View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

ERSA 45th Congress of the European Regional Science Association

23-27 August 2005, Vrije Universiteit Amsterdam

Land Use and Water Management in a Sustainable Network Society

PEARL: THE NEW REGIONAL FORECASTING MODEL OF THE NETHERLANDS

A.H. de Jong and T.J. Leering

Netherlands Institute for Spatial Research
Willem Witsenplein 6
2596 BK The Hague
T +31 70 328 87 83
F +31 70 328 87 99
ajong@rpb.nl

QQQ Delft,
Rotterdamseweg 183c
2629 HD Delft
T +31 15 268 25 81
F +31 15 268 25 91
DLE@qqq.nl

Abstract

The Netherlands has a rather long history of developing models in the field of regional forecasts. Among other things, these forecasts are used as an instrument for planning of house-building. In 2004 the Netherlands Institute for Spatial Research and Statistics Netherlands started with the development of a new model, called PEARL (which stands for ‘Projecting population Events At Regional Level’). It is an integrated model for the forecast of the population (by ethnic group) and households.

PEARL will be used to regionalize the official forecasts of population (by ethnic group) and households at the national level, which are compiled by Statistics Netherlands. The lowest level of the regional forecasts will be the municipal level, which permits the aggregation to larger NUTS regions, such as ‘COROP’ and ‘province’. The forecast-horizon of the regional forecasts will up to 2025, although computations for a longer period are possible.

Assumptions on demographic (growth) components (fertility, mortality, internal and external migration) and transitions rates (with respect to the life course) will be formulated at the municipal level. These assumptions are used as input for PEARL. In this way transparency of the outcomes of the model is promoted.

In order to achieve consistency between population and households, PEARL consists of both a macro and a micro layer. At the macro-layer (the municipal level) the assumptions are applied, while in the micro layer (individual level) the resulting events are administrated. In this way the micro layer consists of approximately 16 million persons and approximately 7 million households. In switching between a macro and micro level, PEARL distinguishes itself from more conventional models.
The primary goal is to use PEARL as a (robust) instrument for forecasting. However, it may also be used as a tool for compiling scenarios. This can be done at the macro level (by formulating alternative assumptions at the municipal level), but also at the micro level (by using alternative figures on risks). In the last application PEARL is used as a micro-simulation model.
The software program PEARL is written in Delphi-5. The intention is to publish first outcomes (with a limited scope) in the second half of 2005.
This paper is a report of a still ongoing project. It gives an overview of important decisions taken during the project and present outlines, that will be implemented in the program in the near future.

KEY WORDS: regional forecast, dynamic model, population events, micro-layer, life-cycle

1. Introduction

In 2004 the Netherlands Institute for Spatial Research (NISR) and Statistics Netherlands (SN) started a cooperation, with the intention to produce a regional forecast on population, foreign groups and households. In 2005 the NISR and SN will publish their first regional demographic forecast. Hereafter, the regional forecast will be published in a sequence of two years. The regional forecast will consist of the number of persons by sex, age, foreign group and household position as well as the number of households by type and size, both at the municipal level.

Insight in regional developments in population and households is a necessity for many organizations, and especially for the national government, provinces and municipalities. For example, a regional forecast of the development in the number in households (by type) is essential in order to know how many houses must be build (and where) in order to accommodate the future population. For the planning of schools it is important to know how many young people will arrive at the age that they must follow compulsory education. The NISR itself, has the task to investigate the spatial consequences of trends in population and households. This was the main reason for the NISR to develop a regional forecasting model. This will be done in cooperation with SN, which has ample experience with compiling forecasts (at the national level) and analyzing regional developments. Furthermore, it has detailed regional databases at its disposal.

The new forecasting model is called PEARL, which stands for 'Projecting population Events At Regional Level'. This name points to the key issues of this model: regional, forecast and population events. The population events incorporated in the model are both the 'traditional' demographic events (birth, death, external and internal migration) as life-course events (transitions between household positions). These events lead to changes in (the size and composition of) population and (the number and composition of) households.

In the process of developing a regional demographic forecasting model many crucial decisions have to be taken. The following paragraph will expand on this topic. The next paragraph will explore important issues in the structure of the model and technical aspects of the model. The last paragraph will illustrate how the model will deal with short-distance migration, being an important component of the regional forecasting model,

2. Main choices taken in the development of PEARL

Regional level

One of the main choices concern the regional level of the forecast. A high regional level, such as the NUTS-2 region (in the Netherlands represented by 'provinces'), compromises the applicability of the forecast (because at lower regional levels the need for forecasts still prolong), while a very low regional level (such as postal code area or district) enhances the complexity of the forecast (because for each specified region assumptions must be formulated). A path between these two extremes had to be found. Two regional levels seemed to be good candidates: COROP (NUTS-3) and municipality (NUTS-4). The choice has fallen on the level of municipalities. The lowest level of government also operates at this level and with respect to many planning activities (such as house-building) this is the most appropriate level. Of course, this choice permits by aggregation to produce outcomes at higher regional levels such as ‘COROP’ and provinces.

Characteristics of persons

In the regional forecast the following characteristics of persons will be considered: sex, age, household position and foreign group. The following household positions are distinguished: child (living home), living alone, living together (with and without children), single parent, other, institutional. With respect to foreign group, first a division is made between natives and foreigners. The latter group is split up in six ‘origin-categories’: Turkey, Morocco, Surinam, Netherlands Antilles and Aruba, other non-western and western. All six groups will be broken down by first generation (persons born abroad, having one or two parents born abroad) and second generation (born in the Netherlands, having one or two parents born abroad).
Although from the literature it is known that educational level is also an important determinant with respect to demographic behavior, this characteristic is not included in the forecast. The reason for this exclusion is the scarcity of regional data on educational attainment.

