CHARLES UNIVERSITY IN PRAGUE

Faculty of Science

Department of Demography and Geodemography

REGIONAL POPULATION FORECAST FOR THE REPUBLIC OF KAZAKHSTAN

Dissertation

Altynay Jazybayeva

Prague 2012 Thesis Supervisor: Prof.RNDr.Jitka Rychtaříková, CSc.

I declare that this dissertation is my own work under supervision of Prof.RNDr. Jitka Rychtaříková, CSc. Where other sources of information have been used, they have been acknowledged.

I agree that if any results gained while working on this dissertation will be used outside the Charles University in Prague, written permissen of the University will be necessary.

I agree to lend this dissertation for study reasons and agree that this thesis will be added to the borrower's database.

In Prague, 2012

. . . .

Signature

ACKNOWLEDGMENTS

I am grateful to my supervisor Professor Jitka Rychtaříková for her support, consultations and valuable comments. I thank all teachers for generously sharing their knowledge and views, it was vital and encouraging in writing this thesis. I would also like to thank officers from JSC 'Center for International Programs' for support I received.

I am thankful to my parents and sisters for helping me.

Regional population forecast for the Republic of Kazakhstan

Abstract

This dissertation has three objectives. The first objective is to present literature review about theoretical background of regional population forecast. The second objective is to analyze demographic situation with relation to past and current fertility, mortality and migration development in regions of Kazakhstan. The third objective is to demonstrate two practical implementations of regional population projections. The first example is a multiregional population projection with population horizon 2009-2029 for 16 administrative divisions of Kazakhstan using period data for the year 2008 and inferring required age-sex specific interregional transition data. The second example is a multiregional population projection for period 2004-2059 of four macroregions using period-observational plan 2004-2008 and imposing internal consistency relations. The second example follows generations of people born during period of recovering fertility when these generations will be approaching retirement ages.

Keywords: multiregional population projections, internal migration, consistency restraints

Regionální populační prognóza Republiky Kazachstán

Shrnutí

Tato disertace má tři cíle. Prvním cílem je představit přehled literatury týkající se teoretického základu regionální populační prognózy. Druhým cílem je analyzovat demografickou situaci ve vztahu k minulosti a aktuální plodnosti, úmrtnosti a rozvoji migrace v regionech Kazachstánu. Třetím cílem je demonstrovat dvě praktické implementace odhadu obyvatelstva v regionech. Prvním příkladem je multiregionální populační odhad obyvatelstva s horizontem na r. 2009 až r. 2029 pro 16 administrativních okruhů Kazachstánu pomocí dobových údajů za r. 2008 a vyvození požadovaných přechodných specifických meziregionálních dat ohledně věku a pohlaví. Druhým příkladem je multiregionální populační prognóza na období od r. 2004 až do r. 2059 pro čtyři makro-regiony s využitím pozorovacího plánu od r. 2004 do r. 2008 a s přihlédnutím k vnitřním stálým vztahům. Druhý příklad se navazuje na generace lidí narozených v době obnovení plodnosti, kdy tyto generace se blíží věku odchodu do důchodu.

Klíčová slova: multiregionální populační prognóza, vnitřní migrace, stálá zamezování

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Despite the proverb ‘no man is a prophet in his own country’ this work is devoted to regional population forecast of the Republic of Kazakhstan. The results of regional population prognosis are very import and can be used by governmental and local authorities in planning tasks, allocation and distribution of different types of resources. For example, pension system, insurance system, building of educational facilities and medical hospitals, all spheres of life oriented and depending on the dynamic and changing structure of population require regional population prognosis. Diversity of interests caused an increasing number of different types of regional projections. Many government statistical offices publish their sub-national population projections, for instance Statistics Canada, Office for National Statistics United Kingdom, Statistics Norway, Statistics of New Zealand etc., international organizations like the United Nations, the Eurostat, the World Bank issue their regional population projections, some scientific research and educational institutions and organizations, for example the International Institute for Applied Systems Analysis (IIASA), the Netherlands Interdisciplinary Demographic Institute (NIDI) publish their works about regional population projections. At the same time demographic prognoses are highly uncertain and suffer from inaccuracy, Keyfitz confirms this point: “The best demographers do it, but none would stake their reputation on the agreement of their forecasts with the subsequent realization. (Booth 2006:548).

That is possible to meet the official caution accompanying regularly published results of population prognosis, for example “Population projections are not forecasts - they simply provide the population levels and structure that would result if assumptions about fertility, mortality and migration were realized. In the subnational projections, these assumptions are based primarily on recent observed demographic trends and so don't reflect, for example, the impact of government policies or likely housing development in an area” (Office for National Statistics United Kingdom).

Prediction of the future of population is not as looking at the crystal ball like people dreamed in the ancient time, now it can be viewed like shooting moving erratically target, the lucky one who shoots before the target will move. Improvements of theoretical models and their complications do not directly result in more precise and accurate prognosis. Changing predictability and historical character of prediction methodology caused many to abjure the idea of forecasting at all. Case in point, the notion of “forecast” was extensively used in the late 1940’s in the United States, but after exposed the gross errors caused by the baby-boom, the term was cautiously replaced with “illustrative projections”, then with “projections”. Nevertheless, demand breeds supply, that’s why increased number of research concerning different methods of demographic prognosis. Accepting the demographic uncertainty this work hold the view expressed by Harold Dorn: “Predictions, estimate, projections, forecast; the fine academic distinction among these terms is lost upon the user of demographic statistics. So long as numbers which purport to be possible future populations are published they will be regarded as forecasts or predictions, irrespective of what they are called by demographers who prepare them” (Alho, Spencer 2005:226).

Producing population forecast and publishing the results can cause feedback effects. These effects can be distinguished between self-defeating forecast and self-fulfilling forecast. People’s decisions concerning health behavior, additional births, or moving from one place to another depend on social or community level values and other compelling forces. Purpose of forecast is to compel such values. For example, in 1920s and 1930 many European countries experienced population decline, then cohort-component forecasts were made in order to prevent population decline. Such type of forecast is called self-defeating. Self-fulfilling forecast are produced with intention to justify the building of new facilities, for example, forecasts of increasing net migration can caused increasing number of future migrants to substantiate the build-up infrastructure (Alho, Spencer 2005).

Nevertheless, attempts to affect fertility in industrialized countries shows inefficiencies of policies, this sometimes true in relation to immigration. For instance, Europe in the 1970s was inspired by enthusiasm concerning prospects of social planning. In Finland by government decision population forecast implemented new assumptions concerning population plan on the regional level. Conceptual ideas were elaborated in accordance with planning tools developed in Sweden, Norway, Italy, France, the Netherlands, the United Kingdom etc. and included estimation of regional economies, population and their change. But this plan was ineffective and didn’t work, and soon was abdicated. Summarizing, population prognosis influence demographic behavior sometimes indirectly due to complexity of people’s attitudes, social norms and implementation (Alho, Spencer 2005).

Inaccuracy of forecast can be caused by expected error and empirical error. Expected error relates to error assessed at the time a forecast made, before the future unfolds. Empirical error relates to errors assessed after the future has unfolded and the attained values of the process have become observed. It is customary to call these as *ex ante* and *ex post* errors. Expected error is usually required by user of forecast, and can be assumed similar to the past errors while *ex post* error helps to improve the methodology of forecast. Expected error is always model based. Misspecification of model leads to the wrong error assessment. For example, overfitting causes an underestimation of *ex ante* error, inversely fitting ARIMA model to twice differenced data series even in case of small residual variance can lead to forecast intervals that eventually cover values that are, in Whelpton's words, "incompatible with present knowledge", and *ex ante* error can exceed *ex post* error (Alho, Spencer 2005:238).

According to Hoem's classification sources of inaccuracy can be divided into three main categories: 1) estimation and registration errors; 2) errors due to random fluctuations; 3) erroneous trends in the mean vital rates. The first type arouses from parameter estimates and basic data (jump-off population and vital rates). The second includes the inherent stochasticity of the vital rates (e.g. binomial or Poisson, and random variation in their expectations). The third embodies model mis-specification.

In compliance with Alho's classification sources of inaccuracy are defined from perspective of statistical modeling:

"(1) model mis-specification: the assumed parametric model is only approximately correct;

(2) errors in parameter estimates: even if the assumed parametric model would be the correct one, its parameter estimates will be subject to error when only finite data series are available;

(3) errors in expert judgment: an outside observer may disagree with our judgments or 'prior' beliefs about parameters of the model;

(4) random variation, which would be left unexplained even if the parameters of the process could be specified without any error; since any mathematical model is only an approximation, one would expect there to be random variation."

These four categories of errors are contingent on each other. Data errors now belong to the third category. The most important category is either model mis-specification or error of judgment (Alho, Spencer 2005:239).

Concerning the consistency of regional population forecast with national population projection Keyfitz noted that when a heterogeneous population is disaggregated by some attributes in this case by region of residence, then "sum of the separate projections will be greater than the projection of the sum of the parts at the average rate of increase prevailing at the start. The aggregation introduces a projection bias" (Rogers 1981:3). In other words, "when one projects a heterogeneous population in disregard of the heterogeneity, which is to say using the average rate of increase for the whole, one underestimates the subsequent population. To project the population of the United States, for example, with the parameters of the country as a whole necessarily gives a lower answer than projecting each state with its own parameters, and then taking the total for the United States." For example, the United States population in 1966 as projected with a life table made from the deaths and births of the same year, results in a total population for 1981 of 230,477,000. Disaggregating population into two separate groups of Whites and Nonwhites, whose 1966 populations, births, and deaths add exactly to the totals for the United States, constructing life tables and age-specific birth rates for the two groups separately, and then projecting each by means of its own life table and birth rates, gives 1981 Whites as 199,287,000 and Nonwhites as 31,441,000, which add to 230,728,000. The difference is 251,000 and after100 years the two separate projections add to 8 percent more than the projection without breakdown by color (Keyfitz, Caswell 2005:278,484).

Many studies evaluated the historical population forecasts by comparison with observed statistics, and they revealed the following empirical evidence about practice of forecasting (Keilman 2005):

1. Forecasts are more accurate for short than for long forecast durations.
2. Forecasts are more accurate for large than for small populations.
3. Forecasts of the old and the young generations tend to be less accurate than forecast of working age groups.
4. Accuracy differs between components and regions.

Outline of the thesis

This dissertation has an objective to evaluate future population development in regions of Kazakhstan by analyzing previous and current regional demographic situation.

The thesis is organized in the following order. Chapter 1 describes mathematical bases including cohort-component method, multistate models, stable population theory and weak ergodicity.

Chapter 2 contains brief review of recent development and achievements concerning modelling and forecasting mortality, fertility and migration.

Chapter 3 designates characteristics of variety of regional population models.

Chapter 4 is devoted to general information about administrative divisions of Kazakhstan and migration processes.

Chapter 5 describes fertility and mortality in regions of Kazakhstan.

Chapter 6 shows methodological aspects of multiregional approach such as disaggregation of migration data by sex and age as well as imposing consistency restraints.

Chapter 7 presents implementation of multiregional population projection for 16 administrative regions of Kazakhstan.

Chapter 8 shows multiregional population projection of four macroregions of Kazakhstan.

Chapter 1

Mathematical background

This chapters starts with introduction to the cohort-component method, which is widely used population projection method for regions of any level. In second part multistate population projections are described, which are more generalized form of projections and produce results simultaneously for different categories of population. In the third part brief introduction to stable population theory and weak ergodicity is given.

1.1 Cohort-component method

Cohort-component method has a longstanding history. Elaboration of this approach was the significant innovation in the evolution of projection methodology. A century ago the English economist Edwin Cannan (1895) first made a cohort-component forecast for England and Wales. At the beginning of the twentieth century this method found application in different countries, for example population forecast for Soviet Union made by Tarasov in 1922, for the Netherlands by Wiebols (1926), for Sweden by Wicksell (1926), for Italy by Gini (1926), for Germany by Statistisches Reichsamt (1926), for France by Sauvy, for the United States by Whelpton.

Declining fertility in the early decades of the past century and overpopulation in the case of Netherlands paved the way for increased interest in developing new methods of population forecasting. Continuing population decline in Germany was connected by Burggdorfer (1932) with “two-child system”, and in Sweden Murdal (1934) substantiated with improved contraception. These opinions accentuated inadequacies of early methods of prediction. For instance, in the first prognosis of Finland Modeen adopted Verhulst logistic model (along with simpler exponential model) because of fixedness of behavior predicting only growth or decline without suitable change in performance.

In 1945 Leslie formalized cohort-component method in mathematical terms however Bernadelli (1941) and Lewis (1942) had earlier considered the matrix formulation (Alho, Spencer 2005:228).

The whole point of the method is that initial populations for countries or regions are grouped into cohorts defined by age and sex, and the projection proceeds by updating the population of each age- and sex-specific group according to assumptions about three components of population change: fertility, mortality, and migration. Each cohort survives forward to the next age group according to assumed age-specific mortality rates. Five-year age groups (and five year time steps) are commonly used (although not strictly necessary) for long-range projections. (O'Neill et al. 2001:211).

There are could be some variations between various types of cohort-component method. The starting point is the launch-year population, population at the beginning of the projection period stratified into age-sex cohorts. Age groups can be specified as one- or five-year groups. Assume that the number of years in the projection interval is proportional to the number of years in the age cohort (e.g., five-year age groups for projections made in five- or 10-year intervals). The oldest age can defined by availability of data and due to the process of population ageing for projections it should maximum available. Many applications of the cohort-component method further subdivide the population by race and ethnicity or any other attribute. This adds to the data requirements and computation, but the logic and whole procedures would be the same.

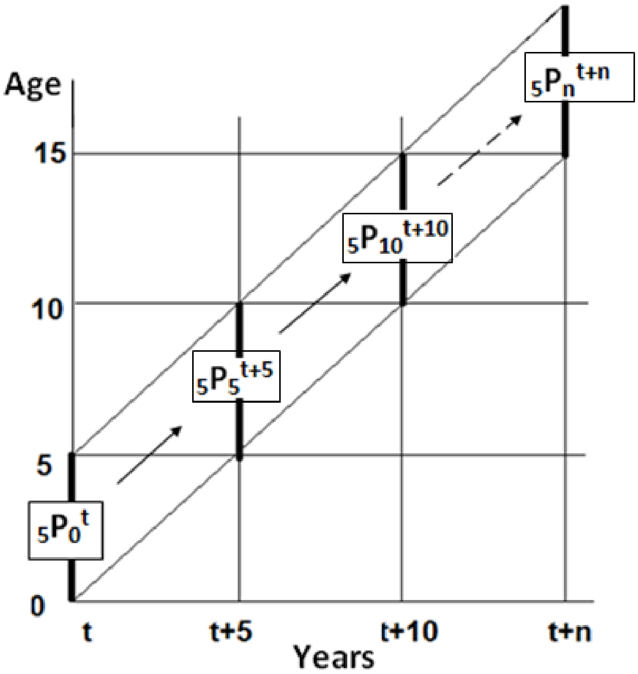
The first step in the projection is an estimation the number of persons survived to the end of the projection interval. For this purpose we multiply age-sex groups of initial population by corresponding age-sex-specific survival rates. These survival rates indicate the probability of surviving over the whole projection interval usually are estimated through the data from life table. Assumptions about future development of survival rates can be based on the extrapolation of historical trends, structural models, simulation techniques, or rates found in other areas.

In the second step estimation of age-sex specific numbers of migrants during the projection interval is performed. These rates can be based on either gross migration data separately calculated in-migrants and out-migrants or net migration data. Gross migration is closer to the true migration process than net migration, but they require more computations and more date often unavailable for small data in multiregional projections. For long projection horizons and rapidly growing areas gross migration models may provide more accurate forecasts than net migration models. Estimated number of migrants are added to or subtracted from the surviving population to provide a projection of persons born before the launch date.

In the third step, expected number of birth is calculated. This is achieved by applying age-specific fertile rates to the female population in childbearing age groups. And finally, the number of births distinguished by sex and adjusted for migration and mortality is added to the rest of the population. This provides a projection of the total population by age and sex at the end of the projection interval. This population serves as the base for projections for the following interval. The process is iterative until the final target year in the projection horizon has been reached (Smith, Tayman, Swanson 2002).

Fig. 1 presents Lexis diagram of survival probabilities for the cohort-component model.

Fig. 1 –Survival probabilities



Source: Denisenko and Kalmykova 2007

Standard period-cohort projection equation, taken from the book Denisenko and Kalmykova 2007, looks in that way

$$P_x^i(t+5) = \left[\left(P_{x-5}^i(t) + \frac{\Delta M_{x-5}^i(t,t+5)}{2}\right) * S_x^i\right] + \frac{\Delta M_x^i(t,t+5)}{2},$$

where survival probabilities S_x^i calculated according to the formula:

$$S_x^i = \frac{L_x^i}{L_{x-5}^i}$$

with adopted notation:

$P_x^i(t)$ – population aged x, gender i at time (t),

L_x^i - is the number of person-years lived between birthdays x and x+5,

i – gender with values m for males and f for females.

t – time point (start of interval)

$\Delta M_x^i(t,t+5)$ – migration population growth.

Number of projected births are calculated according to the following equation:

$$B(t,t+5) = \sum_{15}^{45} F_x * 5 * \frac{P_x^f(t) + \frac{\Delta M_x^f(t,t+5)}{2} + P_x^f(t+5)}{2}$$

Obtained births are introduced into population of the first age group:

$$P_0^i(t+5) = B^i(t,t+5) * \frac{L_0^i}{5 * l_0} + \frac{\Delta M_0^i(t,t+5)}{2}$$

where l_0 –original number of individuals in the cohort, radix of the life table.

The above-mentioned formulas are not the only way for computing projected population. There are several modifications in implementing cohort-component method.

1.2 Multistate models

Multistate demography is the study of population disaggregated into groups (states) by different characteristics, such as age, sex, ethnicity, region of residence, marital status, state of employment, number of children etc. The study of transition patterns between multiple states starts with estimation of missing data, proceeds with calculation of proper rates and corresponding probabilities and comes to the projections about coming prospective on the assumption of unchanging probabilities. Briefly, mathematical demography deals with problems of measurement and dynamics in multistate population systems.

Classical demographic techniques are generalized in the framework of multistate demographic analysis. Projections of populations stratified into multiple states can be accomplished by taking advantage of methodology of multistate projection, where central model of population dynamics represent a multistate generalization either of the continuous age-time model of Lotka (LeBras 1971) or the discrete age-time model of Leslie (Rogers 1966, 1968, 1973, Feeney 1970).

Multistate demography was pioneered by Andrei Rogers with publication in Demography in 1966 and issue of book in 1975. As a specialist in urban and regional planning, Rogers's interest was mainly in regional population dynamics and migration, and changes in multiregional populations are described by systems of simultaneous linear equations written in matrix form. Robert Schoen's research work about population stratified by marital status (Schoen 1975) extended multiregional demography into multistate demography. Philip Rees, the geographer from England, inspired by the work of Richard Stone, the economist who introduced economic and social accounting in the early 1960s, developed an accounting system for multiregional populations (Rees and Wilson 1977). Accounts including population stocks and flows have a great advantage: they must balance. Differences in data type, inconsistencies, and other data problems are easily revealed. Along with accounting approach the multistate life table could be viewed from mathematical statistics point of view as duration dependent life tables where age is considered as a duration variable. Jan M. Hoem, Michael T. Hannan, and others identified common features of the questions demographers try to answer using the life table and those addressed in the fields of survival analysis and event-history analysis with their focus on models of duration dependence (Willekens 2003).

Distribution of people over states defines the population structure, state transitions can be divided into two groups: interstate transitions, for example from being married to widowed or to being divorced, from being diseased to being healthy, or moving from one region to another, and entries from or exits to the rest of the world, as a final or starting points in the chains of transition. The multistate life table shows how the size and composition of a (synthetic) cohort change over time. Multistate projection models describe how the population structure (stock) at a given time depends on the initial population and the transitions people make (flows).

