SP 2
27 April 2010

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(28-30 April 2010, Lisbon, Portugal)

Item 10 – Small population and sub-national population projections

Using national data to obtain small area estimators for population projections on sub-national level

Michael Franzén and Therese Karlsson, Statistics Sweden

ABSTRACT

When forecasting population on small geographical areas it is necessary to have stable and reliable data. In Sweden the population density is small in many areas and the data would not be reliable for forecasting purpose. Furthermore, even if the density is high in some areas the number of inhabitants is too low. Due to few observations, distributions of the demographic components may not be obtained. This is often the case for smaller city districts. Even forecasts on the municipality level may sometimes be unreliable due to few observations, especially when it comes to migration. The data on a national level is however stable and reliable when making population forecasts and the aim of this paper is to suggest how national data may be used on a sub-national level and show how this is used in Sweden.

To obtain these small area estimators we would analyze housing definitions on real-estate level. These estimators should be geographically independent and therefore useable anywhere on sub-national level. The assumption is that people living in for example rentable apartments share the same demographical pattern anywhere in the country.

A ***cluster analysis*** on residential types, using the population register and real estate register in Sweden, is used to identify suitable groups. In general, the idea with cluster analysis is to classify observations with homogenous conditions that separate them from other homogenous groups. The analysis results in a division of a number of residential types with homogenous demographic conditions. The cluster analysis shows that age-structure, fertility and probability of moving varies among different residential types and areas. In a demographical sense, people living in an area with larger one-family houses would "behave" differently from people living in an area with mostly small rentable apartments. The division is presented in terms of housing, tenure, year of construction, size of the city where the building is situated and the attraction of the area. These obtained groups contain sufficient demographic data to be used when forecasting future development in small areas consisting of one or more residential types.

SUMMARY

To obtain reliable data for population projections on small areas the idea is to collect data from the whole country that is applicable on these small areas. It has been shown that people in different residential areas behave differently, in a demographic context. For example migration in multi-family dwellings is more frequent than in single-family dwellings. There are also differences between multi-family dwellings that consist of owned or rented apartments and where the dwelling is situated; in large or small cities. The aim in this project is to describe a division of housing that behave similar and for which reliable and stable data can be obtained using data on the national level. The classification presented in this paper resulted in 33 different 'type codes' all with specific out-migration probabilities, age-distribution of the in-migrants, fertility etc.

1. INTRODUCTION

Sweden is divided into 290 municipalities which are responsible for facilities and services such as schools, elderly and child care, housing, public assistance etc. The population in the municipalities vary from 2.500 in the smallest municipality Bjurholm to 830.000 in Stockholm[1]. All municipalities are divided into subareas set by the municipality council. The number of subareas varies along with the size of the population. For these areas Statistics Sweden provides the municipalities with population statistics that can be used in forecasting the population. Often the statistics on such a low level contains very few observations and the data is not reliable when projecting the population. That is why it is necessary to provide the municipalities with reliable and stable data that can be used on small areas. The idea is to determine how different housing behave, using observations for the whole country, and then apply it to the small areas. The analyze results in a division of housing that behave similar and can be used when projecting the population on small areas. Ten years ago the first division was presented and now Statistics Sweden has started a project that will result in a new and updated division of housing.

2. BACKGROUND

The first type code classification of real estates in Sweden was made in 1991. Then the classification was made using only a sample of real estates in Sweden. In 1998 it was decided that a classification using all real estates was necessary and that the results from the study should be used in yearly production of statistics for population projections on small areas.

In the classification made in the early 90's data from 1986-1988 were used. Data on real estates was manually collected from a sample of 30 municipalities and the collection process was time demanding. In 1998 a full coverage register were used for collecting data on real estates. Data for the years 1992-1994 were used when classifying the real estates for the second time.

In both rounds of collecting data for classification of real estates, it was found that the out-migration probabilities do not depend on the year in which the property was built but on the legal form of the house or apartments. In 1998 it was found that the out-migration probabilities in rented homes had increased with 20 percent. In some ages the increase was as much as 50 percent. However, nor an increase or decrease in out-migration probabilities can be observed for owned apartments or small houses (single-family dwellings). It has been discussed what had happened between 1986-1988 and 1992-1994 that made the out-migration probabilities increase. One explanation was that there were large migration flows within the municipalities. This was due to raised rents in the rented homes. During these years people moved in order to find the apartment with the right rent. An increase in movement could not be seen in owned homes, both apartments and small houses, this was probably due to the fact that capital is often tied up when owning an apartment or house. Also, the price on these homes was in larger extent better adjusted to the market prices.

