	7880	5256	5253	5252

Table 13: OLS regression results for $s2.0.1$ (#7) compared to $s2.0.0$ (#6)

	mod2.7.0 base exp7	mod2.7.1 t-o-l=30	mod2.7.2 t-o-l=40	mod2.7.3 st.dev=1	mod2.7.4 st.dev=2	mod2.7.5 mean=2	mod2.7.6 mean=3	mod2.7.7 mean=4
R-squared (Exp.7)	0.511	0.504	0.522	0.540	0.496	0.408	0.502	0.554
	mod2.6.0 base exp6	mod2.6.1 t-o-l=30	mod2.6.2 t-o-l=40	mod2.6.3 st.dev=1	mod2.6.4 st.dev=2	mod2.6.5 mean=2	mod2.6.6 mean=3	mod2.6.7 mean=4
R-squared (Exp.6)	0.387	0.388	0.388	0.484	0.309	0.281	0.362	0.272
ΔR^2 (Exp.7/Exp.6)	-0.124	-0.113	-0.134	-0.066	-0.187	-0.127	-0.140	-0.282

Table 14: From sensitivity to testing the model

SA	what	Exp. #	MD	comments
$5\times$	residents	1	$2\times$	400 and 800
$3\times$	res. area	1	$3\times$	
$6\times$	similar-wanted	1	$3\times$	$0.3, 0.5, 0.8$
$4\times$	search area	1	$2\times$	only 1 and 9
$4\times$	assessment area	1	$3\times$	$1, 6, 9$
$2\times$	heterogeneity	2	$2\times$	
$2\times$	res. zones	5	$2\times$	
$3\times$	income (mean)	5, 6, 7	–	the value is 2.5
$2\times$	income (st. dev.)	5, 6, 7	$2\times$	
$2\times$	social housing	3	$2\times$	areas set to 4
$2\times$	sh. area	3	–	
$2\times$	enter/exit	4, 6, 7	$2\times$	t-o-l$=40$
$= 276480$			$= 3456$	

calculate $f^2 = 0.4054445$. The other inputs of the formula are the degrees of freedom of the F-statistic (Seri and Secchi, 2018). For the numerator, one needs the number of independent variables in the model X_n. We can take that the full model will have 10 (again, see Table 14). The denominator has the number of observations N (runs, in our case) minus the number of variables in the model $X_n - 1$. N is the number observations one needs. However, given that the effect size is particularly large, one run per configuration of parameters would do. (The number that actually comes out of the calculation is 89.25952.)
If one takes, instead a regression model with an effect size that is micro (such as the ones from Exp. #2 or Exp. #3) then $f^2 = \frac{R^2}{1-R^2} = 0.006/0.994 = 0.006$. The N becomes 5221.217. Given that the number of runs is 3456 then repetitions amount at $5221/3456$ that is 1.5107. This would lead us to determine that the number of repetitions per run is going to be just 2 (to take the safe side) although 1 would also do.

```
> pwr.f2.test(10, NULL, 0.006, 0.01, 0.95)$v + 11
[1] 5221.217
> 2^7 * 3^3
[1] 3456
> 5221 / 3456
[1] 1.510706
```

A relatively different take — i.e. using ANOVA — brings a different result (Secchi and Seri, 2017). Here is the formula and the calculation:

```
> pwr.anova.test(3456, NULL, sqrt(0.006), 0.01, 0.95)

     Balanced one-way analysis of variance power calculation

              k = 3456
              n = 17.16575
              f = 0.07745967
      sig.level = 0.01
          power = 0.95

NOTE: n is number in each group
```

If we were to compare results with an ANOVA-like approach then we would need 17 runs per group. I have done this to see how distant the calculation will be and the number of runs is still small (although more than 2). However, we will be using regression and the calculation made using the smallest effect size from the regression ($\sqrt{f^2}$) is not indicative of the average effect size. Taking a SESOI (smalles effect size of interest) approach (Seri and Secchi, 2017), the smallest ES is the one to take. Again, with a regression this is in between 1 and 2.

3 ADDITIONAL FINDINGS

The following Table 15 features multiple OLS regression analyses to explore results from the simulation. Data is taken from a file called `parso.work.csv`.