Transition rates and transition probabilities are estimated from the data. The estimation of probabilities directly from the data is complicated in the presence of censoring when individuals enter or leave the population during the period of observation for a reason unrelated to the transitions being studied. The estimation of rates does not present that problem since the transitions are related to the time spent in the origin state during the interval. In this approach, people may enter and/or leave a state during an interval. Transition rates must be converted into probabilities (Willekens 2003).

Multistate life tables can have several cohorts (radices) to interact during the process of multistate demographic evolution. There is very important Kolmogorov equation:

$$d\mathbf{l}(x)/dx = -\boldsymbol{\mu}(x)\mathbf{l}(x),$$

where $\boldsymbol{\mu}$(x) is the matrix of out-migration and occurrence-exposure rates and in finite approximation can be defined in the next form:

$$\mathbf{M}(x)=\begin{bmatrix} M_{1d}(x)+\sum M_{ij}(x) & -M_{21}(x) & .. & -M_{m1}(x) \\ -M_{12}(x) & M_{2d}(x)+\sum M_{2j}(x) & .. & -M_{m2} \\ : & : & : & : \\ -M_{1m}(x) & -M_{2m}(x) & .. & M_{md}(x)+\sum M_{mj}(x) \end{bmatrix}$$

Rogers and Ledent (1976) showed that the probability matrix **P**(x) for a 5-year interval could be calculated from matrix **M**(x) using the equation

$$\mathbf{P}(x)=\left[\mathbf{I}+\tfrac{5}{2}\mathbf{M}(x)\right]^{-1}\left[\mathbf{I}-\tfrac{5}{2}\mathbf{M}(x)\right],$$

where

$$\mathbf{P}(x)=\begin{bmatrix} p_{11}(x) & p_{21}(x) & .. & p_{m1}(x) \\ p_{12}(x) & p_{22}(x) & .. & p_{m2}(x) \\ : & : & : & : \\ p_{1m}(x) & p_{2m}(x) & .. & p_{mm}(x) \end{bmatrix},$$

with $p_{ij}(x)$ being the probability of an individual living in region *i* at exact age *x* surviving and living 5 years later in region *j* (Rogers 1995:85).

Multistate models have found applications in many areas. The first were connected with estimation of spatial distribution and calculation of how many years people (synthetic cohorts) spent living in different regions. Specially it ought to be noted the impact of Rogers. Rogers and Frans Willekens (1986) demonstrated multiregional life tables for several countries. For some regions with large migration flows that overweight the effects of fertility or mortality multiregional projections more preferable since their ability to catch migration by origin and destination.

In household and family demography multistate approach allows to go beyond the widely used headship rate method and take into account changes in the number and types of families and households in terms of the demographic events people experience and the transitions they make to new family or household types. During the analysis of marital status tables nuptiality indicators like probability of marriage transition into widowhood, the mean age at divorce, the expected duration of marriage at divorce, and the expected number of divorces in a lifetime can be obtained. Family Demography (1987), edited by John Bongaarts, Thomas Burch, and Kenneth Wachter describes of the marital careers of American women, children's experiences in different models of families, and the variability of family types created within the life span of a cohort (Willekens 2003).

It stands to mention a significant impact of multistate models to epidemiology and public health. People experience different states of health including liability to specific diseases, impairments, disability or handicaps. The life table estimates the probability that a person of a given age develops a disease during the specific period of lifetime, and depending on the availability of data calculates the probability of recovery as well as estimation of expected duration of the disease. Kenneth Manton and Eric Stallard (1988) developed multistate life tables for chronic diseases. Commenges (1999) critically noted that in some multistate models population was not disaggregated by age. Flawless is the work on the cardiovascular life course by Anna Peeters and others (2002) where multistate life used to describe a particular disease history of a cohort and improve the estimates of lifetime risk of the disease and years with the disease attributable to risk factors (Willekens 2003).

1.3 Stable population theory and weak ergodicity

Although the assumptions of asymptotic growth rate and asymptotic age-distribution are unrealistic from forecasting point of view stable population theory is still important since in contrast with exponential and logistic extrapolationg models a population may have unchanging transition rates, a positive current growth rate and negative intrinsic growth rate.

The characteristic that the asymptotic age-distribution and growth rate do not depend on the initial age-distribution is called ergodicity of the process. In case of changing regime of mortality and fertility there is no more ensured specified long term growth rate nor the age distribution that the population might tend. If the vital rates of the population change through time, we cannot expect a time-invariant age structure to exist. However, we can still ask whether the long-run behaivior of the age-distribution is independent of the initial data. This phenomenon is called weak ergodicity of the population, such description was given by Lopez, who first proved the weak ergodicity theorem for the deterministic discrete-time model. Namely, any two population vectors will become proportional if subjected to the same regimes of fertility and mortality.

Suggesting that all transition rates and fertility rates are limited from zero and above, then two multistate population systems exposed to the same sequences of transition rates and fertility rates will have asymptotically the same distribution by age, sex and region despite the common distribution will change eventually just as asymptotic growth rate (Alho, Spencer 2005:183-185).

Rogers (1966) introduced multiregional growth model expressed in matrix form

$$H = \begin{bmatrix} H_{11} & H_{21} & .. & H_{m1} \\ H_{12} & H_{22} & .. & H_{m2} \\ : & : & : & : \\ H_{1m} & H_{2m} & .. & H_{mm} \end{bmatrix}, \quad (1.1)$$

where

$$H_{ij} = \begin{bmatrix} 0 & 0 & b_{ij}(\alpha-5) & .. & b_{ij}(\beta-5) & 0 & .. & 0 \\ s_{ij}(0) & 0 & & : & & & & \\ 0 & s_{ij}(5) & & : & & & & : \\ : & : & & : & & & & 0 \\ 0 & 0 & .. & : & & & s_{ij}(z-5) & 0 \end{bmatrix}. \quad (1.2)$$

In the formulas of this section the following notation will be used:

$S_{ij}(x)$ defines the proportion of x to ($x+4$) year old residents of region i at time t who are alive and x+5 to x+9 years old 5 years later in region j at time t+1;

$b_{ij}(x)$ designates the average number of babies born during the unit time interval and alive in region j at the end of that interval, per x- to (x+4) year old resident of region i at the beginning of that interval;

$K_i^{(t)}(x)$ identifies the x- to (x+4) –year old residents of region i at time t;

α and β specify the first and the last age groups of childbearing;

z denotes the oldest age group.

Alternative form of multiregional matrix growth operator and of the latter as the generalized Leslie matrix was suggested by Feeney (1970):

$$G = \begin{bmatrix} 0 & 0 & B(\alpha-5) & .. & B(\beta-5) & 0 & .. & 0 \\ S(0) & 0 & & : & & & & \\ 0 & S(5) & & : & & & & : \\ : & : & & : & & & & 0 \\ 0 & 0 & .. & : & & & S(z-5) & 0 \end{bmatrix}, \quad (1.3)$$

Using Feeney's form multiregional projection model can be written in the form:

$$\{K^{(t+1)}\} = G\{K^{(t)}\}. \quad (1.4)$$

Summarizing, an uniregional population closed to migration and subjected to an unchanging regime of fertility and mortality will ultimately achieved a stable age composition at a constant intrinsic rate of growth. Rogers (1975) using the similar techniques proposed earlier by LeBras (1971) and Feeney (1971) generalized Sykes' reasons to the external multiregional population systems (Rogers 1995).

Chapter 2

Approaches to forecasting of demographic components

2.1 Modeling and forecasting mortality

In comparison with fertility and migration forecasts in mortality prognosis we know that all people eventually die, the questions arise when and from what causes. Epidemiological transition distinguishes five phases in the evolution of mortality. The first phase, the age of pestilence and famine, is characterized by high and fluctuating mortality with life expectancy at birth about 20-40 years suppressed by epidemics, famine and wars. In the second phase, the age of receding pandemics, mortality declines gradually and population growth begins to describe an exponential curve with life expectancy at birth about 30-50 years. In the third phase, the age of degenerative and man-made diseases, mortality continues to decline and approaches stability at a low level, and can be characterized by infant mortality decline, longevity of generations, decline of infectious and rise of degenerative diseases. In the fourth phase of delayed degenerative diseases probability of death from these causes shifted towards advanced ages, namely cardiovascular disease mortality declined due to several preventive strategies. The fifth phase of the epidemiologic transition, the age of obesity and inactivity, is characterized by increased proportion of overweight with following consequences such as increased risk of coronary heart disease, ischemic stroke, hypertension, joint disease, cancer, sleep apnea, asthma, and a host of other chronic conditions. Such striking changes in mortality profile issued many works concerning what could be future development of mortality.

Extrapolation is one way often used in forecasting specially for short-term prediction. In mortality forecast it could be extrapolation of life expectancy or any other life-table measure, another way is to use empirically based model life tables in order to get the age pattern; this has been facilitated by the expansion of life tables to include older ages (Coale, Demeny, and Vaughan,1983) and lower mortality (Coale and Guo, 1989). The independent extrapolation of age-specific rates usually entails mortality reduction factors or some fraction of the reduction factor (Goss, Wade, Bell, and Dussault, 1998; Pollard, 1987).

Relational Brass logit-life table system (Brass 1968) overcomes shortcomings of model life table. On the one hand, it reflect the patterns found in empirical mortality, on the other hand it's not constrained to represent exclusively the patterns these data embody for. Brass attempted to relate mathematically two different life tables since he discovered that a certain transformation of the probabilities of survival made the relationship between corresponding probabilities for different life tables approximately linear.

Golulapati, De Ravin, and Trickett (1984) applied relation logit life table model to forecast Australian male cohorts (Pollard 1987), and Keyfitz (1991) used it for Canadian data. Flexibility was increased in four-parameter models in Zaba (1979) and Ewbank, Go´mez de Leo´n, and Stoto (1983). The Zaba model adopted by Congdon (1993) found next application in forecast the relatively stable parameters by univariate ARIMA models. Hannerz (2001c) integrated the features of relational models with parameterization functions and model life tables in a regression model (Booth 2006).

Parameterization functions where mortality age patterns are functions of age widely used in forecasting as well as in smoothing data, reducing errors, creating life tables, drawing inferences from incomplete data etc.

In 1825 Gompertz suggested his model of mortality law, where force of mortality increases with age since resistance to death decreases exponentially with age or shrinks like shagreen leather constantly with new birth day. Makeham in 1860 improved the Gompertz model by adding background mortality constant parameter with purpose to overcome of underestimation of actual mortality at youngest adult ages.

In order to overcome overestimation of mortality at the oldest age the logistic model was suggested with good fitting to mortality rates over the entire adult age range with comparatively few parameters (Thatcher 1999, Thatther, Kannisto 1998).

Small overestimation of mortality between ages 60 and 80 and underestimation of women mortality at the highest ages in some countries conform with findings of Himes, Preston, Condran (1994). New version of logistic model, shifting logistic, was proposed by Bongaarts where instead of interpreting mortality as rising or falling the schedule of the force of senescent mortality can be viewed as shifting to higher or lower ages over time. Background mortality is invariant over time and negligible small at high ages, while senescent mortality increases linearly from age 25 to about age 75 and after age 75 approaches 1.0. Since the slope parameter is nearly constant over time, the changes of senescent mortality can be described with only one varying parameter (either the level or shift). In comparison with Lee-Carter method shifting logistic in long-run is better due to weakness of assumption in Lee-Carter method about constancy in mortality declining of age-specific rates (Bongaarts 2005).

In 1827 Theil proposed three-component model for childhood, adulthood and old age. The second term of Theil's model was changed with parabolic function by Mode and Busby in 1982, while Rogers and Plank in 1984 changed it with double exponential curve. In 1980 Heligman and Pollard proposed main model with eight parameters and three modifications of original function. In 1992 Kostaki replaced the middle term of Helligman-Pollard model function by two parameters related to the spread of the accident hump to the left and right of its peak. In 1994, Rogers and Little proposed their multiexponential model consisting of five components and a total of 13 parameters representing in reduced forms simple unimodal curves, ushaped curves, or more complicated bimodal curves with exponentially increasing or decreasing components (Tabeau 2001).

Wilmoth (1993) developed a weighted SVD solution to the Lee–Carter model, providing a good fit without the need to adjust the level parameter, and a maximum likelihood solution; these methods gave almost identical results to the unweighted SVD in the case of Japan (Wilmoth, 1996). Carter and Lee (1992) addressed divergence by sex by estimating a joint level parameter while retaining sex-specific age effects. Lee and Nault (1993) jointly forecast provincial mortality. Lee (2000a) discusses these and other extensions of the method; data reconstruction for years where only total deaths are available is also possible (Lee and Rofman, 1994). Tuljapurkar, Li, and Boe (2000) applied the method (without adjustment of the level parameter) to the G7 countries, finding a common pattern of linear decline in the levelparameter. Lundstrom and Qvist (2004) used the method to examine changing trends in the Swedish mortality decline during the twentieth century.

The stability of the Lee–Carter method to structural change and initial conditions was examined by Carter (1996, 2000) and Carter and Prskawetz (2001). Tuljapurkar (2005) further demonstrated the robustness of the method. Li and Chan (2005) proposed an outlier-adjusted method. Lee and Miller (2001) noted the influence of the adjustment procedure on forecast bias. Three modifications were introduced: the fitting period was restricted to post-1950 to reduce structural shifts, adjustment of the level parameter was by matching life expectancy, and observed rates were used as jump-off rates. Booth, Maindonald, and Smith (2002) also modified the method after finding historical departures from linearity in the Australian mortality decline: they proposed a method for determining the optimum fitting period for use in shorter-term forecasting when the relatively recent trend is linear, and adjusted the level parameter by fitting to the age distribution of deaths (a conditional maximum likelihood procedure).

The Lee–Miller variant has been widely adopted as the standard Lee–Carter method. Recent developments extend the applicability of the Lee–Carter method. Li, Lee, and Tuljapurkar (2004) demonstrate how, by assuming a linear trend in the level parameter, the method can be applied to populations with limited data at unequal time intervals. Li and Lee (2005) develop an augmented common factor method for overcoming the divergence problem, using a common factor to model group mortality and an additive population-specific factor. Such approaches make use of demographic convergence of mean levels; Edwards and Tuljapurkar (2005) note that substantial differences in variances should also be taken into account (Booth 2006).

The Lee–Carter method has close similarities to the principal components approach used by Bell and Monsell (1991), and Bell (1997) discusses the similarities and differences in detail, demonstrating the importance of bias adjustment and the superiority in short-term forecasts of Lee–Carter over both Heligman–Pollard and principal components using all components. Whereas the Lee–Carter method uses only the first component, the principal components approach typically uses several, thereby allowing for greater flexibility in forecasting change. Higher order terms in the Lee–Carter method were modelled by Booth, Maindonald, and Smith (2001, 2002) and modeled and forecasted using univariate ARIMA processes by Renshaw and Haberman (2003a).

Hyndman and Ullah proposed their method as a generalization of Lee-Carter model in application to forecast French mortality and Australian fertility rates. Similar to approaches of Bozik and Bell and Bell and Monsell this method implicated using of principal component decomposition of the mortality or fertility rates and in contrast functional data paradigm was used as well as robust version of principal components to avoid difficulties with outlying years. On the first step constrained and weighted penalized regression was applied to observed data. On the second step decomposition of fitted curves using orthonormal basis was applied. The next steps include fitting univariate time series to get coefficients, construct forecasts and forecast intervals (Hyndman and Ullah 2006).

The method described above was further developed in stochastic population forecast using functional data models for mortality, fertility and migration to predict Australian population 20 year ahead using data for the period 1921-2004. Functional data models with time series coefficients are applied to model age-specific mortality, fertility rates and derived net migration numbers. Then the three models are used in a Monte-Carlo simulation of future fertility, mortality and net migration, which are combined using cohort-component method. The distribution of the forecasts provided probabilistic prediction intervals (Hyndman Booth 2008).

In the modified Lee-Carter model proposed by Wolf in 2004 a first difference specification integrated estimation of the Lee–Carter and time series models, while De Jong and Tickle in 2006 generalized the Lee–Carter method by introducing a state space framework combining model estimation and forecasting through of applying B-splines to build in the expected smooth behavior of mortality over age (Booth 2006).

Integrated estimation and forecasting is a characteristics of modeling within the GLM framework. Renshaw, Haberman, and Hatzoupoulos (1996) proposed a two-factor model with two multiplicative terms: a Gompertz–Makeham graduation term and an age-specific trend adjustment term. This model was used to forecast UK mortality at ages 65+ with qualified success: the optimum fitted model parameters did not necessarily generate plausible forecasts, for which lower-order polynomials are often required (Sithole, Haberman, and Verrall, 2000). This study included a comparison with the standard actuarial practice of fitting the Gompertz–Makeham class of functions. Currie, Durban, and Eilers (2004) employed bivariate penalized B-splines to smooth over both age and time within a penalized GLM framework with extrapolation of the fitted surface over time; comparison with Lee–Carter revealed a much slower mortality decline (Booth 2006).

In modeling mortality reduction factors using GLM, Renshaw and Haberman (2000) identified the conditions under which the underlying structures of the GLM and Lee–Carter models are identical; they later demonstrated the use of the Lee–Carter methodology for forecasting the reduction factors. Renshaw and Haberman (2003c) developed a GLM-based approach that parallels the Lee–Carter method, including matching observed and expected total deaths. The important difference between the two approaches is in the treatment of time: in the Lee–Carter method time is a factor estimated by SVD, while under the GLM approach time is a known covariate. The GLM approach is based on a heteroscedastic Poisson (non-additive) error structure. Brouhns, Denuit, and Vermunt (2002) proposed a similar bilinear approach in which the Lee–Carter model forms the systematic component (predictor) in the Poisson error setting. Renshaw and Haberman (2003a) compare the Lee–Carter, linear, and bilinear approaches with and without age-specific enhancement: in the Lee–Carter case such enhancement is achieved by including the second term, in the GLM case it involves a break point or hinge to allow for greater emphasis on recent trends, and in the bilinear case the two-term Lee–Carter model is implemented as a double bilinear predictor (Booth 2006).

Forecasts based on cohort models are relatively few because of heavy data demands; where data are available the model may depend on the (inappropriate) experience of cohorts born in the nineteenth century if the entire age range is considered (Tabeau et al., 2001). This problem is reduced when only adult mortality is of interest. The cohort approach is free of tempo distortions (caused by changes in timing). Bongaarts and Feeney (2002, 2003, 2005) propose an adjustment for tempo distortions in period life expectancy, with implications for forecasting. Other aggregate measures of mortality may be considered. In developing countries, restricted time series of observations limit the application of most forecasting methods. Lutz, Sanderson, Scherbov, and Goujon (1996) overcome this problem by deriving target life expectancy as the average expectation of experts (Booth 2006).

Girosi and King introduced class of statistical methods for forecasting population death rates using Bayesian hierarchical analysis. They run a set of linear regressions with including covariates for time-series cross-sectional analysis and extend this approach to group continuous variables like age group and other spatial varying variables. They noted that ‘the most common Bayesian method of partially pooling multiple coefficients in cross-sections thought to be similar is often inappropriate as it frequently misrepresents prior qualitative knowledge’. For practical implementation Markov Chain Monte Carlo Algorithm was compared with faster estimation procedure not relying on Gibbs sampling and performance was demonstrated using age-sex-country-specific mortality data (Girosi King 2006).

Regression models are easily extended to three factors, but (as noted above) age–period–cohort (APC) models must accommodate the identification problem (see also Van Hoorn and De Beer, 2001). To address this, Wilmoth (1990, 2001) developed a modified model involving additive age and period effects and several multiplicative interaction terms. Tabeau (2001) concluded that mortality forecasting based on APC models is not feasible because of the difficulty in assuming future period effects (although age and cohort effects can be assumed to be fixed); only in forecasts of specific diseases would sufficient epidemiological knowledge be available. Caselli (1996, 2002) used the APC model to forecast mortality from leading causes (Booth 2006).