Characteristics of households

In the regional forecast households will be split up by type, size and composition. The following types of households are distinguished: one-person households, couples (with or without children), single-parents and other households. The considered sizes of households are 1, 2, 3, 4 and 5 and over. The composition refers to characteristics of the partners of a couple without children or characteristics of the combination of (one or two) parents with the children living at home (1,2, 3 and 4 or over) and possibly the presence another person belonging to the household. So, by looking at characteristics of the reference person of the household (‘the head’) and the children, it is possible to characterize households by age, foreign group and phase in the family-life-cycle (families in the formation phase, expanding phase, contracting phase and empty-nest)

Relationship between the national and the regional forecast

Consistency with the national forecast of Statistics Netherlands on the population, foreign group and households, is considered as a very strict requirement of the regional forecast. A consequence of this choice is that the national forecast will have to be compiled first and that the regional forecast will be made hereafter. This will be done in a cycle of two years: in the one year the national forecast will be drawn up and in the other year the regional forecast. As the regional forecast is subordinate to the national forecast, an important concern in making the regional forecast consistent with the national forecast will be to keep the specificity of regional developments as much as possible intact.

Type of the model

The focus of PEARL is on demographic processes, so the assumptions of the model refer to developments in population events. This means that in the typology of (household) models Pearl can be classified as a dynamic model. Other dynamic household models in the Netherlands are LIPRO (Van Imhoff and Keilman, 1991), PRIMOS (Gordijn, Heida and Den Otter, 1983, Heida, 2002), WODYN (Hooimeijer en Linde, 1988), MUDEA (Willekens e.a, 1988) and NEDYMAS (Nelissen and Vossen, 1989). Where dynamic models deal with processes of household formation and household dissolution, static models extrapolate proportions of population categories (defined by a certain combination of characteristics). The headship rate model is a well known representative of static models. Although static models are relatively simple and practical (due to a modest need for data), these models are not capable of dealing with the dynamics of population processes, since they model changes in (population or household) structures at subsequent points in time. The national household forecasting model of Statistic Netherlands can be characterised as a mixed dynamic and static model (De Jong and De Beer, 2001). In a cohort-perspective changes in the population by household position are modelled. The model does not provide information on household processes. As PEARL does model these processes, this creates a difficulty in making the regional forecast consistent with the national forecast: this has to be done in terms of the population by household position (and not the household processes).
In PEARL the projection of population is based on the well-known cohort-component model. In this model changes in the size and composition of the population are due to the components birth, death, immigration, emigration, internal migration. The modeling of the population by household position will be done by simultaneously dealing with all the distinguished

household events (in contrast to sequentially dealing with these events, as is the case in the PRIMOS model).

Modelling of assumptions

Besides the dichotomy static/dynamic, (demographic and household) models can also be classified on the dimension to what extent, next to demographic variables, also non-demographic (most often socioeconomic) variables are included. Although PEARL itself is a purely demographic dynamic household model, in the process of formulating assumptions on population events behavioural variables play an important role. Assumptions on projected trends in population events will be based, as much as possible, on outcomes of explanatory models. In so-called pre-processor models regional variation in demographic events will be explained by background variables. For instance, De Beer and Deerenberg (2005) have made an explanatory model on fertility based on demographic, cultural and socio-economic variables. In the regional forecast, municipal trends in the total fertility rate will be modelled based on assumed trends in the significant independent variables of the regression model. So, the input of PEARL consist of figures on trends in demographic events (and not on trends in non-demographic variables). However, these trends are based on assumptions how the relevant background variables will develop in the future and what the effect might be on (specific groups of) municipalities.

Limiting the process of making assumptions

An important bottleneck in making regional forecasts, is the multitude of assumptions that have to be made. Imagine, at the municipal level the model must produce results for each category of the population, defined by a certain combination of sex, age, foreign group and household position. Moreover, for each event incorporated in the model for these combination of categories assumptions with respect to future trends must be formulated. It is obvious that the process of making of assumptions must be tackled with skill, otherwise the production of the regional forecast will not be feasible due to the labor-intensive character of the assumptions making process. A way out of this dilemma is that the assumptions do not necessarily have to have to refer to the lowest regional level, namely the municipalities. It is possible to cluster municipalities into meaningful aggregations and then to make assumptions for these clusters. Another solution is to work with key-indicators and translate these into age-specific figures by applying a standard age distribution. For example, in case of fertility the assumptions on trends in the Total Fertility Rate will be broken down into age specific fertility rates by applying the national age distribution. This procedure will be applied for each specified municipality, but because this will be done for each birth order separately the average age at motherhood (all birth-orders together) may differ between municipalities.

The process of making assumptions gets extremely labour-intense with respect to the most critical event of the regional forecast, namely internal migration. In PEARL the choice is made to formulate assumptions on the internal migration flow between each pair of municipalities. So, the number of assumptions that each year has to be made is enormous. At the moment, the number of municipalities in the Netherlands amounts to about 500, which in the worst case would mean modelling a total number of about 250 thousand internal migration flows. It is clear that with respect to internal migration, it is essential to streamline the assumption making process. The solution is to use a model with a limited number of variables. In Pearl a special type of the so-called spatial interaction model will be used. In paragraph 4 this model will be discussed.

Micro layer: individual persons and households as lowest units of the forecast

PEARL will produce outcomes for categories of the population and households (both distinguished by several background characteristics). The model shares this feature with most household models in the Netherlands (like the national forecasts of Statistics Netherlands, PRIMOS, LIPRO, and MUDEA). For this reason, these models can be considered as macro-simulation models. However, PEARL goes further: it will also produce results at the micro level of individual persons and individual households. In this sense Pearl resembles a micro-simulation model. The micro layer consists of approximately 16 million persons and approximately 7 million households.

Three main reasons can be given, for the inclusion of a micro level in PEARL. The first reason is to solve consistency problems, associated with macro-simulation models. Consistency can be defined as a situation in which endogenous variables must satisfy certain constraints (Keilman, 1985). For instance, if individuals are classified by household position the number of males living together should equal the number of females living together (discarding same sex-couples). The consistency problem is especially problematic when dealing with the relationship between persons and households. Household formation and dissolution processes - such as children leaving the parental house, couples formed by singles going to live together, splitting up of couples (with and without children) - makes it very difficult to keep track of the relationships between persons (by household position) and households (by type and size). In PEARL, the effects of these household processes on the size and the composition of households will be modelled in the micro-layer. The micro-layer contains each person and each household residing in the Netherlands, in terms of 'object oriented programming language' each person and household is treated as an individual object. In order to be able to relate persons to households, each persons is given an identification number and its record also contains the identification numbers of relevant others (such as the partner, the children and the parents) and the household he/she belongs to. As in the regional forecast the assumptions of Pearl refer to the level of municipalities, the function of this micro layer nothing more than to administrate the relationship between population and households. For this reason, in principle no specific assumptions will be made at the micro level. In case a certain number of persons from a specific combination of background variables will have to be submitted to a particular event, this will be achieved by picking this number of person at random from the micro-layer.