It is these demographical changes one wants to include when producing statistics on properties. The statistics must be easy to update so that one always have the latest demographical data. Since the results from the project in 1998 are completely register based it is now easy to update the tables with new statistics each year. These tables are provided to the municipalities. The tables are sold to approximately 50 municipality councils each year.

It has now been ten years since the last classification of properties was done. It is now necessary for Statistics Sweden to renew the classification. The project is not yet finished and therefore the results in this paper are those from the previous classification.

3. RESULTS

3.1. Data sources

In this paper 'property' and 'real estate' are used synonymously. Property and real estate can include more than one actual house. For small houses the property and real estate most often equals only one house. For multi-family houses the property or real estate may include more than one house. In Sweden each person is obliged to report to the tax board when moving – even if the person moves to another apartment within the same property. Therefore in the population register in Sweden there are data on migration even within the properties. In the classification migration within properties is included.

The data is retrieved from both the population register and the real estate register. In the previous project population data for 1992-1994 were used in the clustering process. Population and population flows by age and sex are available from the

[1] Stockholm in this context is the municipality, however Stockholm area consists of several municipalities and has a population of 2 million.

population register. From the real estate register the following variables are retrieved: *Household type* (defined as either single-family dwelling or multi-family dwelling), *Tenure type* (tenant rented, tenant owned or small houses), *Year of construction, Size of the urban area* (population in the area) and *Level factor* (describes the attractiveness of the area where the real estate is situated).

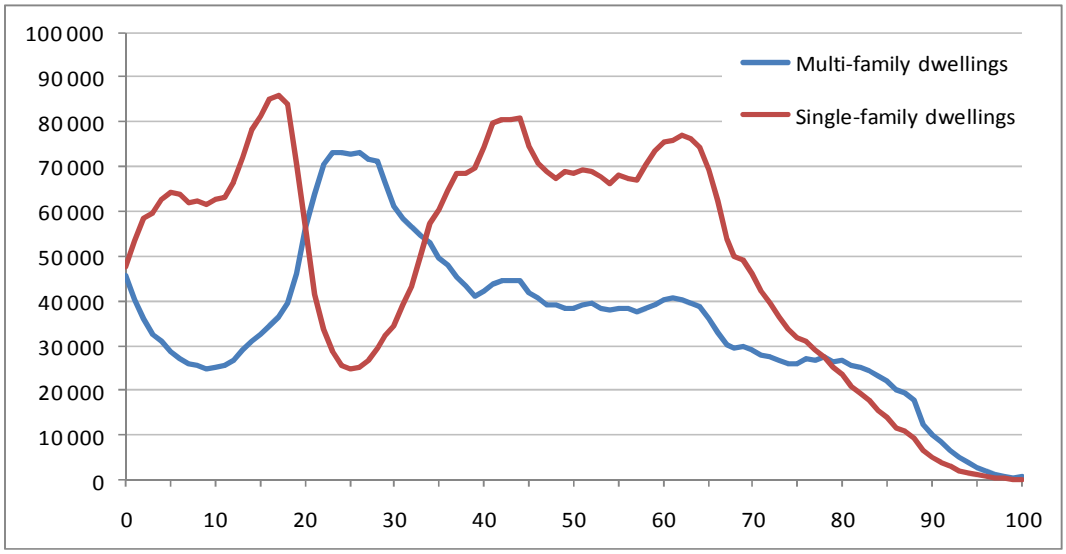
In the new project demographic variables for 2006-2008 in housing units older than 3 years were used (1.8 million properties). This is because the data quality for newly built housing units and real-estates tend to be uncertain. For example, a housing unit may not be reported to the register before the whole building project is completed and the real estate may then be populated without any registered housing (people moving in to the apartments before the building is reported as completely built).

3.2. Demographical results

In this section we use the latest available data for the type codes to show their differences in demography. Data for 2006-2008 has been used.

