Model for Population Projections for Norwegian Regions

By EIVIND GILJE[1]

Central Bureau of Statistics of Norway, Oslo

I. Introduction

The Central Bureau of Statistics of Norway has for some time worked with plans for a large population model which can be used for different analytical purposes. The intention is that it will be a kind of parallel to economic models. As a first step in this direction we have prepared a projection model which is used to produce regional population projections. This model will be discussed in the present paper.

We have constructed a model which gives factor projections. It is quite similar to classical projection models of this type. The most important new elements are our more specified data and the fact that the model provides far more detailed results than any other model I have seen. For the time being, we have refrained from making further refinements. Our main purpose in the initial stage has been to prepare an operable model which can provide us with fair results before we begin working on statistical and technical improvements.

Even though modern data processing techniques have permitted great detail, the methods are still quite unsatisfactory on many points. We shall attempt to improve on this situation in the recently established Study Group for Population Models. Some demographic models of a more restricted scope are being developed by this group, and analyses are being carried out on data collected by the Central Bureau of Statistics. With time we hope to be able to utilise the results of this work for improving the projection model. This will be discussed further in the last chapter of this paper.

Chapter II contains a description of the model we have prepared. I have done my best to give a correct probabilistic justification for the formulae in the model.

In chapter III procedures for the estimation of the structural coefficients are given.

---

[1] I am grateful to Mr. Jan M. Hoem for his advice and guidance, and to Mr. Fridjof Wiese, who has proof-read the manuscript of this paper.

In chapter IV we shall take a look at the restrictions placed on the model by the available data.

II. Description of the model

1. *Conventions and definitions*

I shall use the following symbols as superscripts and subscripts:
Left superscript:
Sex: M for males, F for females
Right superscript:
Municipality no.: k, where $k = 1,2,\cdots,K$.
Right subscript:
Age: x, where $x = 0,1,\cdots,\omega$. (ω is the highest age.)
In addition, n in parentheses after the symbols represents January 1 of the year for stock concepts, and the year for flow concepts.

In order to illustrate the use of symbols I shall give the following examples:

${}^{F}L^{k}_{x}(n)$ represents the actual number of x-year old women in municipality k per January 1 of the year n.

${}^{F}D^{k}_{x}(n)$ represents the actual number of deaths in municipality k in year n among women who were x-years old per January 1 of the year n.

${}^{F}{}_{t}q^{k}_{x}(n)$ is the probability that a woman who per January 1 of the year n lives in municipality k and is x years old, shall die in the municipality before she is $x + t$ years old.

Other stock concepts are defined in a manner similar to ${}^{F}L^{k}_{x}(n)$, and other flow concepts are defined like ${}^{F}D^{k}_{x}(n)$.

Furthermore, we shall introduce the convention that if X is a random variable, $\tilde{X}$ shall denote a predictor for X. (Admundsen, 1962, 2, p. 198, and 3, p. 252.) Thus, for example, $\widetilde{{}^{F}L^{k}_{x}}(n)$ will be a predictor for ${}^{F}L^{k}_{x}(n)$.

A symbol like ${}^{F}{}_{t}\hat{q}^{k}_{x}(n)$ denotes an estimator for ${}^{F}{}_{t}q^{k}_{x}(n)$.

Many of the relations we arrive at are analogous for men and women. We therefore introduce the typographical simplification that symbols with no sex designation represent one single, unspecified sex.

2. *The projection model*

In this section I shall give the argument behind each relation in the model. These relations will be summarised in chapter II.3.

Let L denote the actual population, D the number dead, U the number of out-migrants, and I the number of in-migrants. For $0 \leqq x \leqq \omega$ we then have:

(2.1) $L^k_{x+1}(n+1) = L^k_x(n) - D^k_x(n) - U^k_x(n) + I^k_x(n)$.