The second reason for the inclusion of the micro-layer in PEARL is that this opens the opportunity to run scenario's. It is possible to expose the individual persons in the micro-layer to certain demographic risks. This kind of application of PEARL is similar to that of a full-blown micro-simulation model (like NEDYMASS, Nelissen and Vossen, 1989). However, it is also possible to run scenario's by formulating alternative assumptions for population events at the macro-level.

The third reason is that the micro-layer provides new kinds of information. It is possible to follow individuals (and household) over time during its life-cycle. So, information on all kinds of careers can be produced (for example the career of family formation and dissolution).

Time table and consultancy

For the success of a project, it is essential to keep up the speed and to formulate clear and attainable targets. In this phase of the project, the main target is to compile a first regional forecast in the second half of 2005. Probably it will be done with an 'interim' model, that needs to be developed further. Every month a project team, consisting of employees from the Netherlands Institute for Spatial Research (NISR) and Statistics Netherlands (SN) comes together in order to take important decisions. However, also other institutes are involved in the development trajectory of PEARL. A consultancy group (called 'klankbordgroep') has been established, consisting of specialists coming from an array of institutes familiar with regional forecasts, such as state departments, provinces, municipalities, other planning institutes and scientific organisations. This group has a few meetings each year and advises on the main lines of the project.

3. The regional model PEARL

Cooperation between NISR and SN

The Netherlands Institute for Spatial Research (NISR) and Statistics Netherlands (SN) have agreed in a cooperation with the aim to compile each two year a regional demographic forecast. The results will be disseminated by means of publications, presentations and placing the results on the website of NISR and SN. From the websites users may download the outcomes (free of charge)

Both institutes have agreed on plausibility as the key issue in the development of the regional forecast. This can be translated in two objectives. Firstly, the forecasts must be stable: minor changes in the input may not lead to major changes in the output. Secondly, the forecast must be transparent: the relationship between input and output must be clear. So, it must be possible to explain the trends in the outcomes in terms of trends in the assumptions.

The NISR is also engaged in the development of other models, such as a housing market model and a labour market model. The results of PEARL will constitute an important input for these models. This implies a clear link between PEARL and the other models, not only with respect to the specification of the output but also with respect to the methodology of the models. This is especially of importance when scenario’s are constructed.

Making the regional forecast consistent with the national forecast

The choice to make the regional forecast in two successive steps is largely a consequence of the wish to attain plausible forecasts. At first, Statistics Netherlands compiles at the national level the forecast of the population, foreign groups and households. Then, the regional forecast will be compiled and consistency with the national forecast will be brought about. This means that the regional forecast is hierarchical subordinate to the national forecast. This type of ‘top-down’ approach can be justified by the fact that it is easier to forecast the national population than regional populations. Keilman (1984) mentions that the behaviour of large populations is generally more stable. Moreover, statistical figures on events show minor random fluctuations when larger groups are involved. Finally, at the national level the component of internal migration is irrelevant (after all, for a country the sum of internal migration flows amounts to zero), while for regional populations internal migration is a very important component of population growth.

In PEARL, a first run is done based on region specific assumptions. In a next run, using a method described by Eichperger e.o. (1979), the regional figures will be made consistent with the national numbers. In essence, this method is nothing more than a proportional adjustment of the regional figures on events to the national total. To the events at the municipal level a correction factor will be applied, which is nothing more than the ratio between the national figure and the sum of all the regional figures.

Forecast of the population: cohort-component model

In Pearl the forecast of the population will be based on the cohort-component model. Below this method will be described, for the sake of simplicity internal migration and the subdivision of the population by foreign group will temporarily be ignored. At the municipal level, the input of the model consists of age-specific fertility rates (by birth order), age-specific death risks, age-specific numbers on immigrants and emigrants. For the first three components this specification resembles that of the national forecast. However, in the national population forecast emigration is modelled by using age-specific emigration risks, while in PEARL it will be modelled using age specific number of emigrants. This choice was taken because it resulted in a huge simplification of the assumption making process for this component.

Now, for each municipality the population equation simply is:

$$L_{x+1,t+1} = L_{x,t} - D_{x+1,t} - E_{x+1,t} + I_{x+1,t} \quad (1)$$

where x = 1, 2, …94

$L_{x,t}$ = population aged x on January 1st of calendar year t

$D_{x+1,t}$ = number of deaths aged x +1 at 31 December of calendar year t

$E_{x+1,t}$ = number of emigrants aged x +1 at 31 December of calendar year t

$I_{x+1,t}$ = number of immigrants aged x +1 at 31 December of calendar year t

Again for the sake of simplicity, the subscript t for calendar year will be dropped below. The number of deaths can be expressed as a risk:

$$q_x = \frac{2 \cdot D_{x+1}}{L_x + L_{x+1} + D_{x+1}} \qquad (2)$$

where

q_x = risk of dying at age x at January 1st

Substituting formula 2 into formula 1 leads to the following equation, that can be used in the forecast:

$$L_{x+1} = L_x - E_{x+1} + I_{x+1} - ((L_x - 0.5 * E_{x+1} + 0.5 * I_{x+1}) * q_x) \qquad (3)$$

However, when the death risks would be applied to the population at January 1st (ignoring half of the immigrants and half of the emigrants) this would cause only a small error. At the national level this would lead to an underestimation of about 50 death on a total of 140 thousand, based on the figures of 2004. In order to enhance the speed of the computations in PEARL the simple formula is used (knowing that in the consistency step of the model, the sum of the regional number of deaths will be made equal with the national total according to the national forecast):

$$L_{x+1} = L_x - E_{x+1} + I_{x+1} - L_x * q_x \qquad (4)$$

In a next step, internal migration will to be modelled. Without going into detail, the modelling will be done in the following way. Age-specific risks on moves will be applied to the midyear population in order to compute the number of moving persons in a municipality (in a specific calendar year):

$$M_{x+1} = 0{,}5 qm_{x+1} (L_x + L_{x+1}) \qquad (5)$$

where

M_{x+1} = number of moving people aged x + 1 at 31 December

qm_{x+1} = rate of moving of people aged x +1 at 31 December

(and qm_{x+1} is defined as $\frac{2 \cdot M_{x+1}}{L_x + L_{x+1}}$)