The age-structure in multi-family dwellings differs significantly from the age-structure in single-family dwellings as shown in figure 1. In the single-family dwellings the majority of the population are children or adults in the ages 40-65 years. In Sweden in general young adults leave the parental home at the age 21 for women and 22 for men[2] and this pattern can be seen in the age-structure of the population in the single-family dwellings. When the young adults leave home they, in most cases, move to a multi-family dwelling, which also can be seen in the age distribution of the multi-family dwellings.

Figure 1. Population by age in multi-family dwellings and single-family dwellings 2008



The age-distribution among those moving to and from multi-family dwellings and single-family dwellings differs as shown in figure 2 and 3.

However we see a similar pattern among the people in their early twenties dominating the out-migration from both multi-family dwellings and single-family dwellings. These movements are often strongly correlated to leaving the parental home. There are, not surprisingly, more people in their twenties and thirties moving in to multi-family dwellings than into single-family dwellings. The single-family dwellings have a negative net-migration in the ages 17-27 years while the multi-family dwellings have a positive net-migration in these ages. The multi-family dwellings have a negative net migration of young children and people in the ages 28-45. This pattern comes from people starting a family in a multi-family dwelling and then moving to a single-family dwelling.

[2] Statistics Sweden, Demographic reports2008:5, *Leaving home*

Figure 2. Age-distribution of migrants to and from multi-family dwellings 2006-2008

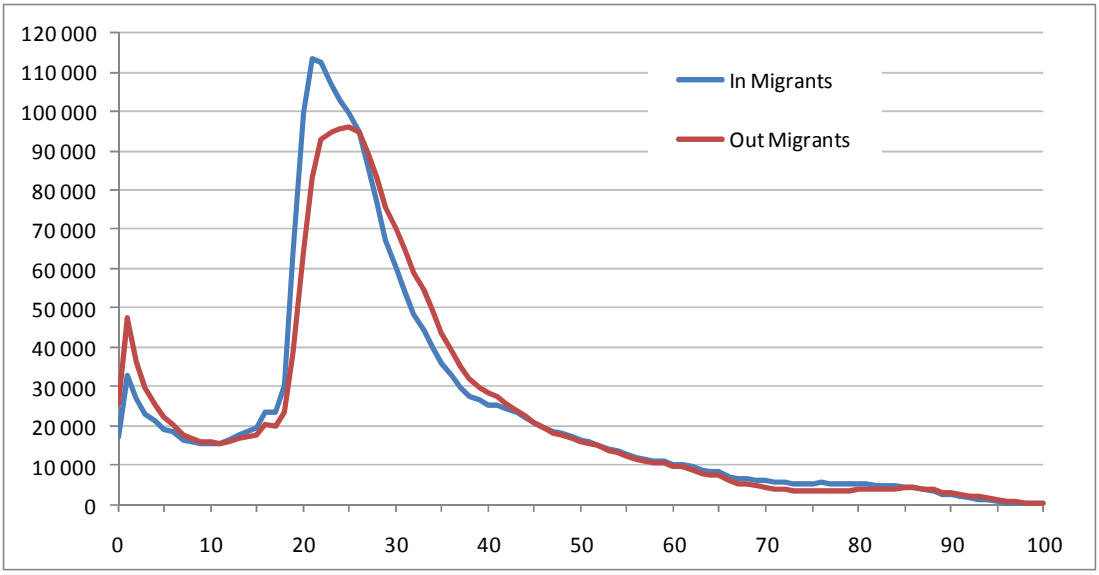
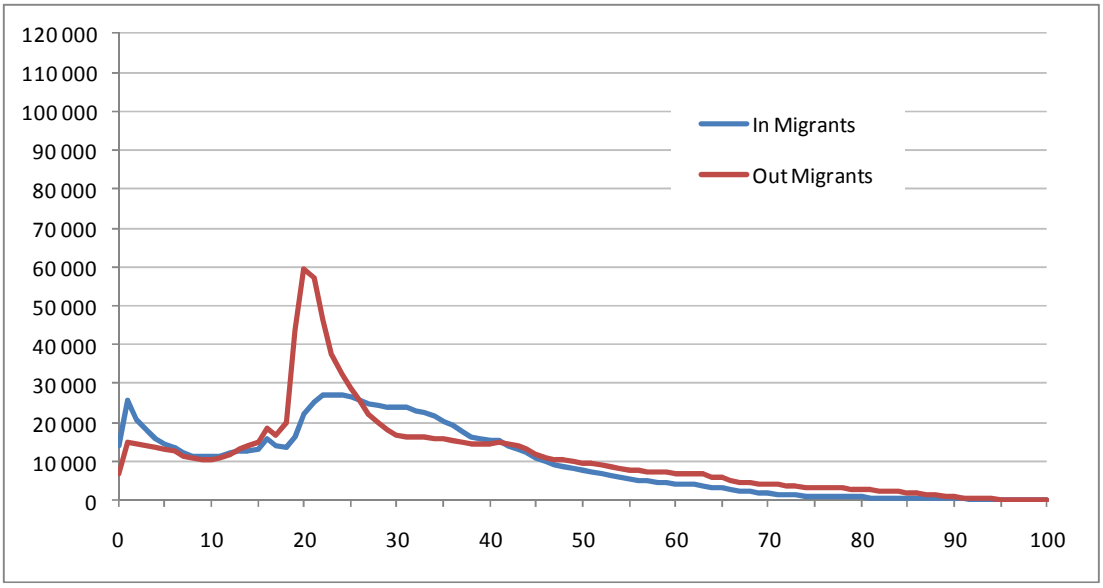