When n refers to future years, however, the items on the right hand side of (2.1) will be unknown. In a projection model one will then replace them with predicted values. In accordance with (2.1) we can then present the *first relation in the projection model:*

(2.2) $\tilde{L}^k_{x+1}(n+1) = \tilde{L}^k_x(n) - \tilde{D}^k_x(n) - \tilde{U}^k_x(n) + \tilde{I}^k_x(n)$

for $n = 0,1,\cdots$; $k = 1,2,\cdots,K$ and $0 \leqq x \leqq \omega$. Here $n = 0$ denotes an initial point of time with a known stock, so that $\tilde{L}^k_x(0) = L^k_x(0)$ for $0 \leqq x \leqq \omega$ and $k = 1,2,\cdots,K$.

It is natural to start with estimators for the expected values for $D^k_x(n)$, $U^k_x(n)$, and $I^k_x(n)$ in order to find predictors for these. Predictors for $L^k_x(n)$ are then calculated recursively by (2.2).

In the following we shall examine these estimators. Let us then first consider $D^k_x(0)$ and $U^k_x(0)$. Both of these random variables are binomially distributed, and we introduce $q^k_x(0)$ and $u^k_x(0)$ as one-year influenced probabilities for death and out-migration, respectively (Sverdrup, 1961, p. 23).

Those who die in a municipality in a certain year can be divided into two categories. The one comprises those who lived in the municipality at the beginning of the year. The other includes those who have migrated to the municipality after January 1st and who have died later in the year. The deaths among the in-migrants in the course of the year are assumed to be of negligible significance, so we have:

(2.3) $E\,D^k_x(0) = q^k_x(0) \cdot L^k_x(0)$.

We similarly assume that a negligible number of persons migrate more than once in the course of one year, and get:

(2.4) $E\,U^k_x(0) = u^k_x(0) \cdot L^k_x(0)$.

We will comment further upon the latter assumption at the end of this section.

We now turn to finding an expression for the in-migrants. First we introduce

(2.5) $U_x(0) = \sum_{k=1}^{K} U^k_x(0)$.

Let $i^k_x(0)$ be the probability that an out-migrant of age x per January 1 of year 0 shall move to municipality k in the course of the same year. Then

(2.6) $E\{I^k{}_x(0) \mid U_x(0)\} = i^k{}_x(0) \cdot U_x(0)$

and thereby:

(2.7) $E\, I^k{}_x(0) = i^k{}_x(0) \cdot E U_x(0)$.

Behind this procedure there is a theory which can be formulated in the following manner: A person's tendency to move from the municipality in which he is living, is only dependent upon the characteristics of *this municipality*. This leads to (2.4). Where the person decides to move to is, on the other hand, not at all dependent upon the municipality from which he is moving, but wholly on the characteristics of the possible in-migration municipalities. Given that an x-year old person decides to move, the conditional probability that he will move to municipality k could be written as a function $i^k{}_x$ which is independent of where he lived at the beginning of the year. This gives a formula like (2.7) (Hoem, 1968, p. 14).

As predictors for $D^k{}_x(0)$, $U^k{}_x(0)$, and $I^k{}_x(0)$ we will use estimators for their expecter values and then get equations (2.8), (2.9), (2.10), and (2.11) below for $n = 0$. Combining these formulae with (2.2), we can let the sets of equations apply for all n.

(2.8) $\tilde{D}^k{}_x(n) = \hat{q}^k{}_x \cdot \tilde{L}^k{}_x(n)$,

(2.9) $\tilde{U}^k{}_x(n) = \hat{u}^k{}_x \cdot \tilde{L}^k{}_x(n)$,

(2.10) $\tilde{U}_x(n) = \sum_{j=1}^{K} \tilde{U}^j{}_x(n)$, and

(2.11) $\tilde{I}^k{}_x(n) = \hat{i}^k{}_x \cdot \tilde{U}_x(n)$,
where $k = 1,2,\cdots,K$; $0 \leq x \leq \omega$ and $n = 0,1,\cdots$.

Age 0 applies to children born in the year n-1. We will set x equal to —1 for children born in the projection year, i.e. year n. With this convention, (2.9) to (2.11) hold also for $x = -1$.