Table 15: Multiple OLS regressions results

	mod1.0	mod1.1	mod1.2	mod2.0
	ps.total	ps.mean.res	ps.mean.res	ps.mean.res
(Intercept)	−0.018	−1.287*	−1.082	−20.122***
	(0.025)	(0.622)	(0.683)	(1.357)
num.res	0.000***	0.002***	0.002**	0.007***
	(0.000)	(0.000)	(0.001)	(0.001)
res.area	−0.009*	−0.283*	−0.256	2.150***
	(0.004)	(0.118)	(0.145)	(0.237)
similar.wanted	0.134***	0.606	0.684	−6.798***
	(0.016)	(0.369)	(0.398)	(0.875)
search.area	−0.001	−0.022	−0.022	−0.276***
	(0.001)	(0.019)	(0.023)	(0.046)
area.assessment	−0.017***	0.294***	0.220**	2.934***
	(0.001)	(0.067)	(0.077)	(0.081)
hetero.tolerance: true/false	0.072***	−0.189		−5.029***
	(0.007)	(0.178)		(0.369)
res.zones: true/false	−0.004	0.314	0.205	1.561***
	(0.008)	(0.285)	(0.338)	(0.420)
income.stdev	−0.002	−0.455**	−0.439*	−1.314***
	(0.006)	(0.139)	(0.171)	(0.348)
social.housing: true/false	−0.008	0.052	−0.100	−0.001
	(0.007)	(0.142)	(0.177)	(0.367)
enter.exit: true/false	−0.014*	0.259	0.166	1.555***
	(0.007)	(0.153)	(0.191)	(0.380)
low	0.028*	1.475***	1.752***	4.902***
	(0.013)	(0.362)	(0.447)	(0.696)
mid	0.055***	0.276	−0.176	8.138***
	(0.015)	(0.567)	(0.712)	(0.846)
high	0.052***	−0.161	−0.052	−6.915***
	(0.014)	(0.499)	(0.605)	(0.790)
ps.mean.income	0.003	0.961***	0.924***	6.652***
	(0.002)	(0.048)	(0.059)	(0.100)
R-squared	0.109	0.635	0.604	0.577
F statistic	60.005	46.770	32.093	672.572
p-value	0.000	0.000	0.000	0.000
N	6912	392	288	6912

Note: ps.total = the number of “alternative” areas in a run; ps.mean.res = the number of residents in “alternative” areas (only when areas become available in model 1.1 and 1.2).

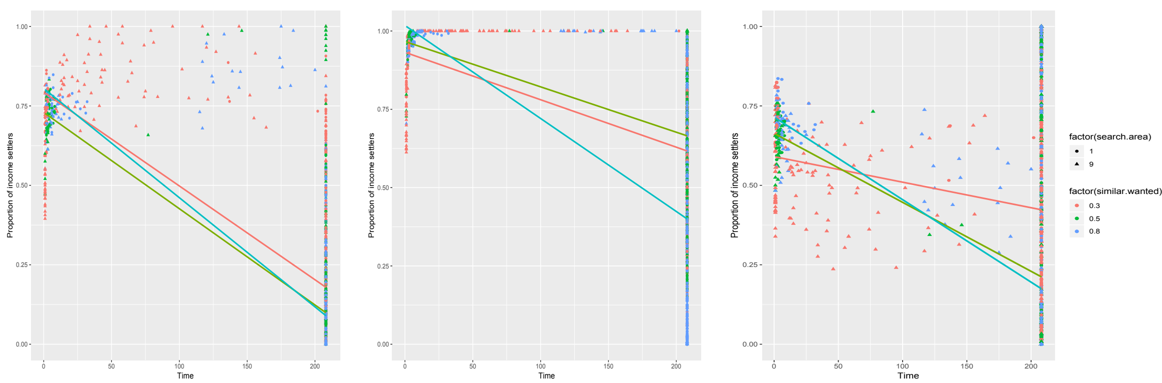
3.1 Exploring the residential terrain

For reasons related to the max length of a paper in the *Journal of Simulation*, we could not include more than five plots. The paper presents only those results that are most relevant to justify the findings. However, the journey we took to get there included drawing a number of plots that we decided were less informative than the ones presented in the paper. Nevertheless these plots have an information value and this is the reason why we have decided to include them in the following pages of this document.
We will not be writing comments on each one of the figures below, but try to highlight the most meaningful trends and results.