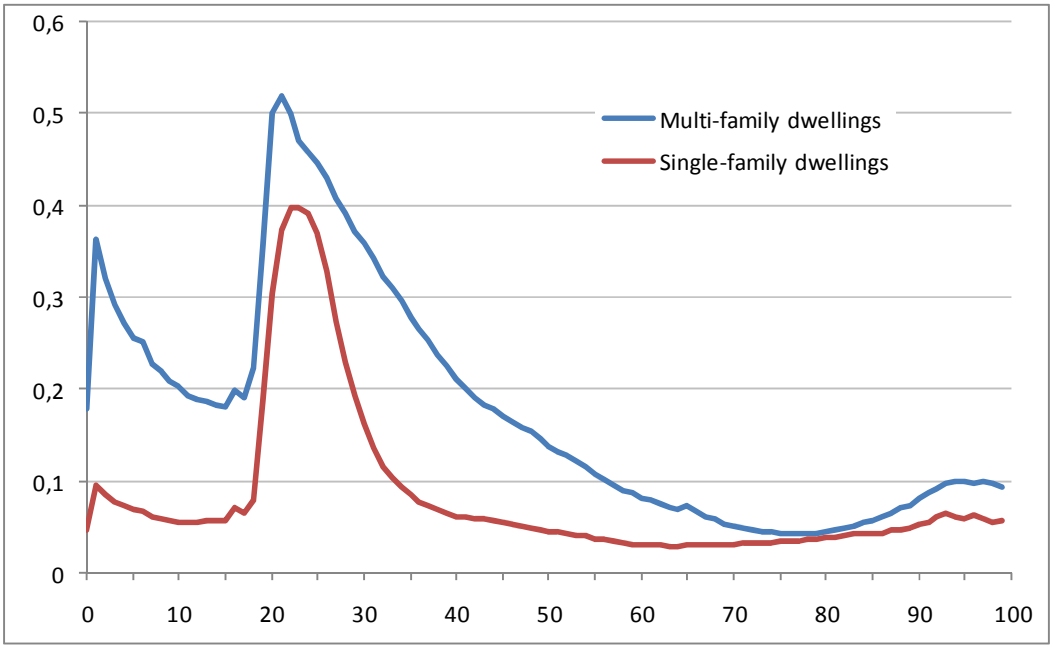
Figure 3. Age-distribution of migrants to and from single-family dwellings 2006-2008



The probability of moving out of a property differs among different types of properties and different ages. This can be seen in the figure below where the probability of out-migration for both multi and single-family dwellings are described. The probability of moving out is age and sex specific, expressed in percentage terms. The probability is calculated using the number of migrants of a certain age and sex proportional to the population in the same age and sex cohort.