Then, the number of moving people will be split into a group of persons moving within the municipality and a group of persons leaving the municipality. The latter group (the internal migrants) will again be split in two parts: a first part consisting of long-distance movers and a second part consisting of short-distance movers. The reason for this split-up is that the two parts are driven by different migration motives and for this reason must be modelled differently. In general, migrations over long distances are connected with reasons of work and study and have a 'structural' character. Certain destinations have a long-term popularity due to the presence of certain (educational) facilities (such as universities), the social-cultural climate (in general large cities are attractive to young people) or economic potential (large cities generally offer more job opportunities). Giving that these kind of attractions do not change much in time, long distance migration to destination municipalities can be modelled by using a constant distribution function. Short distance migrations are connected with housing / living environment reasons and with changes in the life-cycle (such as marriage and divorce). This kind of migration will be modelled by using a spatial interaction model. Paragraph 4 will elaborate on short distance migration. For both short distance and long distance migration the model will provide destination municipalities. Looking from the perspective of the receiving municipality, this means that the number of internal migrants arriving in a specific municipality is also known. So, the population in a municipality can be computed as follows:

$$L_{x+1,r}{}^{*} = L_{x+1,r} - \sum_{i=1,i<>r}^{n} M^{L}{}_{x+1,r,i} + \sum_{i=1,i<>r}^{n} M^{A}{}_{x+1,i,r} \qquad (6)$$

where

$M^{L}_{x+1,r,\ i}$ = number of migrants leaving municipality r age x + 1 at 31 December with destination municipality i (i=1, 2 ….. n, i <> r)

$M^{A}_{x+1,i,\ r}$ = number of migrants arriving in municipality r aged x+1 at 31 December with departure municipality r (i=1, 2 ….. n, i <> r)

At this point, the discussion on the modelling of internal migration will end. However, paragraph 4 will go into more detail on the modelling of short-distance movers.

In a next step fertility can be modelled. The age specific fertility rate is defined as follows (for the sake of simplicity ignoring birth order):

$$f_{x,t} = \frac{2 \cdot B_{x+1}}{F_x + F_{x+1}} \qquad (7)$$

where

f_x = fertility rate of women aged x at January 1st,

$B_{x,t}$ =births from women aged x +1 at 31 December ,

$F_{x,t}$ = female population aged x at January 1st,

Then, in the model the number of live born children by age of the mother can be computed as follows :

$$B_{x.} = \frac{1}{2} f_x (F_x + F_{x+1}) \tag{8}$$

Next, the number of live born children by age of the mother can be aggregated over all fertile ages:

$$B = \sum_{x=15}^{49} B_x \tag{9}$$

At this point, the population aged 0 at 31 December can then be computed as:

$$L_0 = B - B * q_x \tag{10}$$

Note, that in this formula death of immigrants and emigrants aged 0 are ignored.

For the population of the highest age (95+) the death rate is computed in a slightly different way. The death risk is defined as:

$$q_{95+,t} = \frac{2 \cdot D_{95+,t}}{L_{94,t} + L_{95+,t} + L_{95+,t+1} + D_{95+,t}} \tag{11}$$

This death risk is then applied to the population of age 94 and higher at January 1st:

$$D_{95+,t.} = q_{95+,t} (L_{94,t} + L_{95+,t}) \tag{12}$$

The population aged 94 and higher at 31 December can then be computed as:

$$L_{95+,t+1.} = L_{94,t} + L_{95+,t} - D_{95+,t} \tag{13}$$

In the formulas described above, the distinction of the population by foreign group is ignored. However, a forecast of foreign groups is also part of PEARL. This means that computations, according to formulas 1 to 13, must also be done for foreign groups. However, some adjustments must be made to the formulas. For example, fertility rates must be applied to foreigners of the first generation, leading to children of the second generation. Also, children stemming from mixed partners must be accounted for. Children with a first generation father and a native mother are considered to belong to the second generation foreigners. Further, in order to keep the assumptions making process simple, adjusted formulas will only be applied to the specified six foreign group (but not to the natives). This means that figures on natives will be computed as the difference between the total group and the sum of the specified foreign groups.

Forecast of households: simultaneous modelling of transition rates

After the forecast of the population, PEARL will take up the forecast of the population by household position in a next step. Based on the forecast of the population by household position it is possible to derive the number of households. However, the actual computation of household will be done in the micro-layer of the model.

In the forecast of the population by household position a dynamic model will be used. In this step of the model, the following kind of transitions will be modelled (note that several possible transitions are discarded):

- process of leaving home, with destinations living alone, living together and other
- process of returning to the parental home, coming from the positions living alone, living together and other
- process of living together, coming from the position living alone, single parent (and child)
- process of splitting up of couples, with destinations living alone, (child) , single parent (in case children are present in the household) and institutional
- process of becoming single parent due to having children without having a partner
- process of going to an institution, coming from the position living alone or living together

At the municipal level, changes in the population by household position do not only stem from the transitions described above, but also from the demographic events birth, death, immigration, emigration, departure and arrival due to internal migration. This means that for these demographic events the consequences for the population by household position must be determined. This will be done as follows. All births have, of course, the household position child. Deaths will be drawn from the micro-layer and therefore their household position is already known. However, as death risks may vary according to household positions, it might be necessary to give certain household position a higher chance to be drawn. Immigrants will assigned a household position (before they are incorporated in the model). Emigrants will be drawn at random from the micro-layer and therefore again the household position is already known (in case the chosen emigrant belongs to a more-person household all household members will also emigrate). Also persons migrating from other municipalities, are drawn at random from the micro-layer (and again all members of the households will migrate as well). When they arrive in the receiving municipality the household positions of the arriving persons is still known.