Forecasting by cause of death has been advocated from a theoretical perspective as a means of gaining accuracy (e.g., Crimmins, 1981), but experience has largely proved otherwise. Little is gained from decomposition because of similar age patterns in the main causes; cause-of-death reporting is unreliable at older ages where most deaths occur; and cause reduction may have minimal effect on total mortality (Murphy, 1995). Further, model misspecification and the presence of leading indicators (where changes in one cause systematically precede changes in another) can result in reduced accuracy from decomposition (Alho, 1991). The short time series of cause-of-death data also limits extrapolation. Using the multiexponential model, McNown and Rogers (1992) found no consistent discernible gain in accuracy from cause-of-death decomposition. Wilmoth (1995a) demonstrated that, for proportional rates of change models, mortality forecasts based on the sum of cause-specific forecasts will always be higher than those based on aggregate data because causes of death that are slow to decline come to dominate as other causes are more rapidly diminished. Using APC models for ages 60+, Caselli (1996) found this to be true for females but reversed for males. Tabeau et al. (2001) also found this difference between the sexes for France, Italy and the Netherlands, but not for Norway (Booth 2006).

Mortality forecasting based on (partial) cause elimination and cause-delay models make use of targeting and informed judgment (Manton, Patrick, and Stallard, 1980; Olshansky, 1987, 1988); Kunst, Mackenbach, Lautenbach, Oei, and Bijlsma (2002) incorporated competing causes of death. These methods have often led to conservative forecasts of mortality reduction. Le Bras (2005) elaborates a cause-delay model of mortality change. Gutterman and Vanderhoof (1998) argue the case for structural models of cause-specific mortality change that take medical and other factors into account, despite the difficulties involved. Structural models of mortality at older ages relate lifestyle and other risk factors to functional status and mortality using vector autoregression, achieving some improvement over traditional time series and informed judgment methods (Manton, Stallard, and Tolley, 1991; Manton, Stallard, and Singer, 1992). However, their forecasting potential is limited by the short time series of risk factors, the large number of parameters and the non-linear interactions generating the mortality forecast. Epidemiological, structural and multistate approaches to cause-of-death forecasting are reviewed by Van Den Berg Jeths, Hoogenveen, De Hollander, and Tabeau (2001).

AIDS mortality attracts attention of forecasters even complicated due to lack of available data. An early method involves the extrapolation of AIDS cases, while another uses a model of the progression from HIV infection to the onset of AIDS to back-calculate HIV infections, which are then predicted.

2.2 Modeling and forecasting fertility

Fertility rates and births are non-stationary series. A difficulty in fertility forecasting arises from structural change, seen in the trajectory of total fertility, changing age patterns and the complex association between the two. Forecasting success has been limited. Zero-factor models are relatively common in fertility forecasting.

Early forecasts focused on events. McDonald (1979, 1981) used time series methods to forecast total births and first marital births, easily outperforming economic–demographic structural models in the very short term. Improvements were achieved by incorporating transfer functions linking total births to females of childbearing age and first nuptial confinements to marriages (thus, in effect, forecasting rates).

Forecasting fertility rates more beneficial over deriving births since the number of women is known approximately for the first 15 years. Miller (1986) forecast total fertility and the mean age at childbearing by a transfer function model relating past trajectory of total fertility to changing age patterns. Age-specific fertility rates have been forecasted by Congdon (1980, 1989) using regressions and ARIMA models incorporating periodic time and relative cohort size (in line with the Easterlin hypothesis), and by McDonald (1983) using simple time series models, with greater success than structural models (Booth 2006).

In comparison with univariate models and vector autoregression Ortega and Poncela achieved more accurate results in long-run forecast by jointly modeling total fertility trends for a (subjectively-defined) homogeneous group of countries. They use dynamic factor models to estimate one or two common factors capturing the non-stationary average total fertility trajectory and a stationary deviation from the average; they forecast these factors using time series methods.

The approach of Lutz, Sanderson, Scherbov, and Goujon (1996) implicitly deals with the problem of past and potential future structural change by deriving target total fertility as the average expectation of a group of experts. Another improvement of this method includes disaggregation by education. Ahlburg (1982) achieved greater short-term accuracy for total births than US official forecasts; his model included marriage, divorce and female labor force participation. Ahlburg also forecast US births using a simple Easterlin relative cohort size model, identifying cycles of alternate generations (Ahlburg, 1983), and forecast Canadian births using a similar model based on births both one and two generations ago (Ahlburg, 1986). Structural modeling of age-specific fertility rates implicates separate modeling by age but estimation as 'seemingly' unrelated regressions.

Ermisch (1983) modeled three age groups in this way as functions of women's and men's earnings, relative cohort size and female lifetime employment rates, and forecast them. Poorer results for women aged 30–34 than 20–24 were attributed to heterogeneity in fertility responses to economic change by parity with, the changing parity distribution of women producing unstable responses.

Ermisch (1992) supposed a model of fertility rates by birth order; short-term as well as long-term forecasts based on this model were not sensitive to assumptions about the explanatory; the main advantage was not in forecasting but in exploring different scenarios for policy purposes.

Several parametrizations have been applied in the modelling of age-specific fertility rates, including the beta, gamma and Hadwiger functions. Hoem et al. (1981) compared several functions, finding the gamma density and the Coale–Trussell function to be equally superior except for highly parametrized splines; they noted the Coale–Trussell advantage of parameter interpretability for forecasting. Rogers (1986) considered the Coale–Trussell model to be excessively complex and suggested the direct use of the double exponential function; in other words, the third term of the multi-exponential function.

The double exponential fitted better to relatively symmetrical fertility patterns, and less well to the flatter curves of the 1980s (Knudsen, McNown, & Rogers, 1993), whereas the Coale–Trussell model was superior in cases of higher fertility in the 1960s (Rogers, 1986:51). The four parameters of the double exponential are not readily interpretable; in the absence of time series data, Rogers (1986) regressed each on the gross reproduction rate (GRR) and assumed future GRR in forecasting. Knudsen et al. (1993) partially reparametrized the model, the new parameters being the mode and modal value, and related total fertility to the remaining two parameters (via the gamma density), considerably improving forecastability; univariate ARIMA models performed well in out-of sample forecasts of US fertility, despite strong interactions among parameters.

The more tractable relationship between the three pattern parameters of the gamma density and the mean and variance of childbearing was made use of by Thompson, Bell, Long, and Miller (1989) for short term forecasting of age-specific fertility; they directly forecast the level, mean and standard deviation of childbearing using a vector autoregression.

The gamma density was also used by Keilman and Pham (2000) in long-term interval forecasts of Norwegian fertility involving a vector autoregression of three of the four parameters (governing level and age pattern, the parameter representing the minimum age of childbearing being constant).

Congdon (1990) used the Hadwiger function, forecasting the four parameters using univariate ARIMA models; with hindsight, this compared favorably with a structural time series model incorporating cyclical and trend factors, relative cohort size and female job opportunities. Congdon (1993) made a similar comparison using the reduced form Hadwiger function for both period and cohort fertility. He noted that all four-parameter functions are over-parameterized; of the reduced-forms, the beta gave a better fit but the Hadwiger has the advantage of parameter interpretability. Chandola, Coleman, and Hiorns (1999, 2000) used the reduced Hadwiger function to model European fertility: while it provided a good fit for several countries, for others it was unable to capture the slight hump at young ages that has recently developed. For these countries, a mixture model was used to combine Hadwiger functions for non-marital and marital fertility, making use of disaggregation (Booth 2006).

The general approach has potential for forecasting. Brass (1974, 1981) developed the relational Gompertz model which linearly relates observed fertility to a suitable standard. The model is used with incomplete cohort data, or to produce series of the level and two pattern parameters for forecasting. Parameter interpretation was improved by Zeng, Zhenglian, Zhongdong, and Chunjun (2000) in relating the pattern parameters to the median age and interquartile range. Murphy (1982) investigated the use of relational Gompertz models in forecasting. In general, structural change limits the use of parameterization functions and relational models for forecasting, especially where (as is desirable) vector autoregressions are employed.

The Coale and McNeil double exponential has been widely shown to fit first births well (Bloom & Trussell, 1984). Bloom (1982) applied this approach to forecast first births for incomplete cohorts, with limited success for younger cohorts. Trussell and Bloom (1983) allowed the parameters to depend on covariates and then forecast childlessness for incomplete cohorts. The Coale–McNeil function was elaborated by Kaneko (2003) as the generalized log gamma distribution and used to forecast first marriages and parity-specific fertility for incomplete cohorts.

Evans (1986) used linear regression to predict first birth fertility after age 25 from the proportion attaining parenthood by age 25 (quantum) and the ratio of fertility at 15–19 and 20–24 (tempo); overall fertility was similarly forecast. Martinelle (1993) forecast first birth rates and childlessness using a regression model of incomplete cohort fertility that tooks education into account. Chen and Morgan (1991) and Morgan and Chen (1992) showed that the Bloom approach was sensitive to censoring below age 30, while the Evans approach was in fact based on period effects, and concluded that it was preferable to base forecasts on the period life table model which assumes that current rates will persist into the future.

Ryder (1990) also advocated a period approach, incorporating continued patterns of change noting that the dominant effect on fertility change in the short- to medium-term is period rather than cohort. Bongaarts and Feeney (1998, 2005) proposed an adjustment of period total fertility to take tempo effects into account when data by parity are available. Kohler and Philipov (2001) extended this adjustment to include variance effects. While unadjusted fertility remains the basis of population forecasting, adjusted measures aid understanding and inform future trends. Kohler and Ortega (2002) proposed a tempo adjusted period parity progression measure that can be used to forecast the fertility of incomplete cohorts conditional on a level of fertility and a postponement pattern derived from past period trends. Sobotka (2005) used this method to derive tempo-adjusted first birth probabilities, which he used as the low assumption contrasted with the high assumption based on unadjusted probabilities, separating tempo and quantum effects (Booth 2006).

Lee (1992, 1993) modeled age-specific fertility rates over time using a single time varying fertility index (the method parallels the Lee–Carter method for forecasting mortality) by imposing change, lower and upper bounds and an ultimate (average) level. Lee and Tuljapurkar (1994) used this model with a different ultimate level and no bounds. Carter and Lee (1986) used the approach in a joint model of nuptiality (age and period) and marital fertility (duration and period).

The Lee method is a principal components method. Principal components methods were used by Bozik and Bell (1987) to forecast age-specific fertility, using the first four components and the level in a vector autoregression. A principal components approach was also used by Sivamurthy (1987). Bell (1992) discusses the use of principal components and various other models in time series forecasting of age-specific rates.

Cohort forecasting of fertility makes lesser demands on data than mortality, but may be compromised by structural change. Li and Wu (2003) modeled fertility for completed cohorts by age and cohort using the Lee (1993) model, and combined the estimated fixed age effect and incomplete cohort observations to forecast the cohort effect, thereby completing that cohort's fertility. The method is restricted to completing cohort fertility when certain assumptions are met.

De Beer (1985) developed the CARIMA (cohort-ARIMA) model for short-term forecasting which was successful in identifying turning points 6 or 7 years ahead in first and second order births. The model forecasts age- or duration-specific fertility rates for cohorts using time series methods, taking into account error covariances and additive period effects, both of which are modeled as ARIMA processes. The unobserved fertility of incomplete cohorts is forecast on the basis of observations at younger ages and for older cohorts. As with ARIMA models in general, a disadvantage of the CARIMA model is the difficulty in interpreting its parameters. Using post-1950 data for four European countries, De Beer (1989) compared four models: the CARIMA model, a multiplicative APC-ARIMA model (Willekens & Baydar, 1986), cubic spline models of age-specific fertility rates and an ARIMA forecast oftotal fertility. An advantage of the APC-ARIMA model was its greater parameter interpretability, but its usefulness is strictly limited to short-term forecasts. The short-term CARIMA and APC-ARIMA forecasts performed slightly better than the ARIMA forecast of total fertility, while the CARIMA forecasts were more accurate than those from the APCARIMA model when parity-specific rates were used (except anomalously for second births). The relative accuracy of the spline and CARIMA models was highly dependent on the position of turning points in the fitting period: either model could produce large forecast errors (Booth 2006).

Birth expectations might be regarded as a potentially useful exogenous variable in cohort fertility forecasting, particularly in forewarning of changing trends, but the expectations data relatively unstable over time since they also tend to lag rather than lead to actual fertility. Poor correspondence between stated expectations and later births were revealed by using record linkage and longitudinal surveys. At the aggregate level due to effect of compensating discrepancies expectations data are generally regarded as unreliable. Minor fluctuations in desired completed family size will generate major fluctuations in period fertility resulted in annual births and consequent age structures.

Direct use of birth expectations is not favorably for forecasting, as it happened for US official forecasts since discrepancy between expectations and realizations requires consideration of period and involvement of informed judgment, as an example solution could be constructing of bivariate econometric model of children ever born and additional expected births (Booth 2006).

2.3 Modeling and forecasting migration

Rogers, Raquillet and Castro (1977) examined mortality and fertility approaches for capturing regularities exhibited by empirical migration schedules. Observed regularities in migration schedules were expressed in equation with 11parametrs including pre-labor, labor and post-labor curves. Like in mortality analysis where deaths are decomposed by causes, in the next work Rogers and Castro examined how reasons or causes affect the level and age profiles of migration schedules (Rogers and Castro 1979).

Rogers, Willekens and Raymer (2001) addressed the question how to formally represent the spatial structure of an observed origin-destination-specific patterns of interregional migration flows, this problem can be viewed as modeling of multidimensional contingency table. The researchers found that interregional flows exhibit strong regularities over time and can be captured by generalized linear models, that can be useful in situations when data are inadequate or missing to indirectly estimate interregional migration patterns. Origin effects result in differences in the shapes of migration, whereas time effects tend to represent differences in the levels of migration. The log-linear models and logit models decomposed observed patterns of migration to identify the separate effects of all variables considered, namely region of origin, region of destination, time period and age, and their interactions. Summarizing, two principal components, generation and distribution, identify spatial structure of interregional migration patterns over time and then can be used to impose that particular structure onto a different migration setting (Rogers, Willekens and Raymer 2001).

Schoen and Jonsson (2003) compared Iterative Proportional Fitting (IPF) approach with Relative State Attraction (RSA) approach in order to estimate the U.S. interregional migration rates for the period 1980-1990 and showed that both approaches produce similar results. Iterative proportional fitting approach based on entropy maximization provided entropy corresponds to the amount of randomness or lack of structure in the data (Willekens 1999). The maximum entropy solution finds the pattern of flows achievable in the greatest number of ways (Halli and Rao 1992:190). Willekens (1982) showed that IPF is equivalent to estimating an array by log linear modeling, where higher order interactions in the model are ignored. RSA method based on changes in the attraction/repulsion of states influencing the risk of movement and estimates interstate transfer rates from cross-sectional population distributions and an assumed set of standard rates. A basic characteristic of the RSA method is that the product of the transfer rates between two states is the same in both the assumed standard and the resultant estimates. Therefore, relative to the standard chosen, if the rates of transfer from one state to the other increase, then the rates in the opposite direction are assumed to decrease. The comparison showed that RSA yields excellent estimates when there are large, compensating changes in interstate rates. When the rates between two states move in the same direction, the estimates are more in error, but nonetheless preserve the age pattern of behavior in the rates and generally yield age-aggregated summary measures close to actual levels (Schoen Jonsson 2003).

Four groups of summary indices were suggested to overcome these problems at the national level (Bell et al. 2002). The first group, measures of the intensity of migration, evaluates the overall level of mobility within a country and consists of crude migration probability, standardized migration probability, gross migraproduction rate and migration expectancy. The second group, measures of distances of migration, summarizes the effects of distance across the entire migration system and include median distance moved and distance decay parameter, the latter can be obtained by fitting spatial interaction models to matrices of interzonal flows calibrated by using either the entropy maximizing method (Wilson, 1970) or the Poisson regression method (Flowerdew 1991). The third group, measures of migration connectivity, is composed of index of migration connectivity, index of migration inequality, Gini index as a measure of concentration (Duncan and Duncan, 1955; White,1986) and coefficient of variation as measure of spatial focusing proposed by Rogers and Raymer (1998). The fourth group, measures of migration impact, shows the overall effect of migration in redistribution a population across the entire system of regions and consists of the migration effectiveness index and the aggregate net migration rate. The above mentioned indicators were calculated to compare internal migration in Australia and Britain (Bell et al. 2002).

Stillwell (2005) overviewed variety of approaches in modelling of interregional migration. Firstly, the accent in the work was made on the distinction between micro or macro approaches, the rest of work was focused on the macro approaches. Another distinction was noted between causal factors or variables that determine migration (such as marriage or job opportunities) and those factors that have a selective influence on migration (such as age, sex or social class). Then distinction was made between mathematical and statistical calibration techniques in the macro migration modelling. It was pointed on the benefits of the use of the general linear modelling approach in fitting explanatory models of migration and the application of the Poisson model in modelling sub-national migration to explain the relative effects of exogenous explanatory variables on out- or in-migration, and on origin-destination migration. A two-stage migration model based on spatial interaction principles and calibrated using statistical regression developed for use in a policy context in the UK was contrasted with models developed in the context of multi-state demography and used for migration projection in the European Union. It was concluded, that internal migration is influenced by various determinants and experiences historical dependence, and good explanatory model of migration probabilities can be less effective in a projection context in comparison with a model based on historical flows, because of the inadequacies of the projection of the independent variables (Stillwell 2005).

A more detailed description of the above mentioned two-stage model MIGMOD including policy-sensitive explanatory variables can be found in van Wissen (2002). In this work the predictive performance of demographic multiregional model was compared with the extended economic-geographical models using migration data of European countries, the Netherlands, the UK and Sweden, at the NUTS 2 level. The models were nested within a GLM specification permitting both demographic and extended models to be written as specific cases of log-linear models. The explanatory models did not considered prediction of exogenous variables and used observed values, whereas the drawback of demographic model was the absence of causal “drives” producing changes over time in the migration process. The results showed in the short run destination patterns due to their stability can be predicted using historical patterns, and demographic models shows good results. Outmigration patterns were underestimated by both models for the UK, and Sweden, and were satisfactory for the Netherlands due to their stability for this country. Overall migration level probably could be estimated by taking into account business cycles (van Wissen 2002).

Wilson (2010) proposed the way how to include significant age-concentrated student migration into standard parameterized model migration schedule. The example was demonstrated using female in- and out-migration models with and without student peak for two regions of Australia.

In assumptions concerning international migration the informed judgment plays central role and extrapolation sometimes joins the game, while theories of international migration have not often been quantified in forecasting (Howe and Jackson, 2005).

De Beer (1997) proposed time series methods to model aggregate immigration and emigration for Netherlands over the period 1960-1994 with ARIMA (1,0,0)-model, while net migration was modeled by ARIMA (0,0,1)-model. The three models showed consistent results, but informed judgment considered the net forecast to be too low. Informed judgment was incorporated as the target in a five parameter ‘extrapolation-target’ model, parameters can be derived from fitting the model to observations. Three parameters control the extent to which the forecast depends on past observations, the fourth is the target and the fifth is the speed at which the forecast approaches the target.

Keilman and Pham (2004) produced forecast of immigration and emigration for Norway using two time series models estimated on the basis of gross flow data for the period 1958-1997, the ARIMA models were (1,0,1) for immigration and (0,1,0) for emigration. These time series models were adjusted in such a way that they predicted the same migration flows as those used by Statistics Norway in its official population forecast. Next, both flows were broken down by sex, using male/female shares as observed in recent years. Finally, each of the four flows was broken down into one-year age groups on the basis of age-specific shares obtained by means of simple extrapolations of the parameters of Rogers-Castro age schedules as estimated for the years 1967-1997.

In forecasting net migration for California internal migration was modeled using AR(1) model as well as legal part of international migration, adjustments were made for illegal migration while the age distributions were assumed to be fixed (Miller 2002).