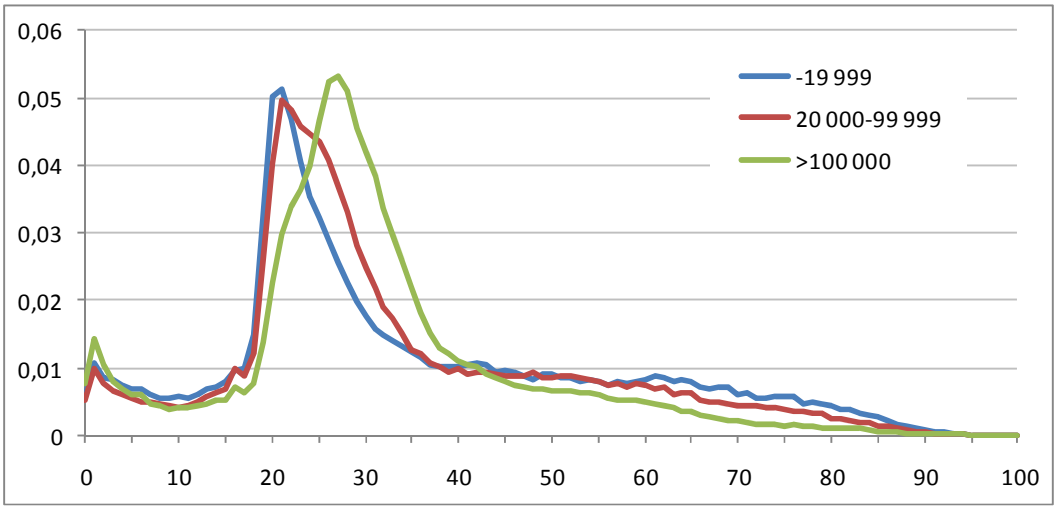
The out-migration probabilities are higher in multi-family dwellings for all ages. For the two types of properties the out-migration probability is at its highest in the ages around 20-30 years. There is also a high risk for young children to move out from multi-family dwellings. Probably due to the fact that their parents, when having children often move from a small multi-family dwelling to a larger one – or from a multi-family dwelling to a small house. The difference in out-migration probabilities between the two types of housing is at its highest in the age of 1-2 years. For both multi-family dwellings and single-family dwellings there is a increase in out-migration probabilities in the highest ages and this is probably due to elderly people moving to an elderly home and leaving their own apartment or house. The out-migration probability for both multi-family dwellings and single-family dwellings are lowest in the ages 70-80 years.

Figure 4. Probability of moving out. Multi-family dwellings and single-family dwellings, 2006-2008



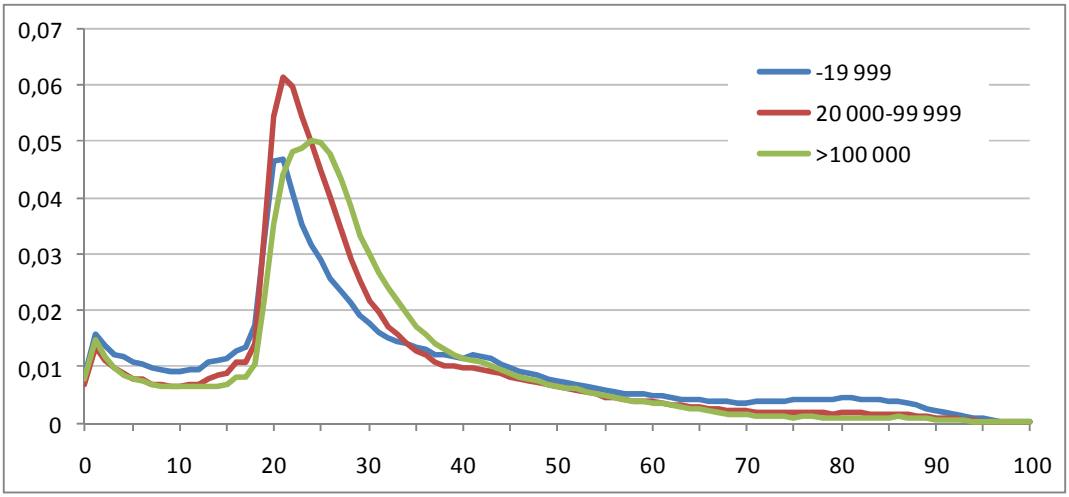
In the figures 5 and 6 below the age-distribution of the in-migrants are shown by the size of the city. In the owned homes there is a significant difference between the three age-distributions. In the smallest cities the in-migration is dominated by young adults in their early 20's. Also in the medium large cities the 21-year olds dominate the in-migration. The in-migration to owned homes in the largest cities with 100.000 inhabitants or more is dominated by people in their middle 20's with a peek for the 26-year olds. These differences may be explained by the fact that the apartments in larger cities often are more expensive than in the smaller cities and therefore the people moving in are somewhat older.