We have not dated the estimators for the structural coefficients $q^k{}_x(n)$, $u^k{}_x(n)$ and $i^k{}_x(n)$. The reason is that in the projection model we let mortality and migration tendency be the same as in the initial year throughout the entire projection period. Below, we will make the same assumption for the expected number of births.

From (2.2), (2.8), (2.9), (2.10), and (2.11) we can now give projections for all age groups excluding those who are born in the projection year. We now consider these births. Let $f^k{}_x(n)$ be the expected number of live births in year n of a woman who per January 1 of the year n is x years old and lives in municipality k (Hoem, 1967, p. 57), and let $B^k{}_x(n)$ be the number of children these ${}^F L^k{}_x(n)$ women actually have in that year. Then

$$E\{E(B^k_x(n) \mid {}^F L^k_x(n)\} = E\{f^k_x(n) \cdot {}^F L^k_x(n)\}$$

which gives

$$E\,B^k_x(n) = f^k_x(n) \cdot E\,{}^F L^k_x(n)$$

and the expected total number of births in the stock in year 0 will therefore be

$$(2.12)\quad E\,B^k(0) = \sum_{x=15}^{44} f^k_x(0) \cdot {}^F L^k_x(0),$$

where we have added over the women's fertile ages. In line with earlier deductions we find the following predictors:

$$(2.13)\quad \widetilde{B}^k(n) = \sum_{x=15}^{44} \hat{f^k}_x \cdot {}^F \widetilde{L^k}_x(n)$$

for $k = 1,2,\cdots,K$; and $n = 0,1,\cdots$. We must be aware that also (2.12) and (2.13) are approximation formuale. Similar to what we have done earlier, we have neglected births by women who migrate to municipality k and give birth in this municipality later in the year.

Let ${}^F c$ denote the proportion of girls in one birth cohort. Since births occur »on the average in the middle of the year», we can present the formula:

$$(2.14)\quad {}^F\widetilde{L}^k_0(n+1) = \{1 - ({}^{F_{1/2}}\hat{q}^k_0 + {}^{F_{1/2}}\hat{u}^k_0)\} \cdot {}^F c \cdot B^k(n) +$$

$${}^F\hat{i}^k_{-1} \cdot {}^F\widetilde{U}_{-1}(n);$$

$$k = 1,2,\cdots,K;\ n = 0,1,\cdots.$$

(For the use of the subscript —1 in the last member, see the explanation after (2.11).) Here ${}^F\widetilde{L}^k_0(n+1)$ is the predictor for the number of live new- born girls who live in municipality k at the end of the calendar year in which they are born. A corresponding prediction formula applies for boys. In the last member of (2.14) we should have had a correction factor for death and migration after in-migration. Since on the average the new-borns run the risk of dying or migrating again after the in-migration only during a short period, we assume that the correction factor will not be of any great significance.

In several places in the model we have made similar assumptions that several »events» can not occur for one and the same person in one and the same year. (»Events» must be interpreted here as something we take into consideration in the model.) We shall briefly summarise these and comment on them.

We assume that one person cannot both migrate and die in one calendar year. A simple arithmetical example gives us an idea of how much this can amount to. The figures are relevant for the situation in Norway today. Out of 200,000 migrants the vast majority are between the ages of 20 and 40. The one-year probability of death is at most approximately 0.002 for these age groups. If we assume that the migrations occur »on the average» in the middle of the year, not more than approximately 200 of the 200,000 migrants will die in the remainder of the year.

In the model we also leave out of account the fact that persons can migrate several times in the course of one year. We are, however, really only interested in where a person is living at the end of each calendar year. We can therefore evade this problem by disregarding all these intermediate moves in estimating the migration probability. As a special case, we will not register any migration for a person who by the end of the year winds up in the same municipality from which he first migrated during the year.

