## Appendix

A common approach for measuring the effect of geographic context is to analyze how individual level attitudes and behavior varies between residential areas that have been classified on the basis of their population composition. The classification can be done in different ways but multivariate cluster analysis tends to be the preferred tool. What is not always acknowledged is that variables that have been computed as aggregates for fixed geographical sub-divisions such as census tracts or output areas can be strongly influenced by the way the boundaries of these areas have been drawn (Openshaw, 1984). Moreover, using aggregates for fixed geographical sub-divisions to classify residential areas implies an assumption that only conditions within the sub-division matters for individual level outcomes. This is a very restrictive assumption for how geographical context can play a role, and it has recently been demonstrated that it can lead to seriously biased estimates.

Using aggregates for individualized scalable neighborhoods to measure context has been proposed as method for circumventing this problem. Such individualized neighborhoods can be constructed by expanding a circular buffer around different residential locations until the population encircled by the buffer corresponds to a selected population threshold. When this threshold is reached, one can compute aggregate statistics on selected socio-economic variables for the encircled population. By varying the population threshold, contextual measures computed in this way can be designed to focus only the closest neighbors or on larger number of neighbors. In the present study, we allow for multiple scales of ecological influence by varying k, the number of nearest neighbors that are included in the computation of population, from 12 to 12 800 in successive doublings of the population thresholds. The computation was carried out using a new software, Equipop, developed by John Östh in order to address the modifiable areal unit problem (MAUP) in segregation measurement (Östh, Malmberg, & Andersson, 2011).

Equipop requires that the input data are geo-coded to a high level of detail. In this case data from the Population Labor Market Chorology database (PLACE) of Uppsala University has been used. PLACE contains register-based, individual level information for the population in Sweden from 1990 to 2010 with geocodes of the residential location in 100 meter squares. 8 different socio-demographic indicators were extracted to use as input for EquiPop. On the individual level these indicators are: (1) Being unemployed, (2) Having a tertiary education (3) Being a single mother (4) Belonging to the top-ten percent of the income earners (5) Arrived in Sweden during the last five years (6) Having received social allowance during the year (7) Without employment during the entire year, and (8) Country of birth in Asia, Africa, or Latin America. Before being imputed these data were aggregated to 100 meter squares based on the geo-coordinates.

8 socio-demographic indicators and 11 different scale levels (k = 12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, and 12800) result in 88 different measures of neighborhood context that can be used to classify residential areas using cluster analysis (the same method and procedure but with 7 indicators and 12 scale levels was used in (Andersson & Malmberg, 2015; Malmberg, Andersson, & Bergsten, 2014). However, since many of these variables will be highly correlated we have subjected the contextual indicators to a factor analysis that compresses the 88 original indicators to 15 orthogonal factors before proceeding with the cluster analysis. The factor analysis was based on correlations and the number of factors was selected based on them having eigenvalues higher than one. The factors were rotated using the varimax method.

 illustrates the results of the factor analysis. The panels in this figure show the loading of the different factors for each indicator The first factor in Figure A1 represents an *elite context* with high values for tertiary education and top ten percent income. The second factor represents *top-ten near* because of high values for top ten percent income for lower k-levels (closer neighbors). The fourth factor indicates a high share of *newly arrived immigrants* and non-European migrants (VM in figure). The fifth factor signals *marginal groups in adjacent areas* with high values for many marginal groups but only for relatively large k-values (distant neighbors). The sixth factor highlights a high share of *non-European migrants* but not newly arrived. The tenth factor represent *white* *non-elite context* with high levels of unemployment and low values for top ten percent income. The seventh factor has high loadings for *social allowance*. The ninth factor has high loadings for *single mothers*. The eleventh factor has high loadings for *non-employed*. This can indicate either low employment shares in the working age population or a high share retirees (65+). The interpretation of the remaining factors can be done from the graphs.

As a result of the factor analysis every populated 100 meter square in Sweden can be assigned factor scores for each of the 15 contextual factors. Together these factor scores can be seen as determining the position of different geographical location in a socio-demographic contextual space. Geographic context in this representation does not only consist of the demographic composition of a residential area. Instead, context is a composite feature that includes characteristics of both the closest environment and of more extended local contexts.