Figure 5. Age-distribution of in-migrants to owned multi-family dwellings by city size 2006-2008



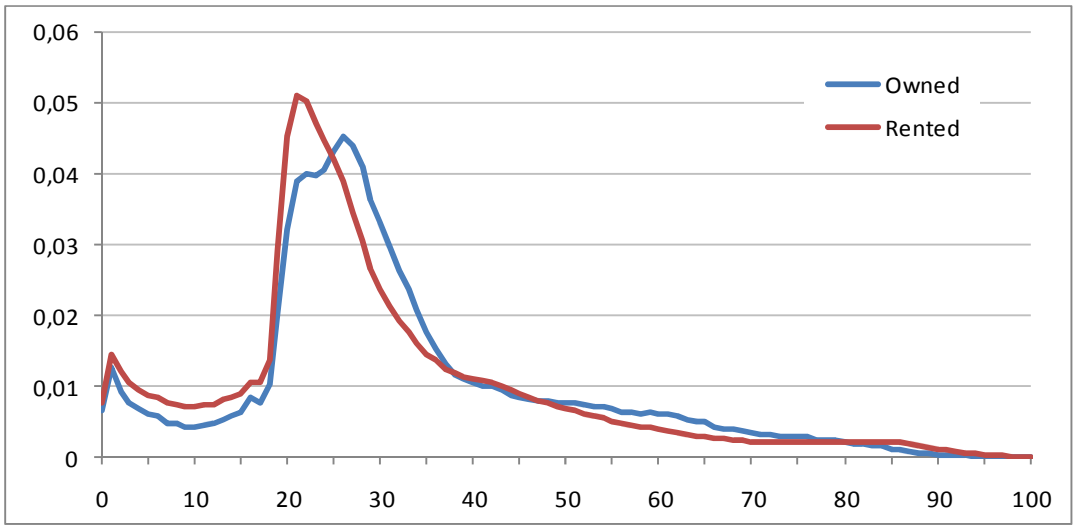
In the rented homes we see a similar pattern, however the age-distribution for the larger cities is somewhat shift to younger ages with a peek for 24-year olds. An explanation to why the people moving in to dwellings in larger cities are older may be that it is more difficult getting an apartment in larger cities and because of that young adults leave the parental home later.

Figure 6. Age-distribution of in-migrants to rented multi-family dwellings by city size 2006-2008



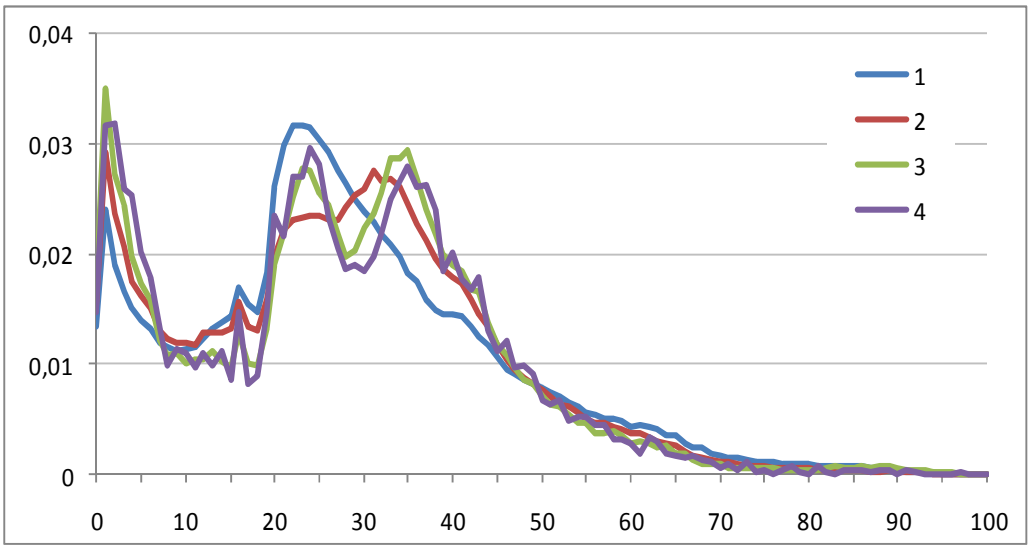
In figure 7 below the age-distributions for in migrants who move to multi-family dwellings is shown. The dominating group of in migrants to rented homes is young adults in the ages 20-25 years. The in migrants to owned multi-family dwellings are somewhat older; the dominating group is 25-28 years. The proportion of children moving to rented homes is higher in multi-family dwellings than for the single-family dwellings.