Working with all the transitions mentioned above, may seem to lead to a very complicated assumption making process. However, this process can be curtailed by working with fixed changes for transitions that are relatively rare.
One of the advantages of working with a micro layer, is that the process of children leaving the parental home is relatively easy to model. Each child living at home is exposed to age-specific risk of leaving home. When the last child has left, the label of the household changes from couple with children to couple without children (or in case of a single parent, it gets the label one-person household). This implies that the transition between position single parent and living alone has not to be modelled explicitly. Likewise, the micro layer eases to job in case of death. In case of a couple, death of one partner implies the assignment of household position living alone to the remaining partner.

In order to be able to model transitions between household positions simultaneously, each household position will receive its own specification. For example, the equation for children living at home is as follows (this and the following equations must be written both for males and females, however this distinction is dropped for sake of simplicity):

$$C_{x+1} = C_x - D_{x+1}{}^C - CLA_{x+1} - TLT_{x+1} - CO_{x+1} + LAC_{x+1} + LTC_{x+1} + OC_{x+1} + I_{x+1}{}^C - E_{x+1}{}^C + MA_{x+1}{}^C - ML_{x+1}{}^C \quad (14)$$

waarbij
C_x = children living at home aged x at January 1st
$D_{x+1}{}^C$ = deceased children aged x+1 at 31 December
CLA_{x+1} = children living alone aged x+1 at 31 December
CLT_{x+1} = children living together aged x+1 at 31 December
CO_{x+1} = children in position other aged x+1 at 31 December
LAC_{x+1} = persons living alone in position child aged x+1 at 31 December
LTC_{x+1} = persons living together in position child aged x+1 at 31 December
OC_{x+1} = other persons in position child aged x+1 at 31 December
$E_{x+1}{}^C$ = emigrating children aged x+1 at 31 December
$I_{x+1}{}^C$ = immigrating children aged x+1 at 31 December

$MA_{x+1}{}^{C}$ = arriving children (from another municipality) aged x+1 at 31 December
$ML_{x+1}{}^{C}$ = leaving children (departure to another municipality) aged x+1 at 31 December

The transitions (persons moving from household position a at age x to household position ß at age x + 1) can be expressed in terms of transition-risks. For instance, the risk for children of leaving home with destination living alone is as follows:

$$u_x{}^{CLA} = \frac{2 \cdot CLA_{x+1}}{C_x + C_{x+1} + CLA_{x+1}} \qquad (15)$$

where
u_x^{CLA} = risk of children living at home at age x to become living alone (at age x + 1)
This risk can also be written as:

$$CLA_{x+1} = \frac{u_x{}^{CLA}}{2 - u_x{}^{CLA}} (C_x + C_{x+1})$$
$$u_x{}^{CLA} * = \frac{u_x{}^{CLA}}{2 - u_x{}^{CLA}} \qquad (16)$$

Formulas 15 and 16 can also be formulated for other transitions. Substituting such formulas into expressions like 14 and working these expressions out, leads to equations as follows:

$$(1 + u_x{}^{CLA} * + u_x{}^{CLT} * + u_x{}^{CO} *) C_{x+1} - r_x{}^{LAC} * LA_{x+1} - r_x{}^{LTC} * LT_{x+1} - r_x{}^{OC} * O_{x+1} =$$
$$(1 - u_x{}^{CLA} * - u_x{}^{CLT} * - u_x{}^{CO} *) C_x + r_x{}^{LAC} * LA_x + r_x{}^{LTC} * LT_x + r_x{}^{OC} * O_x \qquad (17)$$
$$- D_{x+1}{}^{C} + I_{x+1}{}^{C} - E_{x+1}{}^{C} + MA_{x+1}{}^{C} - ML_{x+1}{}^{C}$$

where
u_x^{CLA} = risk of children aged x at Januari 1st to live alone (at age x + 1)
u_x^{TS} = risk of children aged x at January 1st to live together (at age x+1)
u_x^{TO} = risk of children aged x at January 1st to be in position other (at age x+1)
r_x^{LAC} = risk of persons living alone age x at January 1st to live at home (at age x+1)
r^{LTC} = risk of persons living together aged x at Januari 1st to live at home (at age x+1)
r_x^{OC} = risk of other persons aged x at January 1st to live at home (at age x+1)
LA_{x+1}= persons living alone at age x + 1
LT_{x+1}= persons living together at age x + 1
O_{x+1}= persons in position other at age x + 1

The system with those kind of equations for each household position can be arranged more conveniently by using matrix notation (below only for children the transition rates, arrivals and leaves are filled in, for the other positions dots are substituted):

$$\begin{bmatrix} 1+MC_x^{\;C} & -r_x^{LAC} & -r_x^{LTC} & 0 & -r_x^{OC} & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix} \begin{bmatrix} C_{x+1} \\ LA_{x+1} \\ LT_{x+1} \\ SP_{x+1} \\ O_{x+1} \\ I_{x+1} \end{bmatrix} = \begin{bmatrix} 1-MC_x^{\;C} & +r_x^{LAC} & +r_x^{LTC} & 0 & +r_x^{OC} & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix} \begin{bmatrix} C_x \\ LA_x \\ LT_x \\ SP_x \\ O_x \\ I_x \end{bmatrix} + \begin{bmatrix} E_x \\ . \\ . \\ . \\ . \\ ._x \end{bmatrix} \quad (18)$$

Where

$MC_x^C = +u_x^{CLA} * + u_x^{CLT} * + u_x^{CO} *$

$E_x = -D_{x+1}^{\;C} + I_{x+1}^{\;C} - E_{x+1}^{\;C} + MA_{x+1}^{\;C} - ML_{x+1}^{\;C}$

SP_{x+1}= single parents of age x + 1

I_{x+1}= institutional persons of age x+1

The expression above can be simplified by writing it as:

$A_x L_{x+1} = B_x L_x + I_{x+1}$ (19)

In this expression A_x is the first matrix (with transition rates), L_{x+1} the column vector with the population by household position at age x+1, B_x the second matrix with transition rates and I_{x+1} the third column vector with arrivals and leaves from the population

Next, this expression can then be rewritten as follows:

$L_{x+1} = A_x^{-1} B_x L_x + A_x^{-1} I_{x+1}$ (20)

With this formula the population by household position at age x+1 (of each municipality) can be computed. Hereafter, the transitions between household positions can be computed by using formulas like that of 16.