A woman who has given birth, can either die or migrate after the birth, but in the model we have not taken into consideration the fact that she can give birth after the migration. After the in-migration the remaining time during which the woman runs the risk of giving birth will »on the average» be as long as half a year. We cannot therefore assume that this omission has negligible significance. In estimating the expected number of births, it has, however, not been possible to follow the *individual* woman from the beginning of the year to the end of the year. All the births which are registered in a municipality are therefore ascribed to those women who lived there at the beginning of the year. This also holds true for births by women who have migrated there in the course of the year. This estimating error works in the opposite direction from the above-mentioned omission and should therefore completely or partially compensate for this.

Most of the simplifications discussed here are made to facilitate programming. The problems could perhaps be solved in theory in a statisfactory manner, but in such a case the computer time would be considerably longer. The danger does not of course lie in the errors made for a single projection year, but in the fact that the errors are accumulated over the years. They can therefore be of some significance if we prepare projections for many years ahead.

3. *A summary of the relations in the model*

The projection model thus contains the following relations:

(2.2) $\tilde{L}^k{}_{x+1}\,(n+1) = \tilde{L}^k{}_x\,(n) - \tilde{D}^k{}_x\,(n) - \tilde{U}^k{}_x\,(n) + \tilde{I}^k{}_x\,(n).$

(2.8) $\tilde{D}^k{}_x\,(n) = \hat{q}^k{}_x \cdot \tilde{L}^k{}_x\,(n).$

(2.9) $\quad \tilde{U}^k{}_x\,(n) = u^k{}_x \cdot \tilde{L}^k{}_x\,(n).$

(2.10) $\quad \tilde{U}_x\,(n) = \sum_{j=1}^{K} \tilde{U}^j{}_x\,(n).$

(2.11) $\quad \tilde{I}^k{}_x\,(n) = \hat{i}^k{}_x \cdot \tilde{U}_x\,(n).$

(2.13) $\quad \tilde{B}^k\,(n) = \sum_{x=15}^{44} \hat{f}^k{}_x \cdot {}^F\tilde{L}^k{}_x\,(n).$

(2.14) $\quad \tilde{L}^k{}_0\,(n+1) = \{1 - ({}_{1/2}\hat{q}^k{}_0 + {}_{1/2}\hat{u}^k{}_0)\} \cdot c \cdot \tilde{B}^k\,(n) + \hat{i}^k{}_{-1} \cdot \tilde{U}_{-1}\,(n)$

for $k = 1,2, \cdots, K;\ 0 \leq x \leq \omega$ and $n = 0,1, \cdots$. Formulae (2.9) to (2.11) hold also for $x = -1$.

III. Estimation of probabilities for death and migration and expected number of births

We shall first look at the probabilities for death and out-migration which are estimated in an analogous manner.

Let the remaining lifetime for a person of age x be T_x. We introduce the distribution function F_x by

(3.1) $\quad F_x\,(t) = P\,(T_x \leqq t) = {}_tq_x,$

and the force of mortality (Hoem, 1967, p. 97):

(3.2) $\quad \mu_x\,(t) = \dfrac{\dfrac{d}{dt}\,F_x\,(t)}{1 - F_x\,(t)}.$

We assume that $\mu_x\,(t)$ is independent of t within a time interval of one year. Then we clearly have:

(3.4) $\quad F_x\,(t) = 1 - \exp\,\{-\,tu_x\}$

for $0 \leqq t < 1$, which is the exponential distribution function.

Let us now consider a person over the interval [t, 1], i.e. in a period of lenght 1-t. (t is still less than 1.) The number of times the person dies in the period is ${}_tM$, and the number of times he migrates from the municipality in which he is living at time t is ${}_tN$. We have $({}_tM,\ {}_tN)\ \varepsilon\{\ (0,0),\ (0,1),\ (1,0)\ \}$. Moreover, the force of out-migration σ is defined similarly to μ. We can temporarily assume that μ and σ are constants independent of sex, age, municipality, and year. We have

$${}_{1-t}q_{x+t} = E\,{}_tM = P\,({}_tM = 1) = \int_0^{1-t} e^{-(\mu+\sigma)\tau}\,\mu d\tau = \frac{\mu}{\mu+\sigma}(1 - e^{-(\mu+\sigma)\,(1-t)}),$$