3.1.1 Working from Schelling's model

The first plots we present below offer a check of whether the assumptions made work as expected. Each figure presents results comparing the low to the middle and to the high income residents. All results are also split between the area in which residents search a suitable place and by the degree of tolerance (similar-wanted).

At the start of the simulation (step = 0), there are already residents who found a suitable place and settled. There is usually a higher proportion for middle income residents, because they constitute the majority in the system — as per our initial conditions. Figure 1 and 2 show that higher tolerance — i.e. lower intolerance levels, because agents want 0.3, 0.5, or 0.8 other agents that are similar to them in order for them to settle — consistently shows a trend where around 0.70 of residents settle as the simulation start. This happens as the levels are distributed in a non heterogeneous way (see Table 1 above and the paper for details). Allowing individuals to enter and exit the system seem to bear effects only as intolerance grows, i.e. for levels > 0.3, as it becomes more difficult to find suitable settlement. This latter aspect is a departure from the classic model by Schelling (1971).

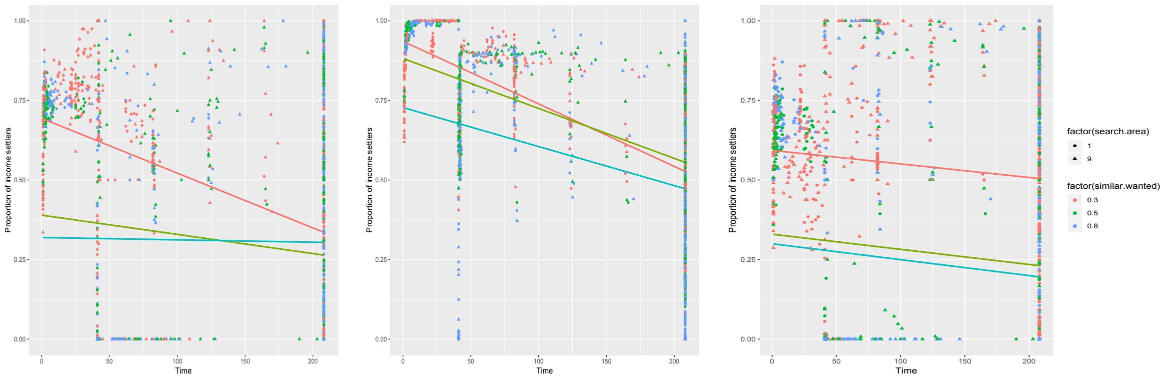


(a) Low income settlers. (b) Mid income settlers. (c) High income settlers.

Figure 1: Proportion of settled residents by time, with `hetero-tolerance` = OFF and `enter/exit` = OFF, (reg. curve on `similar-wanted`, S).

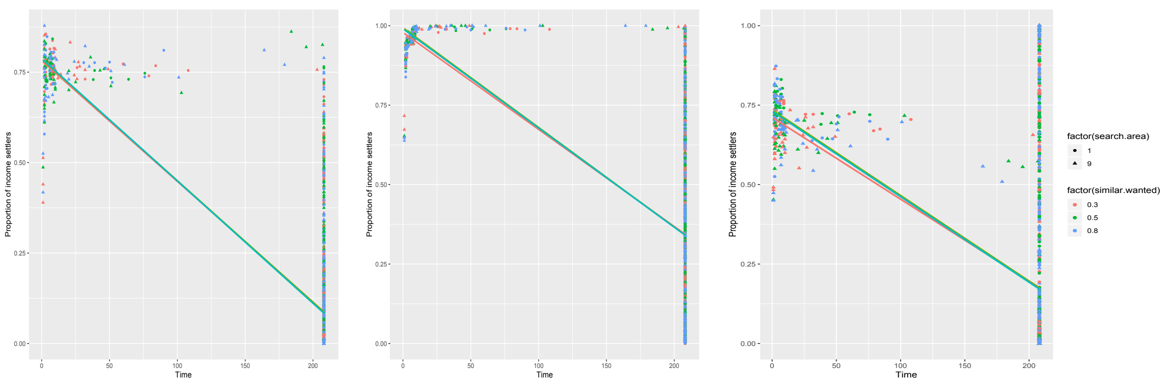
The other figures below, Figure 3 and 4, depart from Schelling in that we allow for heterogeneity to be factored in. This means that intolerance cannot be more than the number shown by the parameter `similar-wanted` and it can take all the other values up to it. The effect of this solution in a closed system — i.e. when the `enter/exit` is OFF — collapses all results in one, canceling the difference imposed by homogeneous preferences (especially Figure 3). In other words, it does not matter where the system maxes out in terms of intolerance, as far as the remaining residents have a heterogeneous set of preferences (tolerance levels).