A hidden assumption behind these formula's is that it is supposed that a change between household positions will not be accompanied by a leave to another municipality (or emigration) in the same calendar year. However, in the next year (being in the new household position) they still have the opportunity to go to another municipality (or emigrate).

At this point, for each municipality all (possible) transitions (between household positions), arrivals and leaves can be generated by the model (for each sex). However, a close look may show inconsistencies. An example is the famous two-sex problem: the number of males living

together may differ from the number of females living together (discarding same sex-couples). A simple solution may repair this error: compute the harmonic mean of the numbers of males and females going to live together (and adjust the original numbers).
Another tactic to resolve possible inconsistencies, is to omit some transitions from the system of equations. For instance, including the transition from single parent to living alone in the model could lead to an inconsistency with the number of children leaving the parental home (when in a single parent family the last child has left the single parent is transformed into a person living alone). For this reason only the transition of children leaving the parental home is modelled and in the micro-layer the effect on single parents is determined. However, a bias may be introduced when the process of leaving home is different for household with two parents and single parents. So, it might be necessary not to draw children at random from the pool of households with children, and in stead give varying drawing chances to different household types. Note, that still no definite figures are computed because further adjustments may arise when the regional figures are made consistent with the national total.

The micro-layer of the model

In the micro layer of the model the effects of the transitions between households positions, arrivals and leaves on the number and composition of households are computed. This can be achieved by assigning each person and each household to a separate record in this layer. The record of each person contains its own identification number and also the identification number of the household he or she belongs to. Each household has of course its own identification number but also contains the identification numbers of all the (actual) members of the household. In the application of PEARL as a regional forecasting model, the function of this micro-layer in the regional forecast is only administrative, because no assumptions will be used in this layer. However, PEARL also offers the possibility to use the micro-level as in a micro-simulation model. In that case, risks will be applied to the individual persons and in a bottom-up way figures will be derived for municipalities, higher level regions and ultimately the national level. In the past a major problem connected with micro-simulation models was that they were very time consuming. However, with the modern computers and by clever programming, this is no longer a serious problem. For example, in PEARL loading her starting population of 16 million records takes about 20 seconds and running a single calendar year only takes a few minutes.

Many events primarily have a significance at the personal level: for example a child leaving the parental home in order to live alone. In other cases, events have effects on pairs of persons, such as two singles going to live together. In a lot of events a group of persons has to be considered. Examples are the events of emigration and internal migration, where the relationship between parents and children must be taken into account. As parents and children migrate together, being members of a specific household, the process of migration must be modelled in the micro-layer. For this reason in the micro-layer emigrating household will be picked in order to be submitted to the event of migration. The might create a consistency problem with the macro-layer where numbers of emigrating adult and children are already determined. In general, this inconsistency must be resolved by picking other households in the micro-layer. However, in some occasions the inconsistency can only be resolved by adjusting the figures in the macro-layer. A constraint associated with this procedure, is that the consistency of (adjusted) municipal figures with the national total (coming from the national forecast) must be kept intact (as much as possible).

The micro-layer permits the creation of new information coming from forecasts. If a paradigm might exist in contemporary demographic analysis, the life-cycle perspective is one. In survey's retrospective questions are used to examine changes in the life-cycle. Also the time series of data (at January 1st) derived from the municipal population registers of the Dutch municipalities (in Dutch called 'Gemeentelijke Basisadminstratie') makes this kind of analysis possible (although the time series goes back in time no further than 1995). The micro-layer in PEARL, also paves the way for examining changes in the life-cycle. Now, it is possible to follow all kinds of demographic careers. For example, the internal migration career can be followed. For persons residing in a certain province, it is possible to see to what extent they move to other provinces and eventually return. Another example is to study (the sojourn time in) phases of the life-cycle. For instance, children can be followed after they have left the parental home: how long do they live alone, to what extent they will have a partner and children and at what age will they face the situation of arriving in an empty nest. Especially, the spatial dimension is in this respect interesting. A consequence of varying regional degrees of aging, might be that in some regions relatively more elderly persons will die (and leave a house behind, what might relieve the pressure on the housing market)

4. Modelling short distance migration

Introduction

Without doubt, the most important component of regional forecasts is internal migration. In paragraph 3 an introduction to the modelling of this component is presented. This paragraph will elaborate on the modelling of short-distance migration. This kind of migration is primarily caused by motives related to housing, the living environment and changes in the life-cycle. Further, it is characterised by a very strong distance relationship: the larger the distance the smaller the inclination to depart. In PEARL for each municipality generating migrants over short distance, the distribution over the municipalities of destination has to be determined. This will be done by using a spatial interaction model. As the number of persons leaving the municipality for short distance migration is already known, the model explains the number of migrants arriving in a certain municipality by the attractiveness of the municipality of arrival on the one hand and the distance between the municipality of departure and the municipality of arrival on the other hand. In the application of the regional forecast, the attractiveness of the municipality of arrival will determined by the housing market (the stock of houses plus newly built houses minus condemned buildings). Using the average size of households the attractiveness can be translated into number of persons.

The advantage of using a spatial interaction model in stead of a constant distribution function (as is used in the modelling of long distance migration) is that the model can react to a changing attractiveness, for instance caused by a large stream of newly built houses. This model gives information on streams of migrants moving between municipalities. This is not the case in co-called migrant pool models. In these models, in a first step outmigration from each region is projected by applying outmigration rates, migrants are added into a pool and then in a second step the migrants in the pool are allocated to various regions of destination (see Van Imhoff e.a. , 1995, for more information on topic).

The spatial interaction model

Following Fotheringham en O'Kelly (1989), the spatial interaction model may be formulated as follows:

$M_{ij}= f\ (\mu W_i, \alpha W_j\ , \beta d_{ij})$ (21)

where

M_{ij}= mobility between municipality i and j (=migrants moving from municipality i to j);

W_i = production factor of municipality i;

W_j = attraction factor of municipality j;

d_{ij} = distance between municipality i and j.