and for $t = 0$

$$(3.5) \qquad q_x = \frac{\mu}{\mu+\sigma}\,(1-e^{-(\mu+\sigma)}).$$

According to Sverdrup (1961) one should first find estimators for μ and σ when we shall estimate q and u. Let us say that we have data for each of the years m to (m+t—1). Estimators for average forces of mortality and migration in the period are then

$$(3.6) \qquad \hat{\mu}^k{}_x\,(m, t) = \frac{\sum\limits_{j=1}^{t} D^k{}_x\,(m+j-1)}{\sum\limits_{j=1}^{t} M^k{}_x\,(m+j-1)}$$

$$(3.7) \qquad \hat{\sigma}^k{}_x\,(m, t) = \frac{\sum\limits_{j=1}^{t} U^k{}_x\,(m+j-1)}{\sum\limits_{j=1}^{t} M^k{}_x\,(m+j-1)}.$$

$M^k{}_x\,(m+j-1)$ is the observed, aggregated lifetimes in municipality k and calendar year (m+j—1) of persons of age x at the beginning of the year. According to Sverdrup (1961) this quantity may be approximated by

$$(3.8) \qquad M^k{}_x\,(m+j-1) \approx L^k{}_x\,(m+j-1) - \tfrac{1}{2}\,(D^k{}_x\,(m+j-1) + U^k{}_x\,(m+j-1) - I^k{}_x\,(m+j-1)).$$

It follows from (3.5):

$$(3.9) \qquad \hat{q}^k{}_x\,(m, t) = \frac{\hat{\mu}^k{}_x\,(m, t)}{\hat{\mu}^k{}_x\,(m, t) + \hat{\sigma}^k{}_x\,(m, t)}.\;\{1-\exp\,(-\hat{\mu}^k{}_x\,(m, t) - \hat{\sigma}^k{}_x\,(m, t)\,)\,\}$$

and

$$(3.10) \qquad \hat{u}^k{}_x\,(m, t) = \frac{\hat{\sigma}^k{}_x\,(m, t)}{\hat{\mu}^k{}_x\,(m, t) + \hat{\sigma}^k{}_x\,(m, t)}.\;\{1-\exp\,(-\hat{\mu}^k{}_x\,(m, t) - \hat{\sigma}^k{}_x\,(m, t)\,)\,\}.$$

$q^k{}_x\,(m, t)$ and $\mu^k{}_x\,(m, t)$ in (3.9) and (3.10) are estimators for average probabilities for death and out-migration in the years m to (m+t—1). Details concerning the estimators' probabilistic properties are given elsewhere (Sverdrup, 1961).

As an estimator for $i^k{}_x\,(m, t)$ we have used:

$$(3.11)\quad \hat{i}^k_x(m,t) = \frac{\sum_{j=1}^{t} I^k_x(m+j-1)}{\sum_{j=1}^{t} U_x(m+j-1)}$$

That this is unbiased can be seen from:

$$E\{\hat{i}^k_x(m,t) \mid \sum_{j=1}^{t} U_x(m+j-1)\} =$$

$$\frac{1}{\sum_{j=1}^{t} U_x(m+j-1)} E\{\sum_{j=1}^{t} I^k_x(m+j-1) \mid \sum_{j=1}^{t} U_x(m+j-1)\} = i^k_x(m,t)$$

and thus $E\,\hat{i}^k_x(m,t) = i^k_x(m,t)$. The equation holds true only for $\sum_{j=1}^{t} U_x(m+j-1) > 0$, but for $\sum_{j=1}^{t} U_x(m+j-1) = 0$, we will have $\hat{i}^k_x(m,t) = 0$ for all k.

$f^k_x(m,t)$ is estimated by:

$$(3.12)\quad \hat{f}^k_x(m,t) = \frac{\sum_{j=1}^{t} B^k_x(m+j-1)}{\sum_{j=1}^{t} {}^{F}L^k_x(m+j-1)}$$

and in the same manner as above we find that $\hat{f}^k_x(m,t)$ is unbiased.