There seems to be some convergence also when `enter/exit` is set to ON, but results are slightly different when one considers low and high income levels (Figure 4). In these cases, settlement levels appear to show very small fluctuations as the system progresses in time.



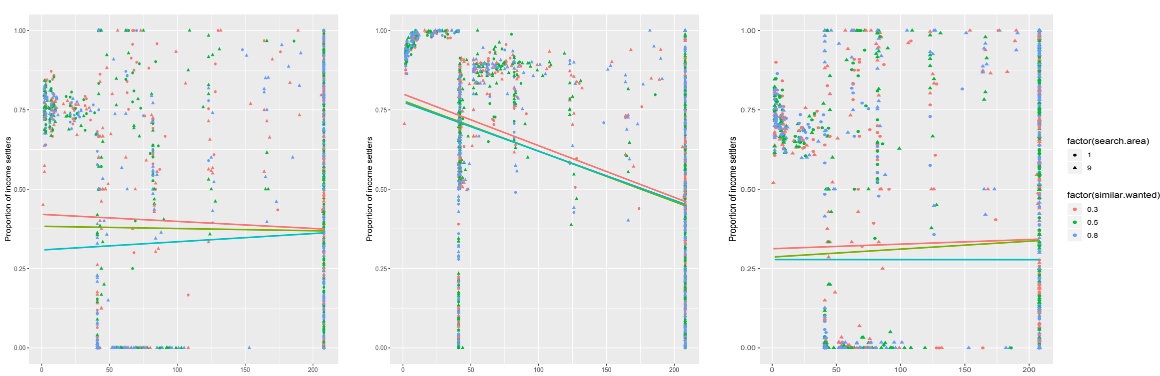
(a) Low income settlers. (b) Mid income settlers. (c) High income settlers.

Figure 2: Proportion of settled residents by time, with `hetero-tolerance` = OFF and `enter/exit` = ON, (reg. curve on `similar-wanted`, S).



(a) Low income settlers. (b) Mid income settlers. (c) High income settlers.

Figure 3: Proportion of settled residents by time, with `hetero-tolerance` = ON and `enter/exit` = OFF, (reg. curve on `similar-wanted`, S).

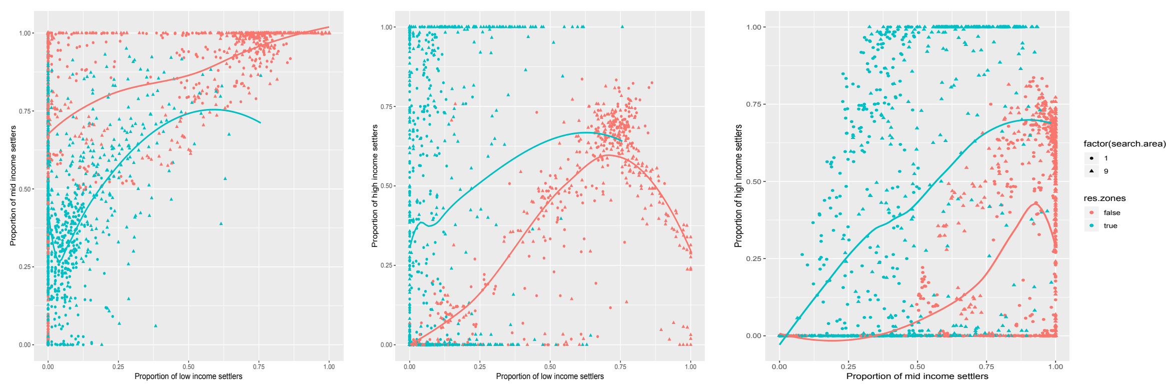


(a) Low income settlers. (b) Mid income settlers. (c) High income settlers.

Figure 4: Proportion of settled residents by time, with `hetero-tolerance` = ON and `enter/exit` = ON, (reg. curve on `similar-wanted`, S).