What we propose in this paper is that these factors can be used to classify residential areas into categories that capture important dimensions in the spatial variation of geographical context. This is achieved by computing residential-area averages for the 15 factors and, then, to use cluster analysis of these averages to group residential areas into classes with varying contextual characteristics. Although this can appear to be almost equivalent to a cluster analysis based on characteristics of the residential area population it is not. Consider for example a residential area where no one has a tertiary education but with neighboring areas that have a high percentage of highly educated. The population composition of these neighboring areas will affect the averages of Equipop-based contextual measures but not measures that are based only on the population composition of the residential area in focus. This implies that our proposed approach will not be as strongly influenced by MAUP (Openshaw, 1984) considerations as the traditional approach.



Figure A1. Description of 15 factors and their loadings.

Using Ward's clustering method we initially grouped the SAMS areas into 20 different area types. However, in order to avoid having cluster with comparatively few SAMS-areas, the number of area types was reduced to 10 by merging smaller cluster into larger ones. The principle used was that area types identified as being too small were merged with the area types having the most similar mean factor scores (minimum Euclidian distance between cluster centroids).

In Figure A2, the average value for the 8 different socio-demographic variables and for different k-values are shown for each of the different area types. Showing the values for the original socio-demographic variables makes the interpretation of the area types easier. The names given to the areas are based on the profile of these values and on the municipality type where the area type is found, see Figure A2.



Figure A2. Average values of socio-demographic variables for the different area types, by k-value.

Table A1 shows the likelihood ratio Chi-Square values for the explanatory variables for each response to the question *‘In a dwelling, what is most important to you?* (Q32). The Chi-Square values indicate the strength of the effect of the explanatory variables, and can be used to calculate significance levels. In Table A1, significance levels below 0.01% are marked in bold, and levels below 1% are shown in italics.

Using Table A1 it is possible to get an overview of the extent to which housing preferences are linked to age, sex, education, income and geographical location. As can be seen in this table, age has a significant effect for most of the response alternatives. Other variables also have effects but these effects are not as pervasive. The idea that age is a major structuring factor for housing preferences is, thus, strongly supported by these results.

Another important finding from Table A1 is that housing preferences among the elderly are only weakly associated with income and education. Thus socio-economic status has effects on the probability of selecting some but relatively few of the alternatives. This is in contrast to an assumption often made in the debate in media that there are large differences in housing preferences between social groups that need to be accommodated, for example by increased options for individual choice in future housing provision for the elderly.

Table A1 confirms that gender is a structuring factor for how dwellings are evaluated but again, the effect is not as pervasive as the effect of age. Gender, thus, is an important aspect to consider in the planning of housing for the elderly.

In addition, Table A1 shows that there, as expected, are clear links between geographic location and housing preferences. This is true both of Municipality type and Area type, with Area type being significantly linked to more response alternatives than Municipality types.

One more observation that can be made is that age seems to be most clearly linked to alternatives that relate to the interior of the dwelling, alternative answers 1 to 9, Table A1. Conversely, the significant response results for Area type concern the location and environment of the dwelling (Q32, alternative 10-15, 16-21).

Table A1. Likelihood ratio chi-square for effects of age, sex, education income, municipality type, and area type on housing preferences. Significance <0.01% in bold and < 1% italics.



# References

Andersson, E. K., & Malmberg, B. (2015). Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies, 52*(12), 2117-2133.

Malmberg, B., Andersson, E. K., & Bergsten, Z. (2014). Composite Geographical Context and School Choice Attitudes in Sweden: A Study Based on Individually Defined, Scalable Neighborhoods. *Annals of the Association of American Geographers, 104*(4), 869-888.

Openshaw, S. (1984). The modifiable areal unit problem, CATMOG (Concepts and Techniques in Modern Geography). *Geo Abstracts*, 40.

Östh, J., Malmberg, B., & Andersson, E. (2011). *Introducing Equipop.* . Paper presented at the 6th international conference on population geographies.