IV. The data

During the work with the model we have of course had to take into consideration the quality of the data at our disposal. These data should be detailed and relatively easy to process. Because of these two requirements, we have, for example, temporarily had to disregard a break-down by marital status. Similarly, we have not been in a position to take into consideration emigration and immigration. (We hope to include both of these features in new versions of the model.) Even so we have more data available than most other countries.

Our initial stock is the Norwegian population of January 1, 1966, which we have obtained from the newly established Population Register for Norway. The main breakdown is by municipalities. Within each municipality the number of persons within each one-year age group distributed by sex have been given. This detailed breakdown exists for all the data which have been used.

Estimates for probabilities for death and migration and expected number of births have been calculated from data for the one year 1966.

(t in the formulae (3.6) to (3.12) is equal to 1). It would have been desirable to use figures for several years, but because of several municipality mergers the figures for the various year up to 1966 are not directly comparable.

It is our intention to utilise new statistics in the current work with projection models as soon as they become available.

V. Application and planned improvements of the model

During the preparation of this model we have always been aware that we would be interested in improvements as soon as the first version was available. We have taken this into consideration during the programming work, and have developed a quite flexible system. Instead of large programmes which carry out many operations at one time, we have broken the system down into a number of blocks where changes made in one block have little or no significance for the rest of the system.

When developing this system we have also had in mind the fact that we know from experience that there will always be an interest for alternative projections. As an example, we are now in the process of producing projections where calculations are made as if all movements between municipalities have been suspended. The migration relations in the model are then of no interest, and we have removed the corresponding blocks in order to save work and computer time. This will not effect the rest of the system.

In the first test runs it appeared that the estimates for the regional mortality probabilities were unsuitable for small municipalities. Even for Oslo with a data basis of approximately 500,000 people the estimates showed large fluctuations in the oldest age groups. We have therefore temporarily used the mortality rates for the whole country for the years 1961—1965 (Vital Statistics and Migration Statistics for Norway, 1965; Tab. XXII) as estimates for the death probabilities in all municipalities. Since we have little emigration in Norway, these rates will be estimates for something which approximates partial death probabilities. When we then predict the deaths in a municipality, we really also obtain deaths among those who have migrated from the municipality. We have not attached any particular significance to this inconsistency.

The first improvements which we have planned are particularly due to the limited data basis for the estimation of probabilities of death and migration and expected number of births in each municipality. Even if we had the data for say 5 years, the variance on an average over the years would be very large for small municipalities. We are now thinking of combining the municipalities into groups in order to obtain a larger data basis per region. We have not yet decided how we shall form these groups of municipalities, but a study concerning the dependency of migrations on variables both of a demographic and of a non-demographic nature has been started.

The projected birth figures are also a matter of uncertainty in population projections. It appears that the birth cohorts in the projections we now have produced will increase more than is reasonable. We have certain theories concerning the reasons for this, but improvements on the model will be postponed until further studies have been made.

REFERENCES

Amundsen, H. T. (1962): »*Innføring i teoretisk statistikk. Bind 1—3.*» (Introduction to Teoretical Statistics, Volumes 1—3.) Memorandum of May 28, 1962, from the Institute of Economics, the University of Oslo, and Oslo University Press.

Hoem, Jan M. (1967): »*Grunnbegreper i formell befolkningslære.*» (Basic Concepts of Formal Demography.) Memorandum of February 7, 1967 from the Institute of Economics, the University of Oslo, and Oslo University Press.

Hoem, Jan M. (1968): »*Befolkningsprognosemodellens flyttingsrelasjoner. I.*» (The Migration Relations of the Population Projection Model). Unpublished Working Paper IO 68/11 from the Central Bureau of Statistics, Oslo.

Norway's Official Statistics XII 220 (1967): »*Vital Statistics and Migration Statistics 1965.*» The Central Bureau of Statistics, Oslo.

Sverdrup, Erling (1961): »*Statistiske metoder ved dødelighetsundersøkelser.*» (Statistical Methods in Mortality Studies), Institute of Mathematics, University of Oslo.