3.1.2 Comparing residential settling by income

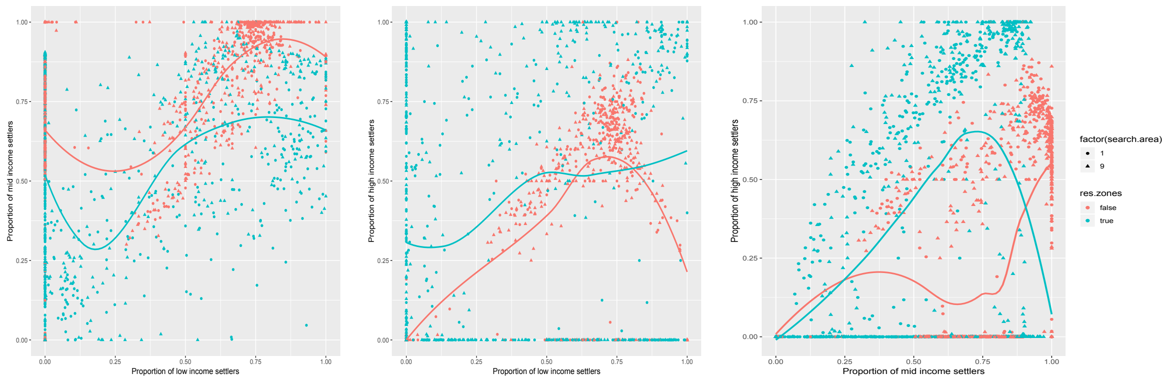
The figures below compare residential settlements by income level, to understand under which conditions opportunities favour one group over the others.
Figure 5 is particularly revealing when one observes low income residents. In fact, when the `search area` is at its highest level 9, there is polarisation, visible by the triangles, mostly concentrated in the middle of Figure 5a and 5b. This means that, under this condition, there seem to be a more balanced position although mid income residents seem to be favoured under all circumstances to low income residents (Figure 5a).
The colour split in Figure 5a indicates that delimitation of `residential zones` is more advantageous for mid income residents, independent of the other conditions, although `residential area` $= 9$ helps raise the number of settlers a bit. In the other two panes, 5b and 5c, the comparison varies although the effect of `residential zones` is clearly visible while the one from `residential area` search is not that clear. The absence of the former clearly favours settlement of both high and low income residents, while their presence favours high income residents (see the upper left side of Figure 5b). The situation is more skewed towards mid income in the absence of `residential zones` while it favours high income residents in the case of their presence (see Figure 5c).



(a) Low vs Mid income settlers. (b) Low vs High income settlers. (c) Mid vs High income settlers.

Figure 5: Proportion of settled residents by income comparisons, with `hetero-tolerance` = OFF and `enter/exit` = OFF, (reg. curve on `res.zones`).

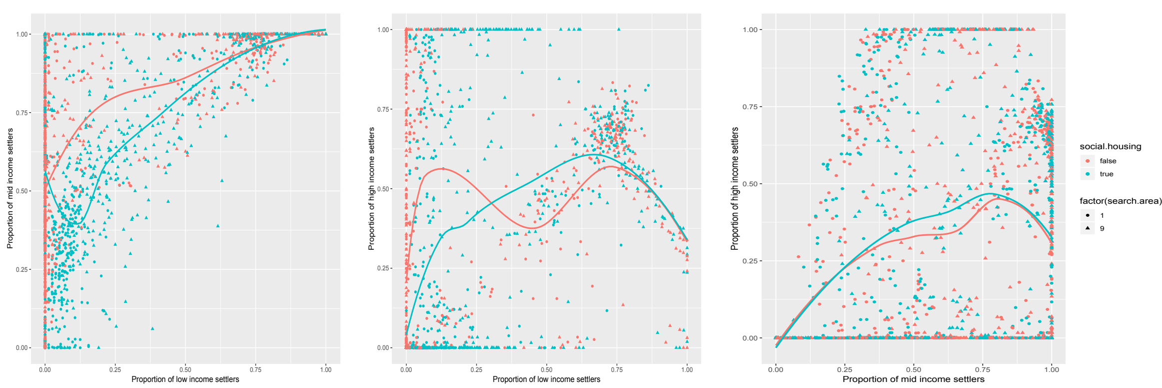
The findings presented in Figure 6 introduce the possibility of unsatisfied settlers to leave the system and it allows for new potential residents to come in at every step (`enter/exit` is set to ON). Most results change, with the exception of high vs mid income (Figure 6c), where there seems to be a higher likelihood of settling when `residential zones` is set to ON. The other two panes, Figure 6a and 6b, show that low income residents benefit from the possibility of entering and leaving the system, more so when `residential zones` are present.
Figure 7 presents results on the introduction of `social housing` and indicates that, as expected, the opportunities for low income residents to settle increase (7a and 7b). Instead, the effect is almost non-existent for the other two categories, again, as expected. The regression curve (Local Polynomial Regression Fitting curve or LOESS) when social housing is introduced is — see the first pane (7a) — lower than the other, indicating that higher proportions of low income