The recent work of Bijak (2010) presented an application of Bayesian time series analysis in forecasting international migration. The Bayesian paradigm ensures the formality of inference, while allowing to include the a priori expert judgment in the analysis, alongside with the observations. Hence, the former can supplement the data-based information for small samples characterizing many time series of within European migration. The explanations are supported by the example of forecasts of both-way migration flows between Poland and Germany for 2005–2015, based on the aggregate data series from German population registers. The analysis covers three sets of forecasting models: simple stochastic processes – sub-models of ARMA(1,1), extensions of an AR(1) model to simplest cases with non-constant conditional variance, as well as propositions assuming a linear analogy to post-accession migratory developments in countries that joined the European Community earlier (Portugal and Spain). In each case, the outcome of the formal model selection in the Bayesian framework allows for the identification of models supported by the data at hand. The Bayesian framework also enables to interpret the results with respect to uncertainty of the forecasted phenomena in coherent, probabilistic terms (Bijak 2010).

Chapter 3

Multiregional population forecast

3.1 Approaches to modelling a system of regions

Approaches to modelling a system of regions differ in their ability to address the three following issues: first, how they expose differences and disparities among regions; second, how they handle interconnections and interdependences among regions; third, how they consistent with national projections, when the value of each variable for the nation must equal the sum of regional values or must be a weighted average (Willekens 1983).

3.1.1 Top-down approach

This approach is commonly used since it ensures consistency. The procedure is simple: first, national variables estimated first and then distributed among regions on the basis of a predefined allocation procedure. In the simplest case allocation procedure is a fixed-ratio technique. Consistency of the approach is founded on the argument that changes on the regional level originated at the national level, or in other words, what happens in a region is very much determined by what happens in the nation. From economic point of view, regional growth is driven by 'basic' sector producing for the national or international market. In reality this principle does not always work. Shortcoming is obvious: regional differences are disregarded. National projections are produced by assuming 'average' values of model parameters. Regional differences are considered only on the second stage during the distribution of national projections (Willekens 1983).

3.1.2 Bottom-up approach

This approach projects regional variables independent of national control totals. Each region is considered as a separate sub-system without explicit links with other regions. This 'uniregional' perspective gives maximum weight to regional differences and can produce inconsistent results. Independence of regions does not mean closeness. Interactions with other regions of the same system may be accounted for through variables representing net exchange implemented in cohort-survival model (Willekens 1983).

3.1.3 Hybrid approach

This approach attempts to combine consistency of the top-down approach and regional differentiation of bottom-up approach by introducing sum-constrained national totals. The predetermined national totals are regionally independent, while the other totals, which are obtained by summation the regional variables, are regionally dependent. Two-sided linkage between top-down and bottom-up approaches can be explained by argument, that changes at the national level cause changes in regions, at the same time changes in regions can spread and extend to the other regions. According to consistency requirement the net migration rate for nation without international migration should be zero. To ensure the consistency, the concept of migration pool model was introduced. The number of out-migrants for every region is computed by applying constant region-specific out-migrants rates. The national totals of out-migrants are obtained by summation. Then numbers of in-migrants from pool are distributed to destinations. Due to relationship between housing construction and in-migration it is reasonable to generate first in-migration pool and subsequently derive out-migration using distribution function (Willekens 1983).

3.1.4 Multiregional approach

In multiregional approach the concern for consistency and regional differences is augmented by a concern for representing and projecting interregional dependences. An essential element of multiregional models is a spatial interaction model. While spatial interaction analysis may formally be defined as the study of observed interregional flows from incomplete data. The dominant characteristic of multiregional models is that they study and project all regions of multiregional system simultaneously. The simultaneous solution of all region variables not only assures internal consistency, but, at the same time exposes regional differences and projects interregional flows (Willekens 1983).

3.2. Subnational population projections as official multiregional population projections

Cohort component method and is refined form – multistate models dominate the field of subnational forecasting according to survey conducted by Kupiszewski and Kupiszewska (2003), while some countries incorporate demographic variables such as nationality and ethnicity, and non-demographic variables such as labour market, school-supply, housing market.

Most of regional population projections in European Union member states appear regularly, for example every year in Austria and every five year in Italy. Number of spatial units varies from 7 for Portugal to 448 municipalities for Finland depending on the chosen level of NUTS classification. According to Tab.1 most of subnational projections are consistent with national projections using bottom up as well as top down approaches (Kupiszewski and Kupiszewska 2003).

For fertility age-specific fertility rates often used as popular measures, as well as occurrence- exposure rates can be used or birth parity dimension. All countries use time series to determine the characteristics of fertility however length of time series vary from three to 49 years. Age-specific mortality rates are widely used for the measurement of morality, as well as probabilities, regional life expectancy, survival rates or occurrence-exposure rates. Most of countries apply smoothing, some use aggregation, and all countries use time series of different length (Kupiszewski and Kupiszewska 2003).

Tab. 1 - Consistency between regional and national population projections for EU member states

Country	Are the most recent regional and national population projections are	Model
Austria	Yes, mix top down/bottom approach	Multiregional cohort component model
Belgium	Yes	Component method
Belgium-Flanders	Not linked	Cohort component method
Finland	Yes, bottom-up consistency	
Germany	No	Multiregional cohort survival model
Italy	Yes, bottom-up consistency	Multiregional cohort component model
Netherlands	Yes, top-down	A hybrid form of multistate cohort survival model
Portugal	Yes, bottom-up consistency	Sequential model
Spain	Yes, they projected separately, and then adjusted	Component model
Sweden	Yes, yearly consistency adjustment to the national forecast	A pure demographic model
United Kingdom	Yes, bottom-up constrained by national population projection	Cohort-component model

Source: Kupiszewski and Kupiszewska 2003

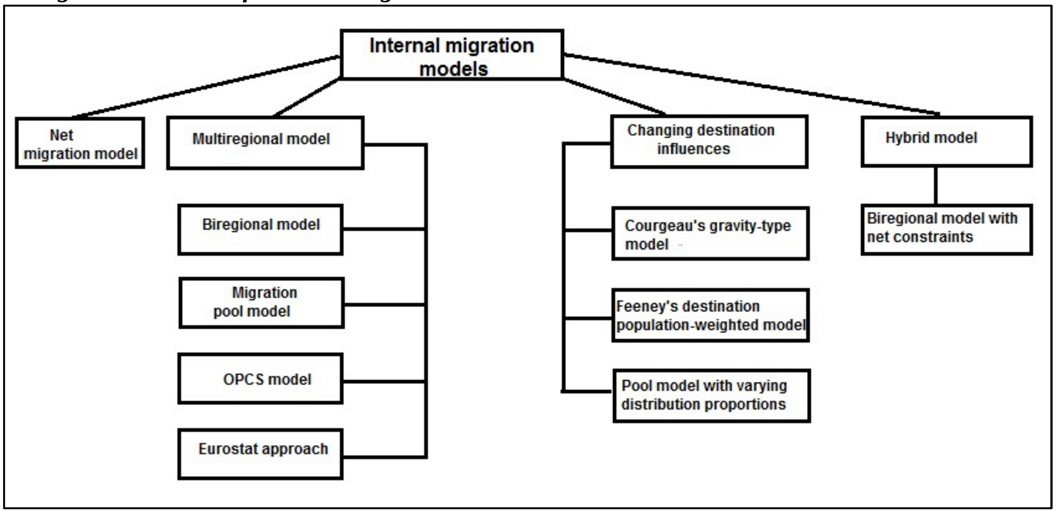
International migration is measured using emigration rates and immigration numbers. Flemish and Scottish forecasters combined external and internal migration into one variable. It should be noted that countries uses different criteria to define international migration and migrants that result in practical consequences. The predominant strategy in forecasting international migration is to set up time-varying totals, some countries use fixed total inflow. Distribution of immigrants to regions and recruitment emigrants from regions relate to the regional shares of existing stock of foreign populations, to labour markets in Austria, to internal migration in Germany, and to housing stock variables and historical time series in the Netherlands (Kupiszewski and Kupiszewska 2003).

The main difference in the methodologies of national and multiregional projections lies in the attempts to take into account migration within country. At the same time internal migration is the most volatile element and difficult for prediction, therefore it’s often neglected and omitted. Diversity of internal migration models are described in the next section.

3.3 Overview of internal migration models

In this section an emphases is placed on how internal migration is handled within subnational population projections. What are differences between models, advantages and disadvantages of every model and how suitable they could be? The mentioned below models as well as formulas and notation are borrowed from the work of Wilson and Bell (2004). These researchers described 10 internal migration models and applied the models to population projection for states and territories of Australia from 2001 to 2051.

Fig. 2 - Overview of internal migration models



Source: Wilson and Bell 2004

All models can be divided into four groups. In first group there is net migration model, the second group is represented by standard multiregional model and by four variants of reduced multiregional model, third group belong to models with changing destination attractiveness, and fourth group presents biregional model with net constraints as an example of hybrid model.

3.3.1 The net migration model

This model is widely used for local, regional, state, and subnational projections in many countries due to its modesty in demanding migration data. Internal migration is presented through of net migration numbers implemented in cohort component method:

$$P_{s,a+1}(t+1) = P_{s,a}(t) - D_{s,a,a+1}(t,t+1) + N_{s,a,a+1}(t,t+1),$$

where P denote populations,

D – deaths,

N – net migration,

s – sex,

a – single-year age group, infants are not included.

t – mid-year date in the projection horizon.

The projection equation has the following form when avoids iterative calculation and uses period-cohort exposure death rates:

$$P_{s,a+1}(t+1) = \frac{\left(1-0.5d_{s,a,a+1}(t,t+1)\right)}{\left(1+0.5d_{s,a,a+1}(t,t+1)\right)} P_{s,a}(t) + \frac{N_{s,a,a+1}(t,t+1)}{(1+0.5d_{s,a,a+1}(t,t+1))}.$$

Net migration model is a poor representation of reality since there is no such thing as ‘net migrant’ and net migration is residual of movements to a region and away from it. In practical implementation this models fails to capture the effects of changes in age structure and geographical distribution on the propensity to move besides when the net migration is negative aggregate loss can exceeds the available population (Wilson and Bell 2004).

3.3.2 The multiregional model

The shortcomings of net migration models are overcome in multiregional model which models origin-destination migration flows and computes them as a function of population responding to changing size, age-sex structure and geographical distribution.

Projection equation takes the following form:

$$\begin{aligned} P_{i,s,a+1}(t+1) = P_{i,s,a}(t) - D_{i,s,a,a+1}(t,t+1) \\ - \sum_{j \neq i} M_{i,j,s,a,a+1}(t,t+1) - E_{i,s,a,a+1}(t,t+1) \\ + \sum_{j \neq i} M_{j,i,s,a,a+1}(t,t+1) + I_{i,s,a,a+1}(t,t+1), \end{aligned}$$

where E is emigration,

I – immigration,

i - origin,

j – destination.

In order to concentrate on the internal migration in the next equation emigration and immigration flows are changed to fixed net international numbers:

$$P_{i,s,a+1}(t+1) = P_{i,s,a}(t) - (d_{i,s,a,a+1}(t,t+1) + \sum_{j\neq i} m_{i,j,s,a,a+1})0.5\left(P_{i,s,a}(t) + P_{i,s,a+1}(t+1)\right) + \sum_{j\neq i} m_{j,i,s,a,a+1}0.5\left(P_{j,s,a}(t) + P_{j,s,a+1}(t+1)\right) + N_{i,s,a,a+1}(t,t+1).$$

Following the description of Willekens and Drewe (1984) and Rees (1984) the above equation can be written in the matrix form:

$$\mathbf{p}_{s,a+1}(t+1) = \left[\mathbf{I} - 0.5\mathbf{M}_{s,a,a+1}(t,t+1)\right]\left[\mathbf{I} + 0.5\mathbf{M}_{s,a,a+1}(t,t+1)\right]^{-1}\mathbf{p}_{s,a}(t) + \left[\mathbf{I} + 0.5\mathbf{M}_{s,a,a+1}(t,t+1)\right]^{-1}\mathbf{n}_{s,a,a+1}(t,t+1),$$

where **p** are population vectors,

n – a vector of international migration,

M – a matrix of internal migration and deaths,

I – the identity matrix.

Multiregional model has several shortcomings. The first problem lies in small numbers of migration matrix. Age specific migration intensities sometimes are too jagged for smoothing and require model migration age profile. Other problems include difficulties of assumption-setting with so many variables in migration matrix.

Usually two solutions are applied to handle these problems. The first solution is to retain the full multiregional model, but to 'inflate' a partial set of migration data into a full matrix, more details given in the Eurostat approach. Another solution to reduce the volume of data required includes partitioning and aggregation. Partitioning is achieved by breaking up the model into small number of semi-independent subsystems with limited direct connections. Aggregation involves merging categories in any of the four principal dimensions: origins, destinations, age groups, sex. Three examples of reduced multiregional models presented here by bioregional model, migration pool model and OPCS model broad period-cohort origin-destination proportions (Wilson and Bell 2004).

3.3.3 The biregional model

Biregional model is an example of reduced multiregional model with spatial aggregation in which migration flows among regions reduced to migration flows between two regions where the second region is an aggregation of all other regions. The chain of such biregional models produces projection for whole country and origin-destination migration rates correspond to in-out-migration rates for each region.

The out-migration rate is computed as the number of migrations from a region in a base year divided by the population at risk in the same region:

$$omr_{i,s,a,a+1} = \frac{OM_{i,s,a,a+1}(T-1,t)}{0.5(P_{i,s,a}(T-1) - P_{i,s,a+1}(T))},$$

While the in-migration rate is defined as all migrants to a particular region from all other regions in the country divided population at risk in the rest of the country:

$$imr_{j,s,a,a+1} = \frac{IM_{j,s,a,a+1}(T-1,T)}{\sum_{i\neq j} 0.5(P_{i,s,a}(T-1) + P_{i,s,a+1}(T))}.$$

These rates have to be corrected to ensure that net internal migration sums to zero (Wilson and Bell 2004).

3.3.4 The migration pool model

Migration pool model represents partitioned model commonly used in European subnational population projections (van Imhoff, van Wissen and Spies 1994). Firstly, the number of out-migrations from each region are projected and gathered in a common 'pool', and secondly, in-migration flows are allocated to destinations.

For example, out-migrations are received by applying out-migration rates to the origin populations:

$$OM_{i,s,a,a+1}(t,t+1) = omr_{i,s,a,a+1} 0.5\left(P_{i,s,a}(t) + P_{i,s,a+1}(t+1)\right).$$

Then received the out-migrations from all regions are placed in sex- and period-cohort-specific 'pools':

$$Pool_{s,a,a+1}(t,t+1) = \sum_i OM_{i,s,a,a+1}(t,t+1).$$

Migration pools multiplied by distribution proportions make number of in-migrations to each destination:

$$IM_{j,s,a,a+1}(t,t+1) = Pool_{s,a,a+1}(t,t+1) p_{j,s,a,a+1}$$

The distribution proportions are fixed and calculated from base year data as:

$$p_{j,s,a,a+1} = \frac{IM_{j,s,a,a+1}(T-1,T)}{\sum_j IM_{j,s,a,a+1}(T-1,T)}$$

Thus, migration pool model has the following compact form:

$$M_{i,j,s,a,a+1}(t,t+1) = omr_{i,s,a,a+1} 0.5\left(P_{i,s,a}(t) + P_{i,s,a+1}(t+1)\right) p_{j,s,a,a+1}.$$

Conceptual shortcoming of this model is that migrants in the pool are at risk of being immediately returned to their origins (van Imhoff et al. 1994). In-migrants distributed to each destination may be assumed to originate from different regions including origin unless in-migration to one origin exceeds total out-migration from all other (Wilson and Bell 2004).

3.3.5 The OPCS model

Multiregional model of the former Office of Population Censuses and Surveys in the UK (OPCS) uses age aggregation in the distribution of migrants to destinations. While out-migration is projected similar to pool migration or biregional models, in-migration is projected using broad period-cohort origin-destination proportions. Number of migrants can be obtained through the migration rates in standard multiregional model:

$$M_{i,j,s,a,a+1}(t,t+1) = m_{i,j,s,a,a+1} 0.5(P_{i,s,a}(t) + P_{i,s,a}(t+1))$$

Then, migration rate can be decomposed using base year data into an out-migration rate and distribution proportion conditional on out-migration from specific origin:

$$M_{i,j,s,a,a+1}(t+1) = \frac{OM_{i,s,a,a+1}(T-1,T)}{0.5(P_{i,s,a}(T-1)+P_{i,s,a+1}(T))} \times \frac{M_{i,j,s,a,a+1}(T-1,T)}{OM_{i,s,a,a+1}(T-1,T)} \times 0.5\left(P_{i,s,a}(t) + P_{i,s,a+1}(t+1)\right).$$

Since distribution proportions vary considerably with life course stages and not so much within them, broad period-cohort (A, A+1) origin-destination distribution proportions are introduced:

$$p_{i,j,s,A,A+1} = \frac{M_{i,j,s,A,A+1}(T-1,T)}{OM_{i,s,A,A+1}(T-1,T)},$$

Newly introduced proportions can be used for each single year a,a+1 period-cohort within broader A,A+1 period-cohort.

$$M_{i,j,s,a,a+1}(t, t+1) = omr_{i,s,a,a+1} 0.5(P_{i,s,a}(t) + P_{i,s,a}(t+1)p_{i,j,s,A,A+1}$$

The above written equation gives number of migrants using broad period-cohort origin-destination proportions (Wilson and Bell 2004).

3.3.6 The Eurostat approach

This approach uses another strategy when partial set of migration data ‘inflate’ into a full matrix. There are two methods of approximating the full migration matrix. The first method was introduced by Rogers and colleagues in which age migration profiles was decomposed into three curves describing pre-labor force peak, labor force peak and retirement peak. The model schedule has 11 parameters, or seven parameters in reduced form when retirement peak is negligible small, a substantial saving on the 100 age-specific migration rates otherwise needed (Rogers, Raquillet and Castro 1977).

In the second method applied in several rounds of European Union regional projections (Rees *et al.* 2001) a full set of mi,j,s,a,a+1 rates was derived from a partial matrix of migration flows. In the refined 1995-based projections combinations of four interregional dimensions without additional information, namely origin, destination, age and sex, were tested in order to find the best representing the full matrix (van Imhoff et al. 1997).

The full matrix of origin-destination migration flows is received from the following elements: sex and period-cohort specific out-migration rates; sex and period-cohort specific in-migration proportions (similar to migration pool model); an origin-destination matrix without age and sex breakdown; populations at risk by sex and period-cohort using the equation (van der Gaag, van Imhoff and van Wissen 1997):

$$\widehat{M}_{i,j,s,a,a+1}(T-1, T) = omr_{i,s,a,a+1} 0.5(P_{i,s,a}(T-1) + P_{i,s,a+1}(T))p_{i,s,a,a+1} f_{i,j}.$$

Where $f_{i,j}$ is the aggregate origin-destination effect defined as the ratio of the observed origin-destination flow to the expected flow in the absence of any spatial effects in migration:

$$f_{i,j} = \frac{M_{i,j}^{observed}(T-1, T)}{M_{i,j}^{expected}(T-1, T)} = \frac{M_{i,j}^{observed}(T-1, T)}{\sum_s \sum_a \left(omr_{i,s,a,a+1} 0.5\left(P_{i,s,a}(T-1) + P_{i,s,a+1}(T)\right) p_{j,s,a,a+1}\right)}.$$

Then origin-destination migration rates are calculated and applied in multiregional model being constant during the whole projection horizon (Wilson and Bell 2004).

3.3.7 Courgeau's gravity-type model

Gravity type model was inspired by Newton's Law of gravitation in physics, which suggests that gravity depends positively on mass and negatively on distance. The basic idea is that larger places attract more people than smaller places, and places closer together have greater attraction.

Assuming that W represents the set of attributes of each region and c is an intervening distance then influence of distance simultaneously with effects of origin and destination can be written in the form:

$$M_{i,j} = f(W_i, W_j, c_{i,j}).$$

In the simplest case W represent the populations of origin and destination regions, in complicated forms can include social and economic variables, in that cases the problem arises how to project these exogenous variables.