Again following Fotheringham en O'Kelly, all three variables will be represented as power functions:

$$f(\mu W_i) = W_i^{\mu} \quad (22)$$

$$f(\alpha W_i) = W_i^{\alpha} \quad (23)$$

$$f(\beta d_{ij}) = d_{ij}^{\beta} \quad (24)$$

In modelling short-distance migration a specific type of the spatial interaction model will be used, namely the 'production-constrained model'. This model only uses an attraction and an distance variable (in our application only one attraction variable will be used, although it is possible to use multiple variables). The choice for the production- constrained model is based on the fact that in the regional forecast the number of persons leaving a municipality is already determined. In the calibration of the model, using observed migration flows, the attraction factor will be represented by the number of inhabitants of the destination municipality. Distance will be measured as distance in bird's eye view. Migrations over a distance of 35 km and less will be considered as short-distance migration (and over a distance of over 35 km as long-distance migration).

The production-constrained model is formulated as follows:

$$M_{ij} = O_i P_j^{\mu} D_{ij}^{\beta} / \Sigma_k (P_k^{\mu} D_{ik}^{\beta}) \quad (25)$$

where

P_j = population of municipality j

O_i = number of persons leaving municipality i

In order to calibrated this model by Ordinary Least Squares, it is transformed in a linear form:

$$\ln M_{ij} - (1/n) \Sigma_j \ln M_{ij} = \mu(\ln P_j - (1/n) \Sigma_j \ln P_j) + \beta (\ln D_{ij} - (1/n) \Sigma_j \ln D_{ij}) \quad (26)$$

The general goodness-of-fit statistic 'R-squared' refers in this case to the transformed interaction model. As we are interested in the difference between the observed and the predicted migration flow also another goodness-of-fit statistic is presented, namely the Average Root Square Error, defined by:

$$ARSE =(1/ \Sigma_i\Sigma_j M_{ij})\sqrt{} (\Sigma_i\Sigma_j M_{ij} - \Sigma_i\Sigma_j M_{ij}')^2 \quad (27)$$

This measure has a lower limit of zero, indicating perfect prediction.

Application of the spatial interaction model

The model is calibrated for each province (of departure) of the Netherlands, based on the observed migration flows in 2002. For illustrative purposes, only the results for two provinces will be shown, namely Flevoland and Noord-Holland. The *table* shows the expected negative effect of distance on migration. According to Ter Heide (1965) this distance deterrence can be explained on the one hand by the fact that in general people have more information on locations at a short distance and on the other hand by the fact that moving over short distances enables people to prolong contacts with families and relatives.

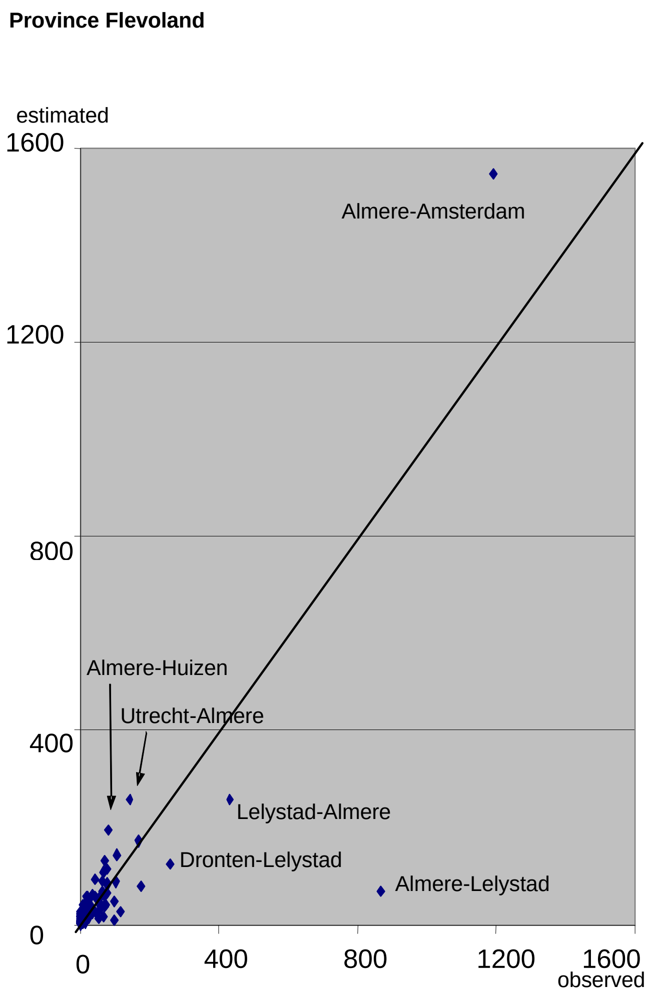
The *graph* show the scattergram of the observed and the estimated migration flows. In 2002 the largest migration flow between two municipalities in the Netherlands was the flow from the capital city of the Netherlands, Amsterdam, to the new-town Almere, namely 3600 persons. According to the spatial interaction model the migration flow should have been much less, namely 600 persons. The unexpected large migration flow can be explained by the amply supply of newly built houses in Almere, stemming from the so-called VINEX task (a large-scale housing program of the Dutch government). These houses are predominantly owner occupied single-family dwellings with a favourable price/quality ratio (Bik en Hooijmeijer, 1997). To a lesser extent, the same situation applies to Haarlemmermeer, although this municipalities also attracts a lot of migrants from Haarlem. In contrast to Almere and Haarlemmermeer, the flow to Zaandam and Amstelveen is overestimated by the model. A tight housing market might be the explanation.