(a) Low vs Mid income settlers. (b) Low vs High income settlers. (c) Mid vs High income settlers.

Figure 6: Proportion of settled residents by income comparisons, with `hetero-tolerance` = OFF and `enter/exit` = ON, (reg. curve on `res.zones`).

residents settle as opposed to the case where there is no social housing. The relation with high income (7b) also improves, with the two curves outlining similar patterns.



(a) Low vs Mid income settlers. (b) Low vs High income settlers. (c) Mid vs High income settlers.

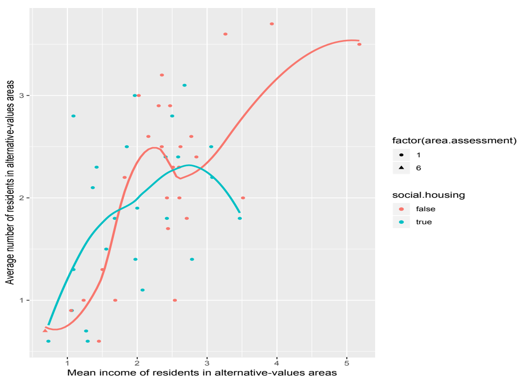
Figure 7: Proportion of settled residents by income comparisons, with `hetero-tolerance` = OFF and `enter/exit` = OFF, (reg. curve on `social housing`).

3.1.3 Exploring parallel society formation

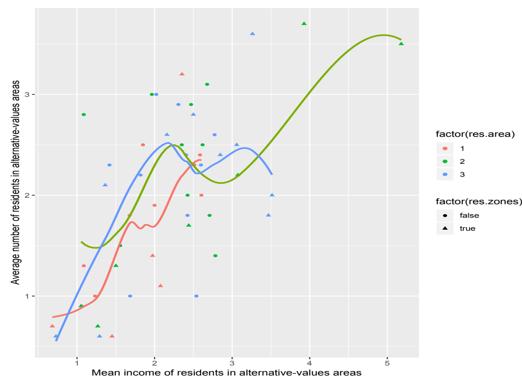
In the following, we do not repeat figures that are already is in the paper. Instead, we present a few more plots that may help clarify the relation between residents in parallel societies and parameters of the PARSO$_{demo}$ simulation.

We have produced a limited number of plots to assess the various conditions that support the emergence of alternative-values areas — called "parallel" societies also in the paper. We have used the criteria of relevance in the selection of the plots to visualise in this document. This does not necessary mean that the plots below are the best possible, just those that, overall, produce a few points of interest. The other plots that we have left out may be considered marginal.

The three figures below — from Figure 8 to Figure 10 — share the same choice of parameters. They present data on the average number of residents in alternative-values areas visualised by their average income, and classified in relation to the `assessment area` and `social housing` for all the (a) figures and in relation to `residential zones` and `residential areas` for all figures labelled with the letter (b).
The parameter `assessment area` is clearly more effective for residents to settle in an alternative values area when its range is very low and equal to 1. This findings is consistent for all the plots, from Figure 8a, 9a, and 10a. On the contrary, `social housing` does not seem to affect the presence of alternative values areas. There are a few variations in the regression curves, but they seem minimal and tend to disappear as `hetero-tolerance` and/or `enter/exit` are set to ON.



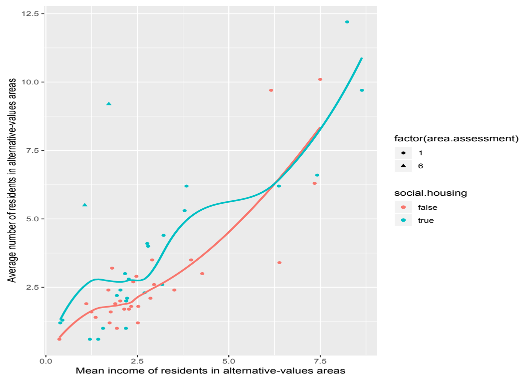
(a) Assessment area and social housing.