Gravity model proposed by Daniel Courageu represents multiregional model where migration rates are explicitly linked to both origin and destination populations, though not to distance. The rates which do not require fitting, are defined as

$$m_{i,j,s,a,a+1} = \frac{M_{i,j,s,a,a+1}(T-1,T)}{P_{i,s,a}(T-1)P_{j,s,a+1}(T)}, \quad i \neq j,$$

Supposing migration rates being fixed migration flows are projected using equation:

$$M_{i,j,s,a,a+1}(t,t+1) = m_{i,j,s,a,a+1}P_{i,s,a}(t)P_{j,s,a+1}(t+1).$$

When regions will grow they will attract increased share of people, number of out-migrants is not fixed proportion of origin population at risk and increase as populations increase (Wilson and Bell 2004).

3.3.8 Feeney's destination population weighted model

In contrast to gravity model there is another approach in which standard multiregional model is used for incorporating destination population influences in migration projections. Feeney proposed model (1973) in which origin-destination migration rates are modified every year in proportion to a ratio that captures the progressive change in the distribution of population between regions:

$$m_{i,j,s,a,a+1}(t,t+1) = m_{i,j,s,a,a+1}(T-1,T) \times \frac{(P_j(t)/\sum_{j\neq i} P_j(t))}{P_j(T)/\sum_{j\neq i} P_j(T)}.$$

According to Plane and Rogerson (1994) the Feeney's model shortcoming is that the projected migration rates could, in theory, sum to more than 1.0 (Wilson and Bell 2004).

3.3.9 A pool model with varying in-migration distribution proportions

In this version out-migration as well as base year proportions used for distribution the migrants to the destinations are calculated similar to the basic pool model. Then, in contrast to basic pool model where these proportions are constant, in this version distribution proportions are modified every year according to changing attractiveness of destination regions. Firstly, temporally values for the distribution proportions are received through multiplication by the ratio of the destination population share (of the national share) to its share in the launch year:

$$p^*_{j,s,a,a+1}(t,t+1) = p_{j,s,a,a+1}(T-1,T)\,\frac{P_j(t)/\sum_j P_j(t)}{P_j(T)/\sum_j P_j(T)}$$

Next adjustment stipulates for each sex- and period-cohort the distribution proportions sum to 1.0:

$$p_{j,s,a,a+1}(t,t+1) = p^*_{j,s,a,a+1}(t,t+1)\,\frac{1}{\sum_j p^*_{j,s,a,a+1}(t,t+1)}.$$

In contrast to Feeney model, the number of out-migrations depends on out-migration rates and origin populations only, whereas in-migration distribution proportions are computed for all regions including the origin. This reveals the shortcoming of model that origin population size affects the in-migration distribution proportions (Wilson and Bell 2004).

3.3.10 The biregional model with net constraints

This model uses biregional model to make first estimates of in-migration and out-migration for each region and then scales the resulting inward and outward flows to match predetermined net internal migration assumptions. The similar approach with varying scaling for every state was used by Australian Bureau of Statistics to produce official population projections of the states and territories (ABS 1995).

As soon as the preliminary biregional in- and out- migration projections are estimated, subsequently the difference between this modeled value of net migration (denoted by *) and predetermined net migration is derived:

$$\mathrm{diff}_i(t,t+1) = N_i^{predet}(t,t+1) - \left(\sum_s\sum_a IM^*_{i,s,a,a+1}(t,t+1) - \sum_s\sum_a OM^*_{i,s,a,a+1}(t,t+1)\right).$$

The obtained difference is distributed evenly between in- out-migration adjusting them:

$$IM_{i,s,a,a+1}(t,t+1) = IM^*_{i,s,a,a+1}(t,t+1) \times \frac{\left(\sum_s\sum_a IM^*_{i,s,a,a+1}(t,t+1) + 0.5\ \mathrm{diff}_i(t,t+1)\right)}{\sum_s\sum_a IM^*_{i,s,a,a+1}(t,t+1)}$$

$$OM_{i,s,a,a+1}(t,t+1) = OM^*_{i,s,a,a+1}(t,t+1) \times \frac{\left(\sum_s\sum_a OM^*_{i,s,a,a+1}(t,t+1) - 0.5\ \mathrm{diff}_i(t,t+1)\right)}{\sum_s\sum_a OM^*_{i,s,a,a+1}(t,t+1)}.$$

On the one hand the biregional model provides that the age profiles of in- and out-flows reflect changes in the age composition of the origin populations. On the other hand imposing the net constraints means that underlying migration flows are no longer consistent function of origin or destination population size (Wilson and Bell 2004).

In conclusion, every above-mentioned method has its own advantages and drawbacks. Chapters 7 and 8 of this thesis show how multiregional method can be applied for population projection of Kazakhstan. This approach reveals migration impact on the growth of regional population. It is a good method when interregional migration flows are intensive and heterogeneous, as in case of Kazakhstan. Multistate approach has a wide variety of applications, for example nuptiality, education and household projections. In analogy with multiregional projection it could be useful for population projection when population is broken down into different categories and this division between people could be dynamic.

Chapter 4

General information about Kazakhstan

The Republic of Kazakhstan is a country located in Central Asia, independent since the year 1991. Kazakhstan is the ninth largest country in the world and has a territory of 2.7 million square kilometers which is equivalent to the size of Western Europe. The distinctive feature of this country in comparison with other Asian countries namely neighbor China is very low population density 5.8 per sq km. According to the preliminary results of the last census which was held in 2009 the population of Kazakhstan consists of 16,004,800 people that exceeds 14.95 million announced after the previous census in 1999 by 6.8 percent and designate long-expected progress in the growth of population.

Composition of the population includes 54 percent rural and 46 percent urban inhabitants, and proportion of men makes up 48.3 percent against 51.7 percent of women. Life expectancy of males is 67.9 years as a rule lower than 73.5 years life expectancy of females, and infant mortality rate is 25.7 infant deaths for 1000 live births, that is still high in comparison with achievements of other countries. Crude Death rate is equal to 9.4 in relation to crude birth rate 16.6 for 1,000 inhabitants, and total fertility rate is 2.68.With regard to age structure children of 14 years old and under constitute 21.8 percent; share of people older than 65 years makes up 7.9 percent, proportion of people of age group 15-64 amount to 70.2 percent. Mean age of the inhabitants is 31.7 years.

4.1 Development of Kazakhstan by regions

Kazakhstan is divided into 16 administrative regions. This division includes 14 oblasts and two cities. The first column of the Tab.2 shows that the lowest value of population density equal to 2.4 people per sq km belongs to Aktyubinskaya oblast. While the highest value appertain to Almaty city with population density 4550 comparable with London's 4761 people per sq km. New capital Astana is not so populated like former capital but possesses the highest life expectancy in country equal to 73.75 years combined for both sexes.

Among provinces South-Kazakhstanskaya oblast has the highest population density 20.3people per km^2, as well as the biggest population 2,381,543 inhabitants, also the highest value of TFR 3.93, and the youngest mean age of population in the country 27 years, and as a borderline province, distinguished by high level of immigration, in-migration and out-migration.

Fig. 3 - Administrative divisions of Kazakhstan

Source of map: //www.nationsonline.org/oneworld/map/Kazakhstan-administrative-map.html

In contrast two northern provinces, Kostanaiskaya oblast and Severo-Kazakhstanskaya oblast, show the lowest value of TFR 1.69 and 1.77 correspondingly, and the highest mean age 35.4 year, and as borderline regions with Russia indicate high level of migration.

Another pair, Almatinskaya oblast and Karagandinskaya oblast have both high populations exceeding 1 million, but by the character of production they differ considerably, and Karagandinskaya oblast has the highest proportion of urban population 78 percent while Almatinskaya oblast has the highest proportion of rural population 77 percent (Table 2). This is actually happens since cities, towns and settlements in Karagandinskaya oblast were created and developing on the base of industrial factories and plants, while other branches are dependent on the main factories.

Regarding socio-economic indicators 16 administrative regions of Kazakhstan can be grouped into four categories:

1. Oil-and-gas producing regions in the western part of Kazakhstan (Atyrauskaya oblast, Mangystauskaya oblast, Aktyubinskaya oblast, Zapadno-Kazakhstanskaya oblast, Kyzylordinskaya oblast);

2. Industrial regions in the central, northern and eastern parts (Karagandinskaya oblast, Vostochno-Kazakhstanskaya oblast, Pavlodarskaya oblast);

3. Agricultural regions in the south and north of Kazakhstan (Akmolinskaya oblast, Almatinskaya oblast, Kostanaiskaya oblast, Zhambylskaya oblast, Severo-Kazakhstanskaya oblast, Yuzhno-Kazakhstanskaya oblast);

4. New and old capitals of the country (Astana, Almaty).

Tab. 2 - Demographic indicators by regions, 2008

Region	Population density (per sq km)	Mean age of population (years)	TFR	Life expectancy at birth for both sexes (years)
Kazakhstan	5,8	31.7	2.68	67.1
Akmolinskaya oblast	5,1	33.7	2.11	65.0
Aktyubinskaya oblast	2,4	30.9	2.62	67.6
Almatinskaya oblast	7,4	31.2	2.85	67.5
Atyrauskaya oblast	4,2	28.9	3.26	67.4
Zapadno-Kazakhstan. oblast	4,1	32.9	2.29	67.3
Zhambylskaya oblast	7,1	29.6	3.35	67.4
Karagandinskaya oblast	3,1	34.1	1.98	64.6
Kostanaiskaya oblast	4,5	35.4	1.69	65.6
Kyzylordinskaya oblast	2,8	27.8	3.51	67.5
Mangystauskaya oblast	2,6	27.5	3.73	66.8
Yuzhno-Kazakhstan. oblast	20.3	27.0	3.93	67.9
Pavlodarskaya oblast	6,0	34.4	2.01	66.5
Severo-Kazakhstan. oblast	6,6	35.4	1.77	65.4
Vostochno-Kazakhstan. oblast	5,0	35.0	2.05	65.8
Astana	913,0	32.8	2.38	73.75
Almaty	4550.0	33.6	2.55	70.35

Source: The Agency on Statistics of the Republic of Kazakhstan

Disparities in regional development are intrinsic characteristics of every country, in Kazakhstan they caused by following conditions:

1. Concentration on the extractive industry;

2. Necessity to introduce innovations into different branches, especially in agricultural sector where one third of a nation is working,

3. Centralization of economic planning and isolations of regions from centers due to long distances.

Oil-and-gas producing regions and industrial regions are dependent on the price variation of oil and metals on the world level and on the extension of resources. World financial crisis which began at the end of 2007 year adversely affected on the wellbeing of Kazakhstan and increased the unemployment level, especially it has had negative impact on banking area, building constructions, etc. Regional policy in order to solve social and economic problems tries to redistribute revenue by flattening economic disparities between regions.

Tab. 3 - Socio-economic indicators by regions, 2008

Region	Urban population (%)	Rural population (%)	Gross Regional Product per capit (bill. ten.)	Main industry branches
Akmolinskaya oblast	45	55	478	bulding construction and agriculture
Aktyubinskaya oblast	54	46	872	mining and miling, oil and gas development, agriculture
Almatinskaya oblast	23	77	677	building construction, agriculture
Atyrauskaya oblast	49	51	1799	oil and gas extraction, agriculture
Zapadno-Kazakhstan. oblast	45	55	827	gas production and agriculture
Zhambylskaya oblast	42	58	325	phosphor production and mineral fertilizers, agriculture
Karagandinskaya oblast	78	22	1463	coalmining, flat-rolled products, refined copper, agriculture
Kostanaiskaya oblast	49	51	704	are metal ore mining and agriculture
Kyzylordinskaya oblast	35	65	685	throughput of crude oil, building construction and agriculture
Mangystauskaya oblast	54	46	1096	throughput of crude oil and gas, agriculture
Yuzhno-Kazakhstan. oblast	38	62	731	agriculture
Pavlodarskaya oblast	66	34	862	coalmining, steel discharge, ferroalloy smelting, agriculture
Severo-Kazakhstan. oblast	35	65	403	agriculture
Vostochno-Kazakhstan. oblast	54	46	890	coal mining, chemical industry, building constructions, agriculture
Astana city		-	1372	building industry, government sector
Almaty city		-	2950	building constructions, communication infrastructure

Source: The Agency on Statistics of the Republic of Kazakhstan

Constant efforts are made already to achieve balanced development of regions and progressive steps in this direction continue. Since the beginning of 2000s several government programs were implemented, based on different approaches they cover many aspects of people’s life. The current program “The concept of Territorial Development Strategy 2015” which has come after the program “Conception of the regional policy of the Republic of Kazakhstan for the period 2002-2006” puts the accent on diversification of economics and reform of regional development by forming the centers of economic advancement using clustering approach.

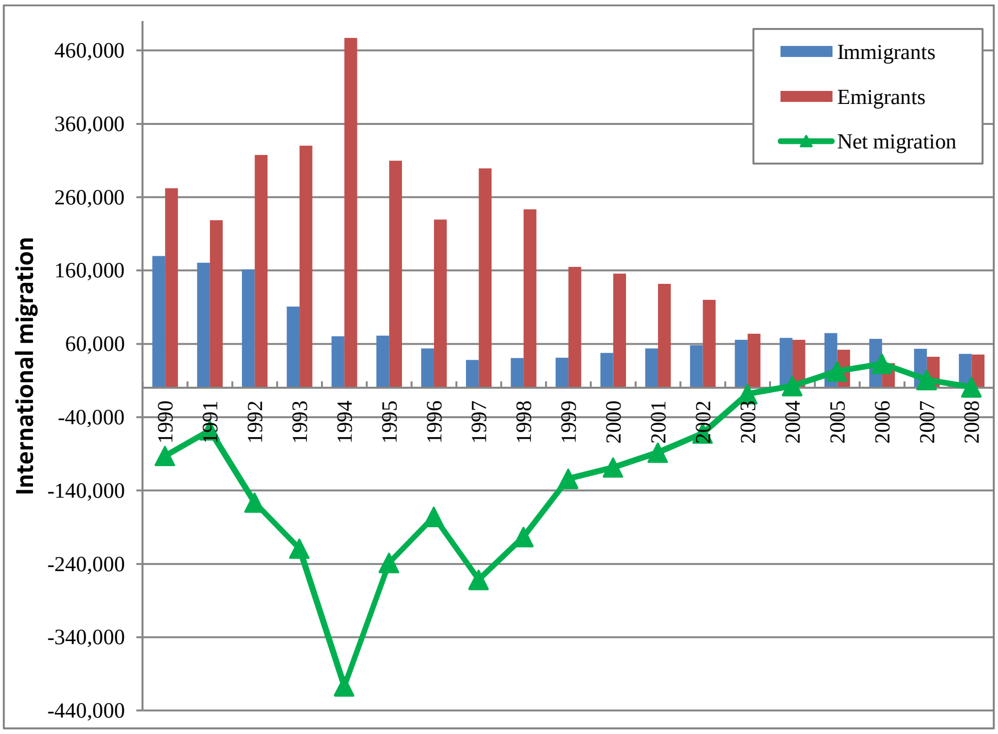
4.2 Migration processes

Migration policy in Kazakhstan is based on the following legislative documents: the Constitution of the Republic of Kazakhstan vesting citizens with rights of free movements and choice of place of residence, and defining rights and duties of foreigners and stateless people; Migration regulatory acts accepted during the period 1993-1995; Population migration Law of the Republic of Kazakhstan № 204 dated 13th December, 1997, Concepts of migration policy of the Republic of Kazakhstan for the periods 2001-2007 and 2007-2015 (Bapakova 2010).

4.2.1 Emigration

The fall of Iron Curtain, disintegration of Soviet Union, proclamation of independence of Kazakhstan and ethnic return migration policies of other countries had a significant impact on migration flows in Kazakhstan, both external and internal. Since the beginning of 1990s Kazakhstan encountered exodus of various ethnic groups, specially German, Russian, Jewish, Estonian people, who returned to their historical homeland Germany, Israel, Russia, Estonia, some of migrants moved to Poland, Hungary etc.

Fig.4 - Number of immigrants, emigrants and net migration of Kazakhstan, 1990-2008

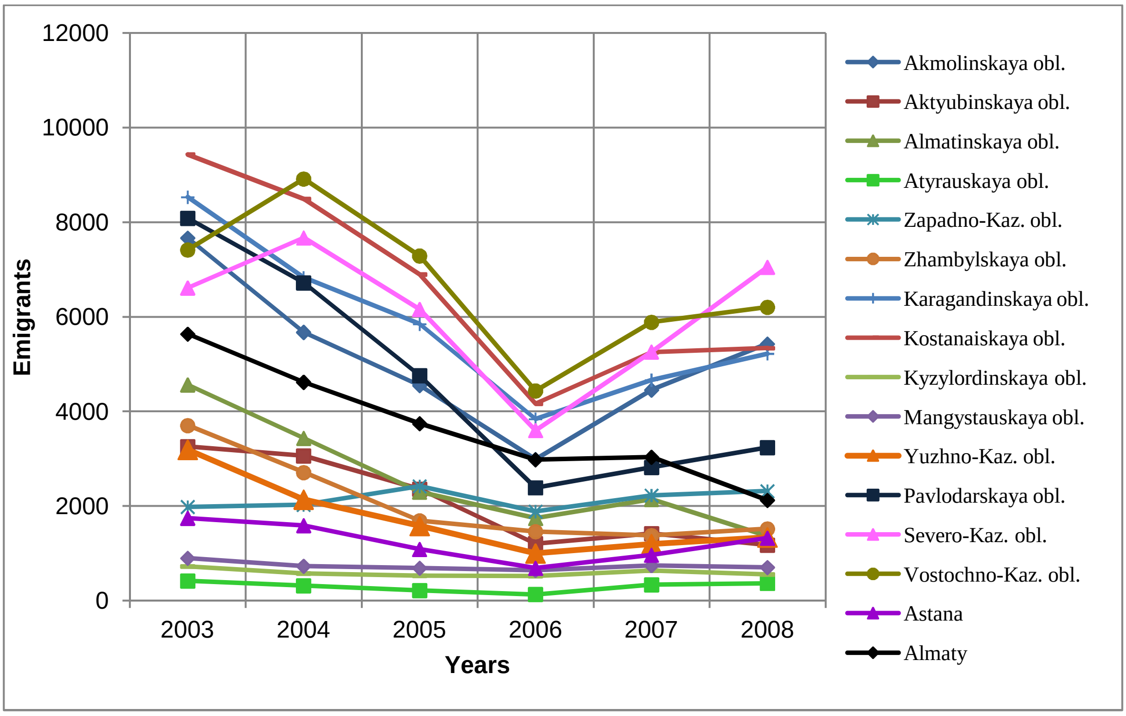


Source of data: The Agency on Statistics of the Republic of Kazakhstan

The peak of these emigration flows was in 1994 when about half of million people left Kazakhstan and two thirds of them of working ages. These outflows are still continued, but the intensities are slow downed, since 2004 Kazakhstan has positive net migration (Fig.4). Nevertheless it's resulted in depopulation of North Kazakhstan and Kootenaiskaya oblasts, and in general it decreased the proportion of these ethnic groups in the national composition of country.