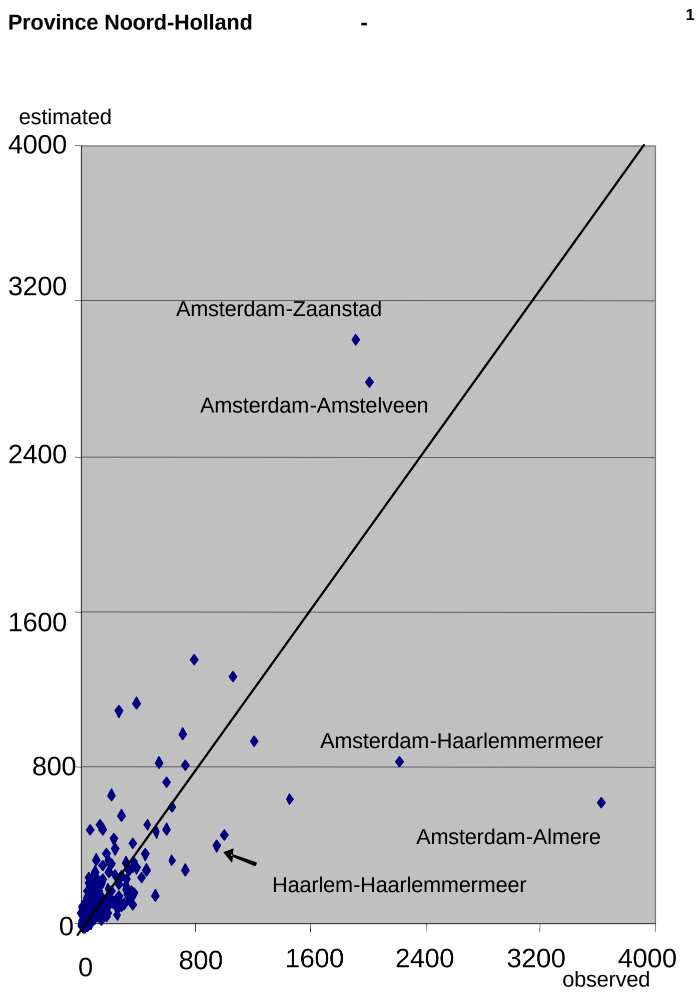
With respect to Flevoland, the largest migration flow concerns the migration from Almere to Amsterdam with some 1200 persons. This large flow can be explained by the escalator-region theory of Fielding (1992). According to Fielding large cities in Western countries function as an escalator for educational and labour career of young people. They move to large cities for their education and start their working career there. Later they move to smaller cities in the surroundings (suburbanisation). Despite an error of about 350 persons the model has predicted the migration between the two municipalities fairly good. This is surely not the case for the migration flow from Almere to Lelystad: the model estimated about 100 persons while the observed number was 850 persons. Lelystad is like Almere a newtown with a large supply of newly built houses. However, the prices are much lower than in Almere. This might explain the unexpected large flow from Almere to Lelystad.

Results of the regression analysis for the provinces Flevoland and Noord-Holland

	Regression coefficient		R^2	Average Root Square Error
	Number of inhabitants of municipality of arrival	Distance		
Flevoland	1,23	-1,56	0,58	0,14
Noord-Holland	0,76	-1,83	0,68	0,06

Scattergram of (short-distance) migration flows in 2002: observed versus estimated flows





Concluding, in general the simple production-constrained model seems to produce rather good estimates of observed migration flows. However, in the calibration attractiveness is solely represented by the number of inhabitants, whereas in the real world a lot of pull-factors might be working. Especially in municipalities with a large number of newly built houses the model produces serious errors. However, in the actual application of the regional forecast attractiveness of a municipality will be represented by the housing market (translated into inhabitants by using average household size), so this kind of errors will be minimal. Moreover, in the model this flow will be considered as a demand for houses in a certain municipality. This demand will then be confronted with the supply of free houses and in an iterative procedure supply and demand will be matched by the model. Other estimation errors are due to the fact that certain municipalities exhibit a strong attraction on surrounding municipalities. In these cases the results from the interaction model can not be applied directly. If for some destiny municipalities such forecasting errors apply, estimated migration flows will be adjusted by a correction factor (computed as the ratio between the estimated number of migrants and the observed number of migrants).

References

Beer, J. de and I. Deerenberg (2005), Regionale verschillen in vruchtbaarheid: een verklarend model. Bevolkingstrends 53 (1), pp. 46-55.

Bik, M. en P. Hooimeijer, 1997, Concurrentie op de woningmarkt: Almere en Amsterdam. Rooilijn, n.7, blz. 335-341, Amsterdam.

Eichperger, Ch.L., B.A. van Hamel and W.P. Nieuwenhuis (1979), Een methodiek voor de vervaardiging van regionale bevolkingsprognoses. Studierapport RPD no. 15, ’s-Gravenhage, Staatsuitgeverij.

Fielding, A.J., 1992, Migration and social mobility – South East England as an escalator region. Regional Studies 26(1), blz. 1-15.

Fotheringham, A.S. and M.E. O’Kelly, 1989, Spatial Interaction Models: Formulations and Applications. Kluwer Academic Publishers, Dordrecht/ Boston/ London.

Gordijn, H., H. Heida, H. den Otter (1983), PRIMOS, prognose-, informatie- en monitoringssysteem voor het volkshuisvestingsbeleid. Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, Directoraat-Generaal van de Volkshuisvesting, Directie Onderzoek.

H.R. Heida (2002), Primos, prognosemodel voor bevolking, huishoudens en woningbehoefte. Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer.

Heide, H. ter, 1965, Binnenlandse migratie in Nederland. Staatsuitgeverij ’s-Gravenhage.

Hooijmeijer, P. and M. Linde (1988), Vergrijzing, individualisering en de woningmarkt, het WODYN-simulatiemodel. Drukkerij Elinkwijk BV, Utrecht.

Imhoff, E. van, N. van der Gaag, L. van Wissen, P. Rees (1995), Model Selection For Internal Migration At The NUTS-2 Level. Paper submitted to the international seminar ‘New Long-Term Population Scenarios for the European Economic Area, Luxembourg, 8-10 November 1995.

Imhoff, E. van and N.W. Keilman (1991), LIPRO 2.0: An Application Of A Dynamic Demographic Projection Model To Household Structure In The Netherlands. Swets&Zeitlinger B.V. Amsterdam/Lisse.

Jong, A.de and J. de Beer (2001), Het huishoudensprognosemodel. Maandstatistiek van de bevolking 49(7), pp. 16-20.

Keilman, N. (1984), Integratie en consistentie van regionale en nationale bevolkingsprognoses. NIDI, intern rapport nr 36.

Keilman, N.W. (1985), Internal and external consistency in multidimensional population projection models. Environment and Planning A, 17, pp. 1473-1498.

Nelissen, J.H.M. and A.P. Vossen (1989), Projecting household dynamics: A scenario-based microsimulation approach. European Journal of Population, 5, pp. 253-279.

Willekens, F.J., F.W.A. van Poppel, N.W. Keilman, Ch. Eichperger and P. Drewe (1988), MUDEA-prognoses van de bevolking van de COROP gebieden en de vier grote steden. Rijksplanologische Dienst.