Figure 7. Age-distribution of in-migrants to multi-family dwellings by legal form 2006-2008



In figure 8 below the age-distribution of the in-migrants to single-family dwellings by the areas level factor is shown. The level factor describes the attractiveness of the area where the dwelling is situated. The attractiveness also describes the price of the house which may be an explanation to why the distributions are more and more shift to the right for the young adults when looking at a higher level factor. The age-distribution of level factor 3 and 4 resembles each other and it may therefore in the future be a good idea to merge these two groups of real estates.

Figure 8. Age-distribution of in-migrants to single-family dwellings by level factor 2006-2008



Given these differences the project resulted in 33 groups of real estates (type codes) for which demographical data can be produced and used when forecasting the population on small areas. A complete list of the 33 different type codes is found in Appendix 1.

3.3 Classification method

The aim was to find a division of housing that is similar within each group considering the following demographical variables:

- Age distribution in in-migration
- Age specific out-migration probabilities
- Housing density
-

The real estate variables used were:

- Household type
- Legal form
- City size (population density)
- Year of construction
- Level factor (attractiveness of the area)

The demographic variables where aggregated by the real estate-variables. These variables are categorical and we need a certain amount of observations to obtain stable distributions of in-migration and the out-migration probabilities**.** The age-groups that were used was 0-5, 6-16, 17-20 21-30, 45-54 and 60-80 years of age. The clustering was carried out with chi2-test on the demographic variables for each category of real estates. If the two categories where similar in the distribution of the demographic variable they were grouped into the same cluster.

This process was carried out for the in-migration, out-migration and housing density separately. The three separate analyses gave different clusters but the final categories that were chosen are shown in figure 1.

The break down in multi-family and single-family dwellings is obvious because many studies have shown demographical differences such as age distribution, probability of moving and housing density between the two types of housing. Other break downs were discussed within the project and the result was as follows.

Figure 9

Multi-family dwellings	
Legal form	
	Tenant owned housing
	Tenant rented housing
City size	
	-19 999
	20 000-99 999
	100 000-
Year of construction	
	1900-1954
	1955-1964
	1964-
Level factor	
	1
	2

Single-family dwelling	
Legal form	
	Small houses
	Tenant owned houses
City size	
	Not densely populated
	Densely populated
Level factor	
	1
	2
	3
	4

In the new project, a complete clustering is carried out without any chi2-tests. Before the clustering it is necessary to determine some large groups of real estates as in the previous classification. This is made as objectively as possible. For example the construction year is set into 10 year groups before clustering.

4. CONCLUSIONS

The results show that people living in real estate with different properties have different demographical behavior. Migration from both multi-family dwellings and single-family dwellings are dominated by people in the ages 20-30. Single-family dwellings have negative net-migration in these ages while multi-family dwellings have a surplus. Multi-family dwellings also have a surplus of people in the ages above 65 years. Multi-family dwellings have negative net-migration of children in the ages 0-10 years whilst single-family dwellings have a surplus.

The age structure in multi-family dwellings and single-family dwellings differs significantly. The population in single-family dwellings is dominated by children and adults in the ages 40-65 years. The population in multi-family dwellings is dominated by young adults in the ages 20-30 years.

Also the out-migration probabilities are different in multi-family dwellings and single-family dwellings. The out-migration probabilities are higher in all ages in multi-family dwellings.

In-migration to multi-family dwellings is dominated by people in the ages 20-30 years and there is relatively few children moving in to multi-family dwellings. The in-migration to single-family dwellings however is dominated by children and people in the ages 20-35 years. The proportion of in-migration in the ages 65 and above is larger in multi-family dwellings. This is probably due to the retirement age which in Sweden is 65 years. Many people living in their single-family dwelling move to a multi-family dwelling when retiring.