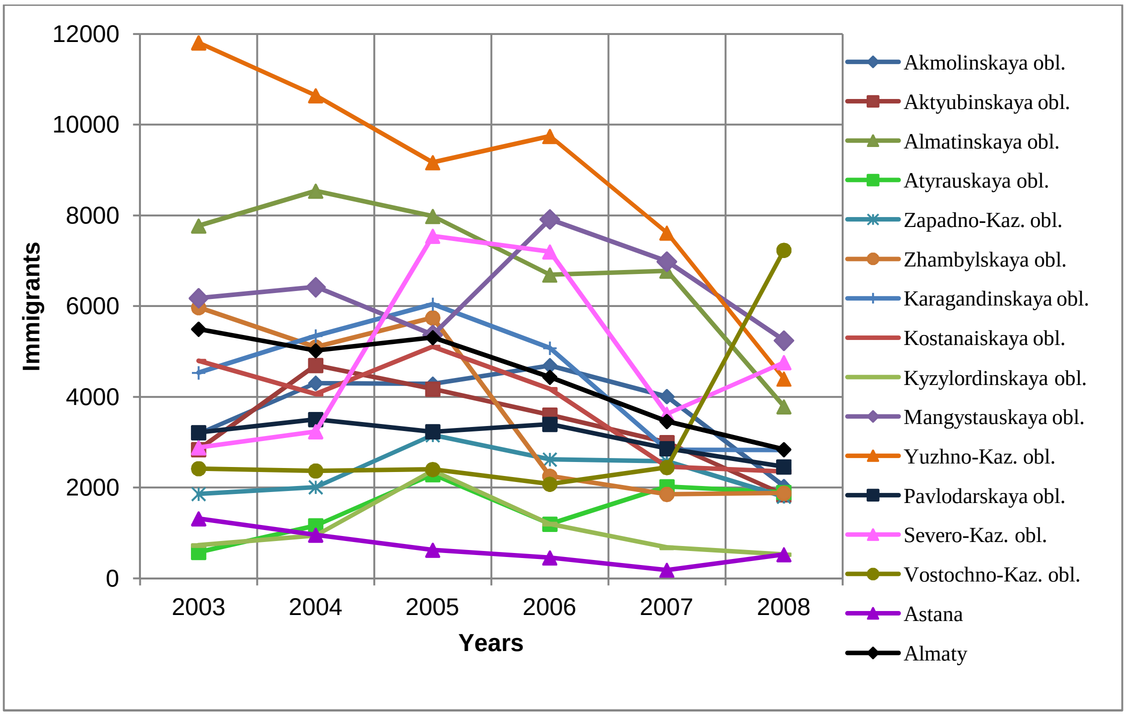
Fig. 5 presents number of emigrants for the period 2003-2008 by regions, while Figure 6 shows number of immigrants by regions for the same period.

Fig.5 - Number of emigrants by regions of Kazakhstan, 2003 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 6 - Number of immigrants by regions of Kazakhstan, 2003 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

4.2.2 Ethnic migration policy

According to the last census of the year 2009 the proportion of Kazakhs has increased by 26% in comparison with previous census, and comes to 63% or 10,099 thousands of people. The increase in the titular nation population by 800 thousand people was achieved through of immigration of oralmans from Mongolia, China, Russia and other neighboring countries.

During the process of legitimizing the territorial sovereignty and nation-building Kazakhstan has actively promoted the ethnic return migration policy in order to ingather Kazakh diaspora living abroad. Under the Law on Population Migration accepted in 1997 foreign or stateless persons of Kazakh nationality are granted status of oralmans. The status of oralman give them priority in obtaining the permanent residence permits, afterwards Kazakhstan citizenship, and broad spectrum of benefits, including financial support from the state budget as a lump-sum grant, reimbursement of travel and transportation expenses, funds helping to purchase of housing, priority support in job placement and getting education. Oralmans coming from Mongolia, China and Russia can improve their knowledge of Kazakh and Russian languages in 14 adaptation centers, also they can get there legal advice or (re)training of skills required for professional development.

The quotas for accepting oralmans were established in1993 and since then increased gradually.

For the period 2000-2001 annual quotas were 500-600 families, for the period 2005-2008 the quota increased to 15 thousands families. Current quota for the period 2009-2011 is 20 thousand families. In 2009 year Kazakhstan accepted 16,336 families. By age composition 54 percents constitute people of labor age, while proportion of person under age 18 is 41 percents (Bapakova 2010).

About 50 percent of repatriates are living today in overpopulated areas with high level of unemployment among residential population. According to the state program Nurly Kosh for the period 2009-2011 years adaptation and integration of repatriates is based on rational distribution, and takes into account internal migrants living in unfavorable regions and former citizen of Kazakhstan coming back to work. New policy makes accent on redistribution of oralmans by providing financial and accommodation support differentially depending on regional coefficients.

4.2.3 Refugees

Political persecution at home forces some migrants from developing countries to cross national borders. The 1951 Geneva Convention defines a refugee as a person outside his or her country of citizenship who does not want to return "owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership in a particular social group, or political opinion." Countries that sign the Geneva Convention pledge not to "refoul" or return those recognized as refugees to places where they could be persecuted (Martin and Zürcher 2008).

The Law on Adherence by the Republic of Kazakhstan to the 1951 Geneva Convention Relating to the Status of Refugees and its 1967 Protocol was adopted in 1998. The number of registered refugees by the 1 October 2010 was 618 people; 594 of them are from Afghanistan, 1 is from Somalia, 2 are from Ethiopia, 15 are from Uzbekistan, 2 are from Kyrgyzstan and 4 from China. A total of 509 people seeking asylum have applied to regional offices of the Committee on Migration this year (Abishev 2010).

In concordance with the Law on Refugees accepted in 2009 the status of refugees is given for one year period, after one year the applications of people are considered again, and persons whose applications were denied have to leave the country. Most of the refugees work, and those with a per capita income that is not above the poverty line receive targeted social assistance. Refugees have the equal rights with citizens of Kazakhstan in respect of health care, social security, money transfer and ownership rights to housing and other property (Abishev 2010).

4.2.4 Labor migration

During the recent years Kazakhstan has become a country accepting migrant workers from other countries attracted by relative economic growth, high price level, social stability, geographic location and liberal migration policy in Kazakhstan.

The Government of Kazakhstan supports temporary labor migration with the purpose to maintain competitiveness of labor market and to recruit specialists of high qualification and workers of professional skills. At the same time annual quotas are established in order to control situation, distribute foreign workers according necessity and to secure local workers. This year the quota is 63,700 people represented 0.75 percent of economically active population. There are 19,800 foreign workers currently working in the country under legal permits, and 73 per cent of them are managers or specialists at the managerial level. China, Turkey and the United Kingdom continue to be the main countries of origin for labour migrants. Companies carrying out main projects relating to industry- and innovation-based development reported that they are able to meet only about half of their manpower needs, and that they need government assistance to train 108,000 workers (74,900 during the construction period and 33,100 during the operational period), which means that the need for government assistance is the same size as the manpower shortage. Particularly, construction of the transportation and energy infrastructure require 39,900 and 6,500 workers, respectively, while the oil-gas sector, the tourism industry and machine-building lack for 24,100, 6,800 and 6,000 workers respectively (Abishev 2010).

According to current legislation, foreign nationals visiting Kazakhstan as tourists or coming to study or by invitation of relatives or acquaintances may not be employed and do business other than stated in visa unless otherwise provided in the Kazakhstan laws. Foreigners staying temporarily in Kazakhstan can’t bring unlicensed labour.

The procedure for providing authorizations for use of migrant labour is based on giving priority to the gradual replacement of foreign workers by local personnel. The process of obtaining foreign workforce permission consists of three steps:

1. Searching for job seekers on the domestic labour market with mandatory use of prescribed forms and in compliance with their terms;
2. Obtaining an authorization for a definite number of foreign workers by the categories and qualifications specified in it;
3. Certification of a nominal list of foreign workers being hired by the same state authority.

The first step requires from employers to ensure a preferential principle for employment of Kazakh nationals and to present a proof that there are no appropriate applicants on the domestic labour market. Step two requires inclusion of several documents such as an application specifying the number and category of foreign workforce to be hired by individual occupations and skills, job qualifications for each position in accordance with the relevant regulatory documents of the Ministry of Labour and Social Protection and justification for number of hired foreign workforce requested. On the third step the employer concludes contracts of employment with foreign workers, ensures migrant workers' departure to places of their permanent residence upon expiry of the permit term, and makes a list of hired migrant workers which should be approved by the local executive body. Since 1 January 2008 assessment of qualifications of migrant workers by employers is based on scoring system taking into account such characteristics as education, relevant service record, and demand for a specific profession occupation on the labour market (Ni 2008).

For mobilization of investment the recruiting of the following categories of foreign workers is exempted from preferential searching on domestic labor market: the heads of corporation with foreign persons shares in authorized capital no less than 50 percent, board directors of corporations with state or foreign persons shares no less than 50 percent, academicians and scientists of higher education and research institutions, foreign workers of companies participating in projects of the state program 'The 30 corporate leaders' (Bapakova 2010).

In case of family reunification of labour migrants according to the changes made in the Regulations for Arrival and Stay of Foreigners in the Republic of Kazakhstan by Government Resolution No. 688 dated 10 August 2007 it's necessary to present confirmation of the availability of funds for the stay and departure of family members. This requirement does not apply to family members of the oralmans, the persons who were born or had been the nationals of the Republic of Kazakhstan or the Kazakh Soviet Socialist Republic, and members of their families. Besides, Kazakhstan maintains a visa-free regime with many countries supplying migrant workers such as Kyrgyzstan, Uzbekistan and Tajikistan (Ni 2008).

In case of loosing job legal migrant worker person can look for a new job according vacancies of employers with obtained permissions to employ foreign workers of appropriate category and qualification. In case of admission this employer has to receive the approval to new amended lists of recruited foreign workers under the simplified procedure (Ni 2008).

According to special rules approved by Government Resolution of No. 997 dated 16 October 2006 foreigners and stateless persons temporarily and permanently residing in Kazakhstan including irregular migrants can be provided with emergency medical care and medical treatment on a free basis in case the disease is dangerous without requirements of visa or registrations. Other medical services are provided to foreign nationals and stateless persons at their own expense, at the expense of their employers, the voluntary health insurance system and from other sources not prohibited under the Republic of Kazakhstan legislation (Ni 2008).

Kazakhstan legislation grants the right to purchase housing only to the nationals of Kazakhstan. The general practice in a number of labour receiving countries with respect to seasonal and lower skilled migrant workers that employers provide workers with access to decent and low cost housing on a free, subsidized or non-subsidized basis. According to the Law of the

Republic of Kazakhstan No. 319-III ZRK dated 27 July 2007 “On Education” the foreigners permanently residing in Kazakhstan have the same rights as Kazakhstan nationals in access to pre-school, primary, basic secondary and general secondary education. During the enrolment of children of migrant workers to grade one, it is required that parents provide only a document confirming their place of residence (registration), while older children may face difficulties in accessing school education due to differences in the curricula (Ni 2008).

The issue of migrant workers’ right to social protection is linked to pension schemes, mandatory social insurance, and compensation of damage due to occupational accident. Since migrants workers are not recognized as permanent residents in Kazakhstan employers do not make pension fund contributions for them. According to the letter of the Ministry of Labour and Social Protection No. 07-02-17/6 dated 10 November 2006 migrant workers may not take advantage of the services of Kazakhstan defined-contribution pension funds even if they or their employers make voluntarily contributions. The same reason with social insurance, employers need not make social contributions for migrant workers. The Labour Code establishes the employer’s liability to compensate for the harm caused to the employee’s life and health while performing his or her duties, including as a result of occupational accidents. This requirement extends to the migrants workers as well. In order to ensure fulfillment of the obligation to compensate for the harm caused to the worker’s life and health when performing his or her duties, the employer’s mandatory civil liability insurance system is established in Kazakhstan. The provisions of Republic of Kazakhstan Law No. 30-III ZRK dated 7 February 2005 on the “Employer’s Compulsory Civil Liability Insurance to Cover any Harm Caused to the Worker’s Life and Health in Performing His/Her Labour (Working) Duties” extend to all workers, including migrants (Ni 2008).

Differences in the assessment of labor inputs can cause some tension between local and foreign workers as it happened in 2006 in West Kazakhstan between Turkish and Kazakh workers. International organizations such as World Bank assert that remittances of legal migrant workers reduce poverty and speed development in the migrants’ countries of origin. But the remittances of legal migrant workers are only visible parts of iceberg, whereas other parts relating to money transfers of illegal migrant workers are much bigger and hidden.

From 1 August to 31 December 2006, a one-off action to legalize irregular labour migrants from CIS countries was held in Kazakhstan. Subject to registration with internal affairs bodies, labour migrants, employed in Kazakhstan without proper authorizations obtained by their employers were granted the right to register and subsequently work in Kazakhstan on legal grounds for a period not exceeding three years. The regularized migrant workers were entered into a special register and issued special migration cards as a confirmation of their right to employment. Following the registration of migrant workers, the employers were to pay all payroll taxes. The regularized migrant workers and their employers were relieved from administrative and criminal responsibility for previous irregularities linked with illegal labour recruitment activities. This procedure extended only to the persons who had arrived 60 calendar days before the enactment of the Law on Amnesty due to Legalization of Illegal Migrant Workers, and who since then had been engaged in a labour activity in Kazakhstan. At the same time, legalized labour migrants were not granted the right of free access to the Kazakhstan labour market, but only an opportunity to formalize their labour relations with the employer that used the procedure for their registration (Ni 2008).

4.2.5 Unauthorized migration

The Law on Migration of the Population defines illegal migrants as foreigners or stateless persons who have entered and stay in the Republic of Kazakhstan in violation of the legislation regulating the procedure for entry, departure, stay and transit via the territory of the Republic of Kazakhstan, for example illegal entry, evasion of departure at a stated time, nonconformity of the purpose of entry to the purposes specified in the visa or at registration or in the migration card, infringement of the transit rules (Ni 2008).

The vast territory of Kazakhstan is divided into 14 provinces (oblasts), and 12 of them are near-border provinces. Yuzhno-Kazakhstanskaya oblast, Kyzylordinskaya oblast, Mangystauskaya oblast and Aktyubinskaya oblast share common borders with Uzbekistan, 2300 km. Almatinskaya oblast and Zhambylskaya oblast have common borders with Kyrgyzstan about 980 km. Seven provinces are neighbors with Russia: Atyrauskaya oblast, Zapadno-Kazakhstanskaya oblast, Aktyubinskaya oblast, Kostanaiskaya oblast, Severo-Kazakhstanskaya oblast, Pavlodarskaya oblast and Vostochno-Kazakhstanskaya oblast. The Kazakhstan-Russia border is equal to 7500 km. China borders on Almatinskaya oblast and Vostochno-Kazakhstanskaya oblast about 1740 km. The total length of national boundary is 12,187 km including 600 km by the Caspian Sea. Diversity of routes, roads and byroads, and dense railway traffic cross frontier.

According to geopolitical location Kazakhstan is situated between Europe and Asia. Range of “pushing” forces in countries of origin such as poverty, unemployment, discrimination and social instability compel mobile groups from South Asia, for example Sri-Lanka, India, Afghanistan, to move in other countries, while Kazakhstan due to above mentioned factors can be chosen as a transition country on the way to Russia and European countries. Some stray migrants from transit flow ‘sink’ in Kazakhstan.

Kazakhstan accepts citizens of Turkey and CIS countries except for Turkmenistan without requirements of visa. Most migrants come from Central Asia countries. In addition to “pushing” factors at home there are several “pulling” factors in Kazakhstan, such as relatively high level of living standards in comparison with other CIS countries, closeness, absence of linguistic barrier, common religion - Islam and network of relatives and countrymen, who came earlier, more successful and sometimes employers. Such migrants can be divided into three groups. The first group, daily migrants, is composed from citizens of other countries from near-border regions delivering vegetables using passing transport and returning back towards evening, namely increased number of Uzbeks commute to work in the south of Kazakhstan. The second sizable group consists of seasonal workers, such as builders involved in construction works and home repairs services, pickers working in the tobacco fields, shepherds etc. In contrast to two first groups of hired employees, the third group represents illegal foreigners being self-supporting entrepreneurs.

On the one hand illegal migration provides supply of manpower for difficult and dangerous jobs with small wages that reflected on the prices of final products. On the other hand illegal migration has severe consequences such as human trafficking, labor and sexual exploitation, shadow economy, international terrorism.

4.2.6 Internal migration

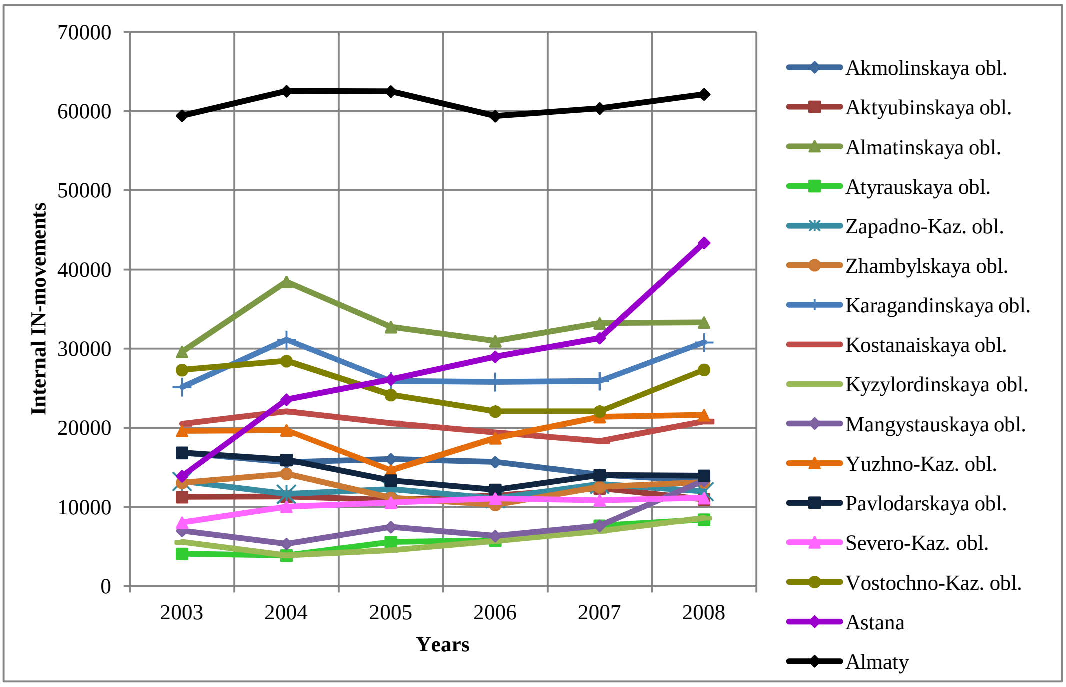
Since the 1990s Kazakhstan has experienced strong internal migration flows. Internal migration events surpass external migration events, it confirms axiom of migration, the less are the distances, the more occur movements. Regarding gender comparison female mobility within country exceeds male's till the age 40-44, while male's migration is slightly increased with retirement specially for age group 60-64 years.

Due to process of urbanization more than 4 million people were involved in migration from countryside to towns for the period 1993-2003. Proportion of Kazakhs in last three years constituted 76 percent (Bapakova 2010). Almaty, the commercial center of the country with registered 1.5 million population and assumed 1 million of unregistered population, where are situated many educational, cultural and financial organizations, is the biggest attractor. Fig. 7 shows outstanding position of Almaty, while Fig. 8 reviles that Out-flow from this city is also high, maybe due to high competition between new comers and local population as well as saturated labor market and house limitation. On the second place is Astana, fast-growing capital of young independent country, attracts all people from businessmen, civil servants, specialists to people without any qualifications. Fig. 7 shows steep ascent of In-flow movements, while according to Fig. 8 Out-flow movements from this city are increasing still relatively low. According to Fig. 7 in 2004 internal In-flows in Almaty (62518 movements) were in 16 times higher than internal In-flows in Kyzylordinskaya oblast (3883 movements), in 2008 the biggest difference was between Almaty and Atyrauskaya oblast more than 7 times, 62113 and 8419 movements correspondingly. According to Fig. 8 all regions of origin show steady increase of internal Out-flows, in 2008 Out-flows in Almaty were more than 5 times greater than in Atyrauskaya oblast, 44280 and 8293 movements, correspondingly.

On the one hand, urbanization has positive effect and means modernization or development, when young people come to big cities to study in institutions and build their careers in prosper companies, and negative effects, when there are no jobs, no infrastructure in remote rural areas and people force to move to cities. Another reason of internal movements is depletion of minerals deposits, small part of migrants moved due to ecological problems of Semipalatinsk Test Site and shallowing Aral Sea.