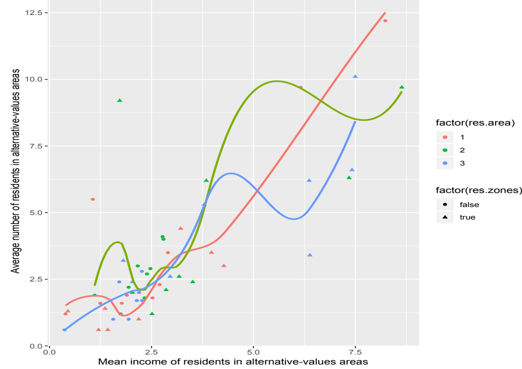
(b) Res. areas and zones.

Figure 8: Average number of residents in alternative-values areas by their average income, with `hetero-tolerance` = OFF and `enter/exit` = OFF, (LOESS curves on `social housing` and `residential areas`).

Results related to the variation of parameter `residential area` and `residential zones` are almost indistinguishable, especially as parameters `hetero-tolerance` and/or `enter/exit` are set to ON (Figure 8b, 9b, and 10b). In particular, when `enter/exit` is set to ON, the emergence of alternative values areas is particularly concentrated around 2.5 residents and low income, i.e. below 2.5 income.

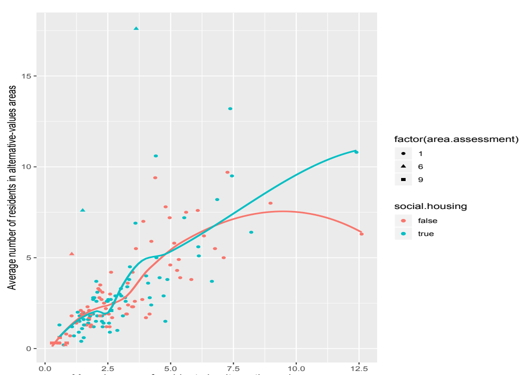


(a) Assessment area and social housing.

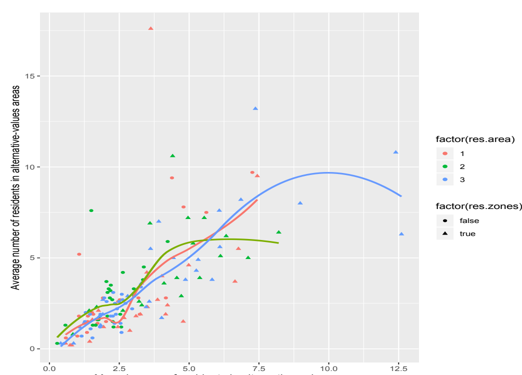


(b) Res. areas and zones.

Figure 9: Average number of residents in alternative-values areas by their average income, with `hetero-tolerance` = OFF and `enter/exit` = ON, (LOESS curves on `social housing` and `residential areas`).



(a) Assessment area and social housing.



(b) Res. areas and zones.

Figure 10: Average number of residents in alternative-values areas by their average income, with `hetero-tolerance` = ON and `enter/exit` = ON, (LOESS curves on `social housing` and `residential areas`).

REFERENCES

Broeke, G. t., Voorn, G. v., and Ligtenberg, A. (2016). Which sensitivity analysis method should i use for my agent-based model? *Journal of Artificial Societies and Social Simulation*, 19(1):5.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Secchi, D. and Seri, R. (2017). Controlling for ‘false negatives’ in agent-based models: A review of power analysis in organizational research. *Computational and Mathematical Organization Theory*, 23(1):94–121.

Seri, R. and Secchi, D. (2017). How many times should one run a computational simulation? In Edmonds, B. and Meyer, R., editors, *Simulating Social Complexity. A Handbook*, pages 229–251. Heidelberg: Springer, 2nd edition.

Seri, R. and Secchi, D. (2018). A power primer for agent-based simulation models. Determining the number of runs in linear and polynomial regression. In *European Academy of Management Annual Conference*. Reykjavik, Island.