There are also differences in the age-structure among the people moving in to multi-family dwellings depending on the properties of the real estate. We have seen that people moving in to owned apartments are somewhat older than the migrants to rented homes. Also the dominating groups of migrants to multi-family dwellings are older in larger cities than in smaller cities which can be explained by the fact that homes are on average more expensive in larger cities.
For single-family dwellings the age-structure for the in migrants are different depending on the attractiveness of the area where the property is situated.

The results show that it is better to use distributions for single-family dwellings than a distribution for all houses when it is known that the small area, for which we would like to project the population, consists only of single-family dwellings.

These differences in demographical behavior between different real estates has resulted in 33 groups; (type codes) that can be used when projecting the population on small areas.

For forecasting purposes it would be optimal to have subareas containing only one type code and then be able to use one set of demographic data. However this is not the case in many municipalities in Sweden. Therefore a common method is to weigh the distribution by population size in each type code to get a distribution for the entire subarea.

APPENDIX 1

Type code	Type of housing	Legal form	City size	Year of construction	Level factor
FBR128	Multi-family dwelling	Tenant owned housing	-19 999	-1954	<7,50
FBR138	Multi-family dwelling	Tenant owned housing	-19 999	1955-1964	<7,50
FBR148	Multi-family dwelling	Tenant owned housing	-19 999	1965-	<7,50
FBR228	Multi-family dwelling	Tenant owned housing	20 000 -99 999	-1954	<7,50
FBR238	Multi-family dwelling	Tenant owned housing	20 000 -99 999	1955-1964	<7,50
FBR248	Multi-family dwelling	Tenant owned housing	20 000 -99 999	1965-	<7,50
FBR328	Multi-family dwelling	Tenant owned housing	100 000-	-1954	<7,50
FBR329	Multi-family dwelling	Tenant owned housing	100 000-	-1954	7,50-
FBR338	Multi-family dwelling	Tenant owned housing	100 000-	1955-1964	<7,50
FBR339	Multi-family dwelling	Tenant owned housing	100 000-	1955-1964	7,50-
FBR348	Multi-family dwelling	Tenant owned housing	100 000-	1965-	<7,50
FBR349	Multi-family dwelling	Tenant owned housing	100 000-	1965-	<7,50
FHR128	Multi-family dwelling	Tenant rented housing	-19 999	-1954	<7,50
FHR138	Multi-family dwelling	Tenant rented housing	-19 999	1955-1964	<7,50
FHR148	Multi-family dwelling	Tenant rented housing	-19 999	1965-	<7,50
FHR228	Multi-family dwelling	Tenant rented housing	20 000 -99 999	-1954	<7,50
FHR238	Multi-family dwelling	Tenant rented housing	20 000 -99 999	1955-1964	<7,50
FHR248	Multi-family dwelling	Tenant rented housing	20 000 -99 999	1965-	<7,50
FHR328	Multi-family dwelling	Tenant rented housing	100 000-	-1954	<7,50
FHR329	Multi-family dwelling	Tenant rented housing	100 000-	-1954	7,50-
FHR338	Multi-family dwelling	Tenant rented housing	100 000-	1955-1964	<7,50
FHR339	Multi-family dwelling	Tenant rented housing	100 000-	1955-1964	7,50-
FHR348	Multi-family dwelling	Tenant rented housing	100 000-	1965-	<7,50
FHR349	Multi-family dwelling	Tenant rented housing	100 000-	1965-	7,50-
SBR000	Single-family dwelling	Tenant owned houses	-	-	-
SER801	Single-family dwelling	Small houses	Not densely populated	-	<5,50
SER802	Single-family dwelling	Small houses	Not densely populated	-	5,75-9,50
SER803	Single-family dwelling	Small houses	Not densely populated	-	10,00-12,00
SER804	Single-family dwelling	Small houses	Not densely populated	-	13,00-
SER901	Single-family dwelling	Small houses	Densely populated	-	<5,50
SER902	Single-family dwelling	Small houses	Densely populated	-	5,75-9,50
SER903	Single-family dwelling	Small houses	Densely populated	-	10,00-12,00
SER904	Single-family dwelling	Small houses	Densely populated	-	13,00-