In general, internal migration flows are forced, and affect all age groups, because many remote settlements are in poor conditions, many buildings and houses are ruined, others houses look desolate among such ruins, people of working ages along with their families compel to migrate to cities, towns, in abandoned settlements live usually pensioners, old or sick persons, or people who can't move due to some reasons. Internal migration caused the problem of spontaneous settlements near big cities, in Almaty well-known settlements Shanyrak and Bakai, in Astana settlement Ondiris.

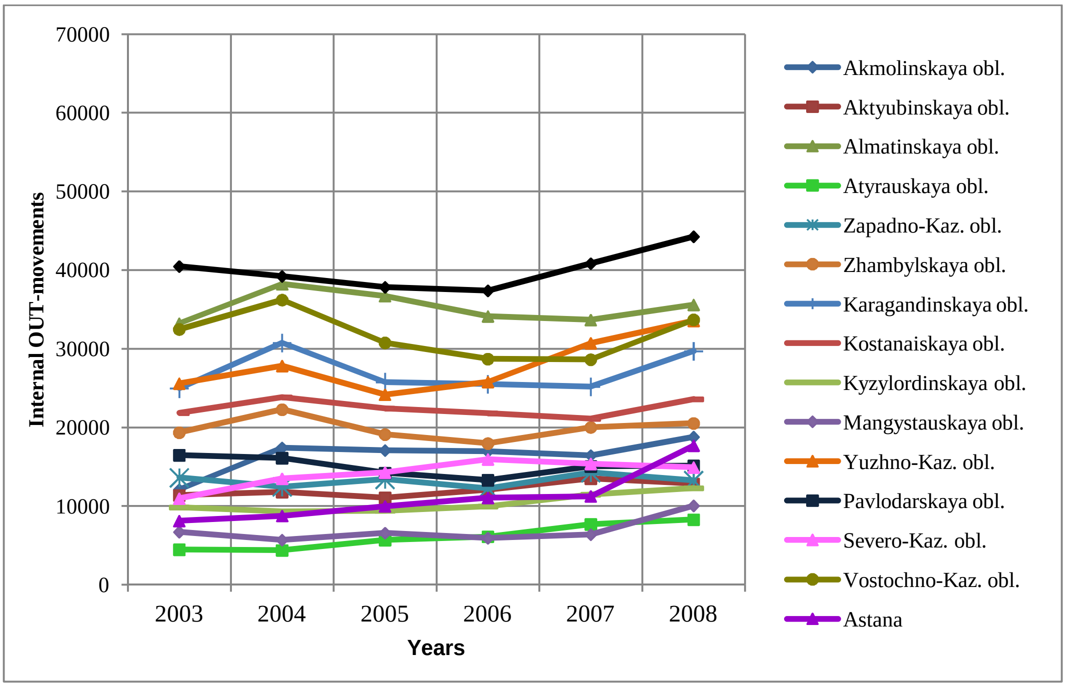
Fig. 7 – Internal In-flow movements by regions, Kazakhstan, 2003-2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Note: Regional Internal In-movements consist of interregional incomings plus regional incomings, in other words, arrivals from other regions and arrivals from the same region.

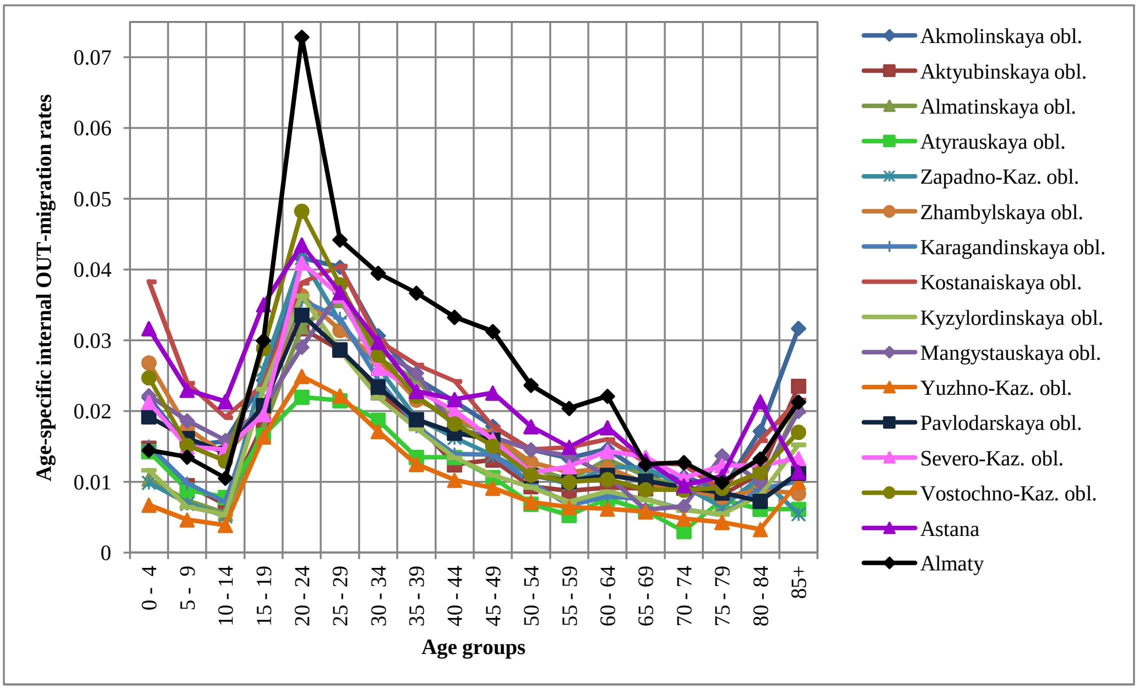
Fig. 8 – Internal Out-flow movements for the period 2003-2008 by regions of Kazakhstan.



Source of data: The Agency on Statistics of the Republic of Kazakhstan

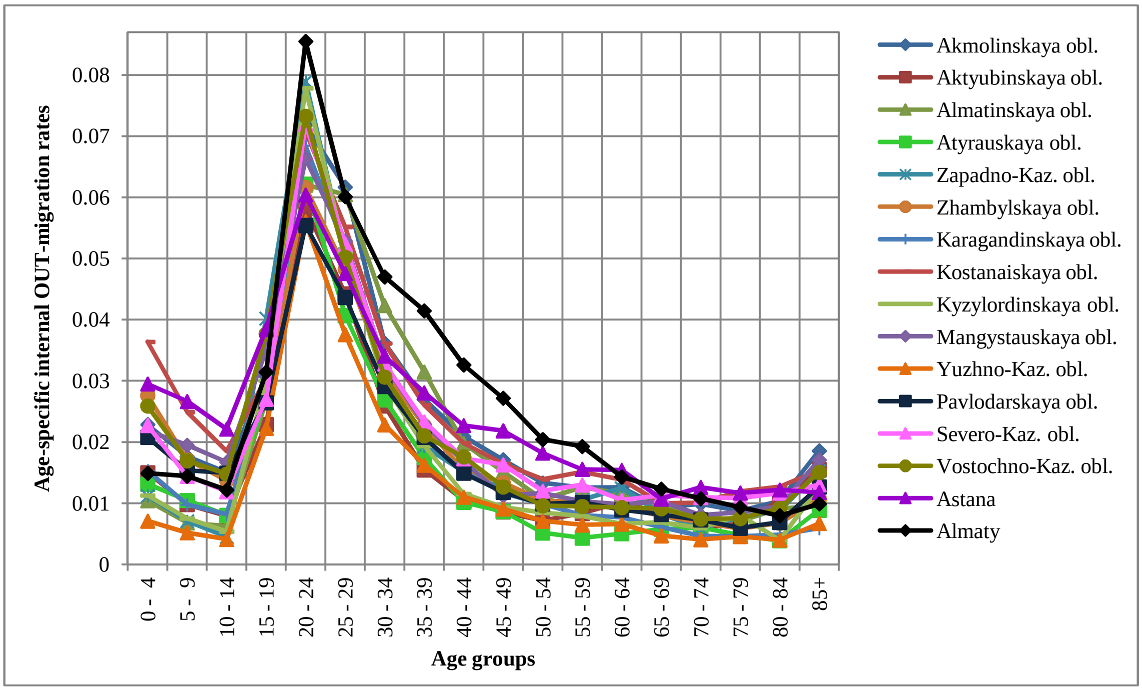
Note: Regional Internal Out-movements consist of interregional departures plus regional departures, in other words, departures for other regions and departures without leaving region.

Fig. 9 – Age-specific internal Out-migration rates by regions, Kazakhstan, 2008, males



Source of data: The Agency on Statistics of the Republic of Kazakhstan

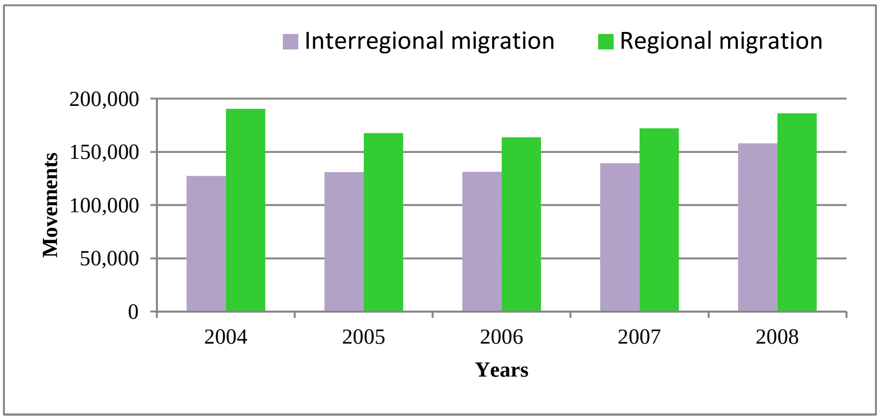
Fig. 10–Age-specific internal Out-migration rates, Kazakhstan, 2008, females



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Internal migration data consists of two types of data: interregional migration, when people move between administrative regions, and regional migration, when people move within region. Figure 11 shows steady increase of interregional migration during the period 2004-2008.

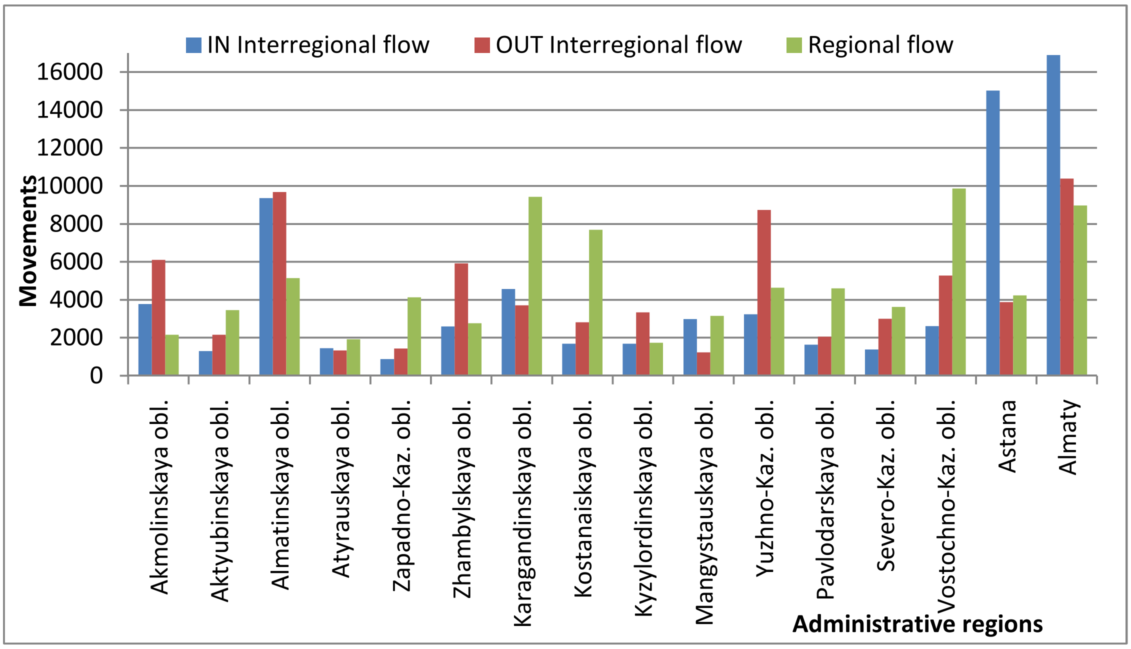
Fig. 11 – Proportions of interregional and regional migration of Kazakhstan, 2004-2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

At the same time proportions of interregional and regional migration vary considerable among administrative regions. For example, Fig. 12 presents proportions of regional flows, interregional In- and Out-flows separately for every region of Kazakhstan in year 2008.

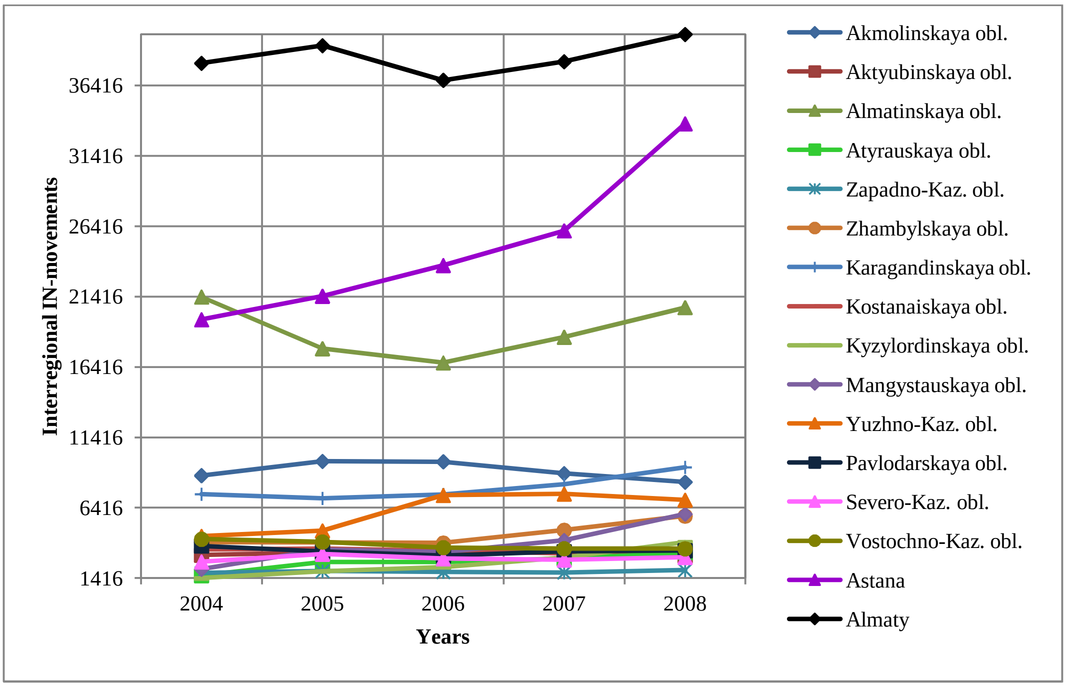
Fig. 12 – Proportions of interregional and regional migration by regions, Kazakhstan, 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Since for multiregional projections of administrative regions the interregional migration has a top priority Fig. 13 and Fig. 14 present pure interregional In- and Out-movements by regions for the period 2004-2008. According to Figure13 there are three regions of destinations absorbing interregional migrants: Almaty, Astana, and Almatinskaya oblast. Outflows are less sizeable; Figure 14 reveals 6 regions of origin: Almatinskaya oblast, Almaty, Yuzhno-Kazakhstanskaya oblast, Akmolinskaya, Zhambylskaya oblast, and Vostochno-Kazakhstanskaya oblast. In general, both figures shows increase of interregional mobility expressed through of In- and Out- movements.

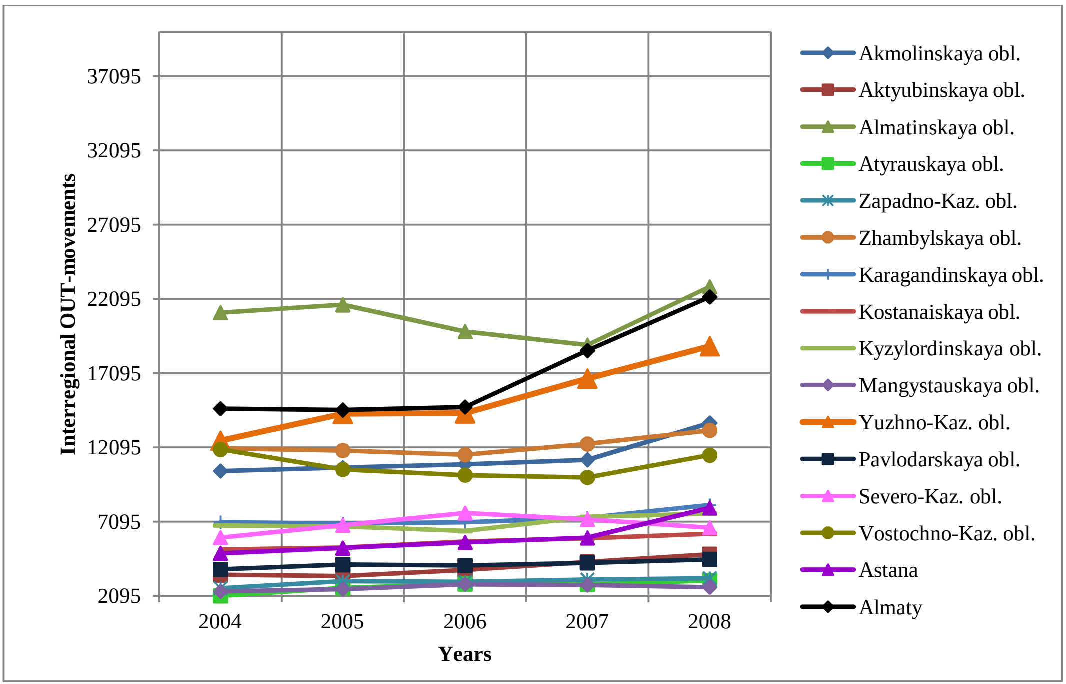
Fig. 13 – Interregional In-movements by regions, Kazakhstan, 2004-2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Note: Interregional In-movements consist of interregional incomings, in other words, when particular region gains aggregate sum of arrivals from other regions.

Fig. 14 – Interregional Out- movements, Kazakhstan, 2004-2008

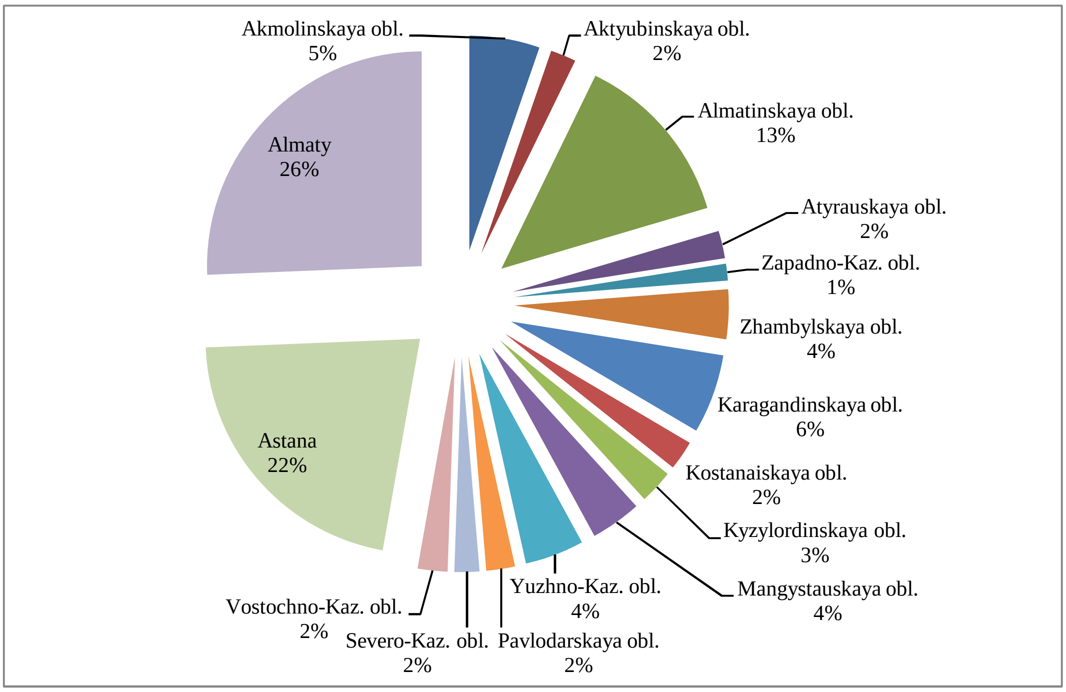


Source of data: The Agency on Statistics of the Republic of Kazakhstan

Note: Interregional Out-movements consist of interregional departures, in other words, aggregate sum of departures for other regions.

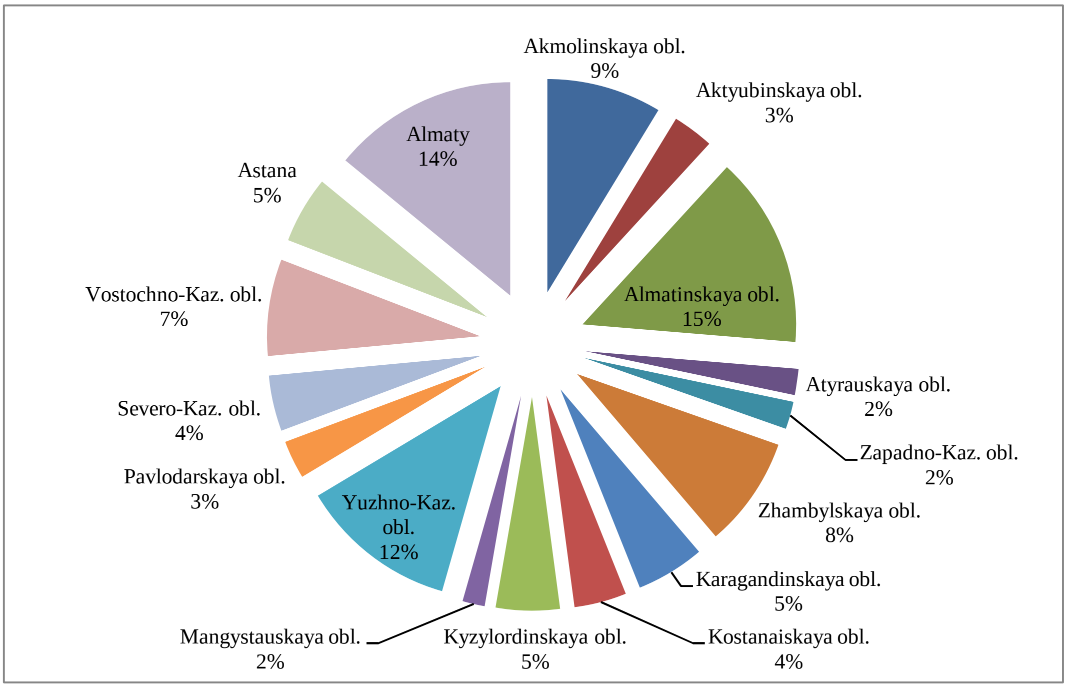
The previous two pictures show dynamic development of interregional migration, whereas the next two are snapshots in motion-picture films or diagnostic cross-sections.

Fig. 15 –Proportions of interregional In-movements by regions, Kazakhstan, 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig.16 – Proportions of interregional Out-movements by regions, Kazakhstan, 2008.



Source of data: The Agency on Statistics of the Republic of Kazakhstan

The Fig. 15 shows impact of every region in the distribution of interregional In-movements, while Fig. 16 reveals percentage proportions of interregional Out-stream. According to these pictures three regions, in 2008 year Almaty, Astana, Almatinskaya oblast registered 61 percent of In-migration movements, while the other 13 regions of destination received 39 percent. At the same time Almatinskaya oblast made a contribution of 14 percent to the pool of Out-stream, while share of Almaty was 14 percent, and Yuzhno-Kazakhstanskaya oblast contributed 12 percent.

4.2.7 Ongoing work

According to statements of official persons in recent newspaper publications attention will be paid to measures on strengthening of immigration control and regularizing internal migration through of tightening residency rules and redistribution of migrants. New changes are related to distribution of oralmans with orientation to the north regions, implementation of labor migrant quotas by countries and permissions given to labor migrants independently on employers, as well as reducing the quotas due to world economic crisis.

Chapter 5

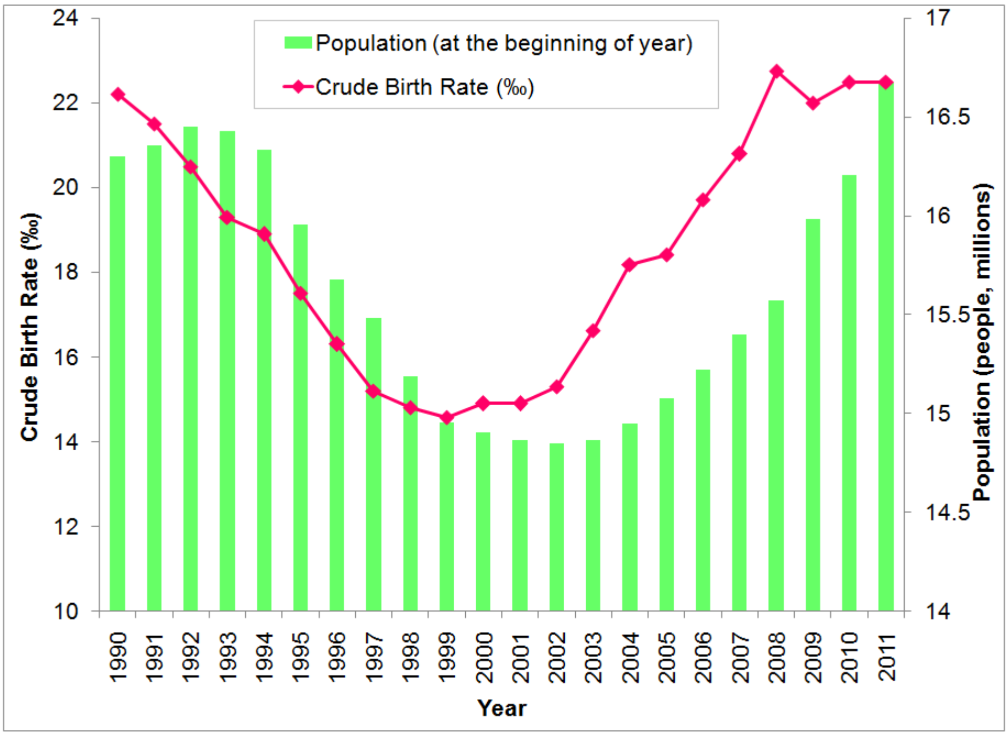
Fertility and mortality in the regions of Kazakhstan

Fertility and mortality are components responsible for natural growth of population. This chapter considers both of these components, their spatial differentiation on the vast territory of Kazakhstan, and how they were developing during previous time period.

5.1 Fertility in the regions of Kazakhstan

Kazakhstan is an example of the country that has overcome previous drop of fertility after the disintegration of the Soviet Union. Political stability, growth of economy as well as bonus system encouraging births have contributed to increased proportion of new-born generations (Fig.17).

Fig. 17 – Crude birth rate and population at the beginning of year, Kazakhstan, 1990-2011



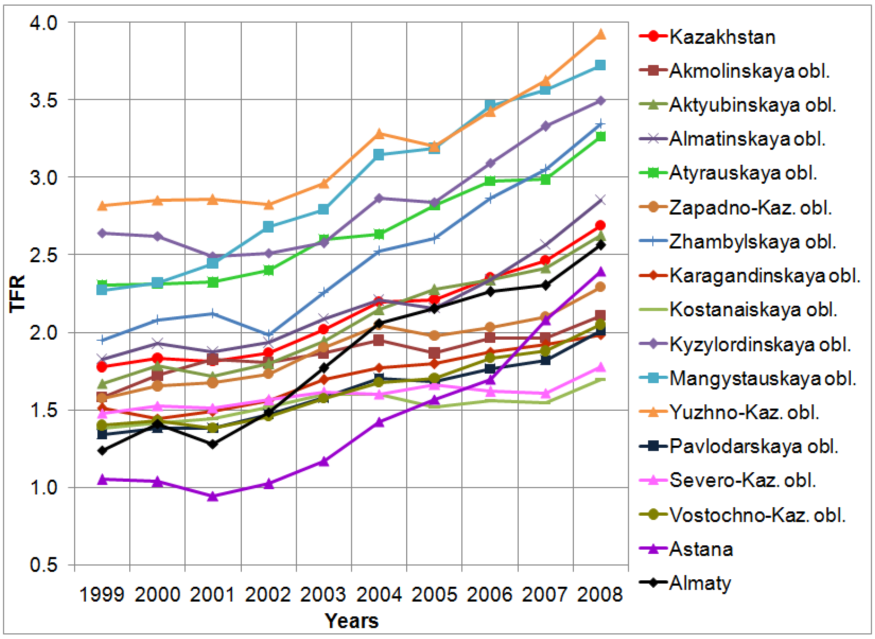
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig.18 presents trends of regional total fertility rates with possibility to compare with total fertility rates for the country level for the period 1999 -2008. It is obvious that all regions have demonstrated increase of fertility for this period.

The next several groups of pictures reveal more detailed analysis about spatially and dynamically distributed information concerning countrywide population reproduction.

Bunch of charts displayed from Fig. 19 to Fig. 24 show trends of regional age-specific fertility rates separately for each age group for the period 1999 -2008. As it seen from Fig.19, regional ages-specific rates for the youngest age group 15-19 are characterized by fluctuations with noticeable concavity in the middle of period. Possible explanation is that in such transitional period young generations are more susceptible to changes in cultural values, new west style of life, and they are studying and preparing to move to cities with educational institutions. The next age group 20-24 displays high level of fertility but progress not sharp rather moderate. The trends for the three age-groups 25-29, 30-34, 35-39 demonstrated steep slopes along the whole period as a confirmation of ongoing demographic transition when fertility shifted to older ages due to prolong period of study and building career, other economical and social reasons, such as absence of own house and high rent payment. Age-specific fertility curves for age group 40-44 indicate slight increase over time whereas fluctuations in trends for the last age-group 45-49 are minute and insignificant due to small occurrence of such events for this age group.

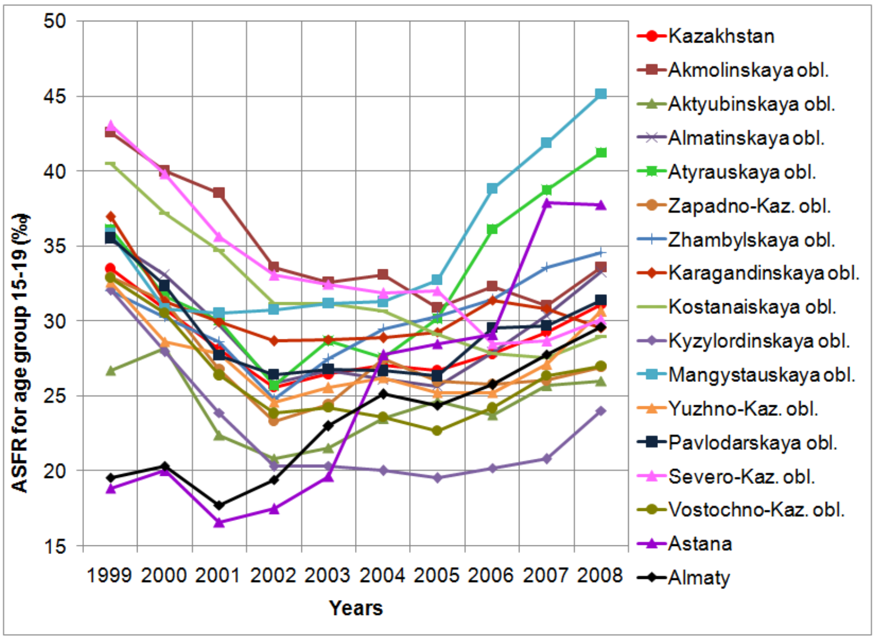
Fig.18 – Total Fertility Rates, by regions , 1999-2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

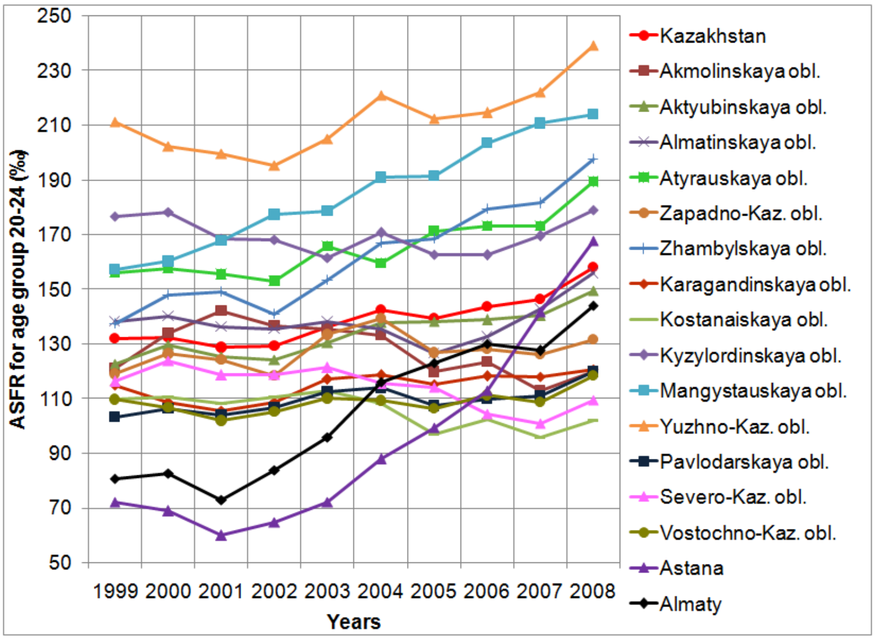
.

Fig.19 – Age-specific fertility rates by regions of Kazakhstan (‰), age group 15-19, period 1999 - 2008



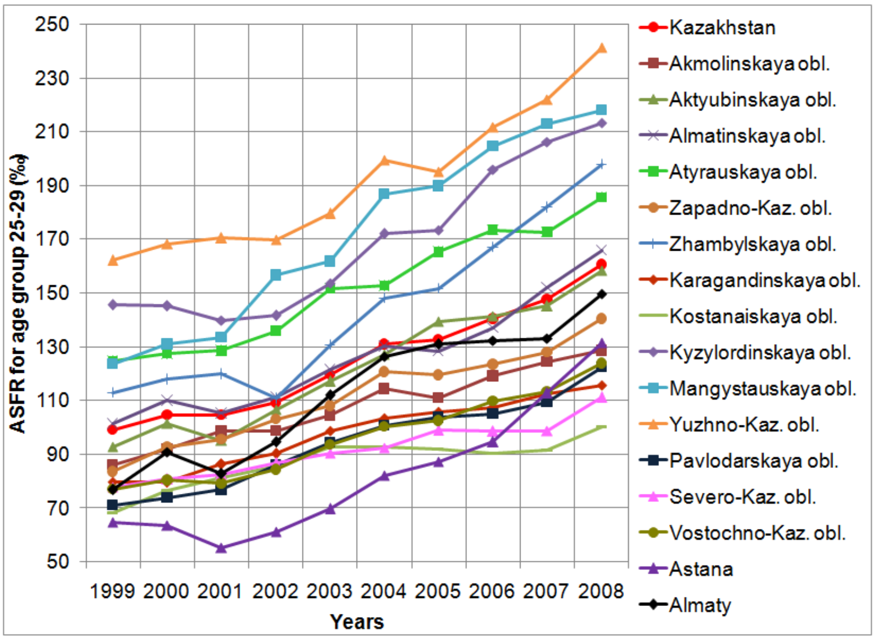
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 20 – Age-specific fertility rates by regions of Kazakhstan (‰), age group 20-24, period 1999 - 2008



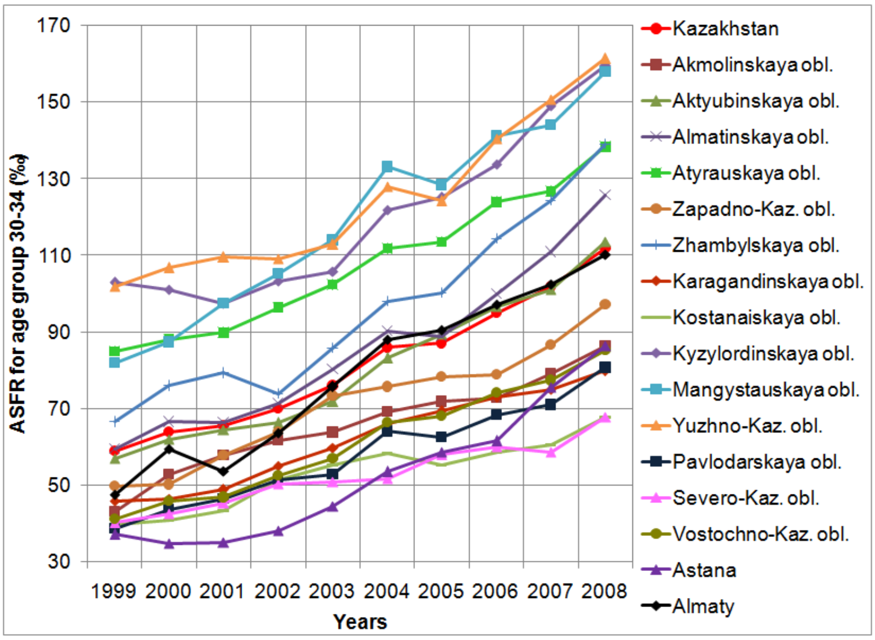
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 21 – Age-specific fertility rates by regions of Kazakhstan (‰), age group 25-29, period 1999 - 2008



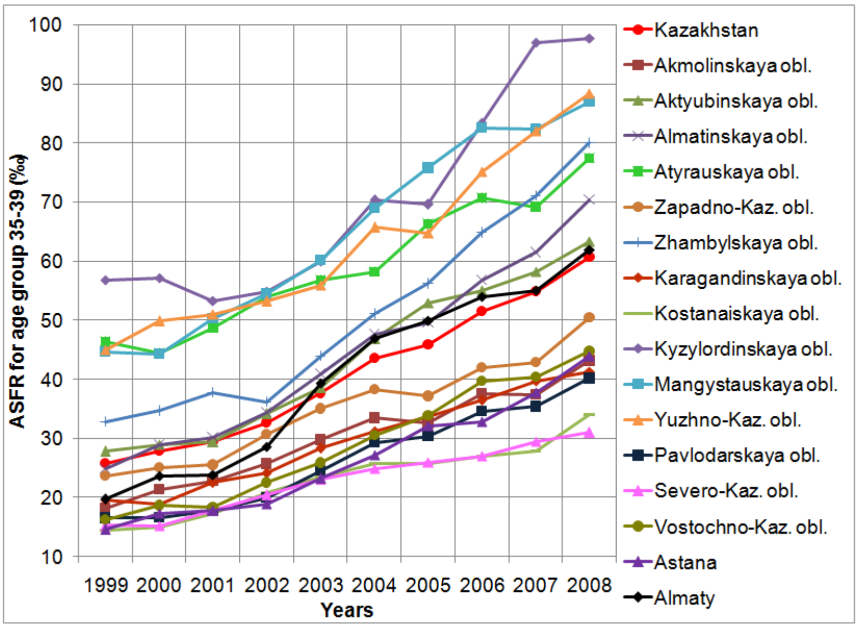
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 22– Age-specific fertility rates by regions of Kazakhstan (‰), age group 30-34, period 1999 - 2008



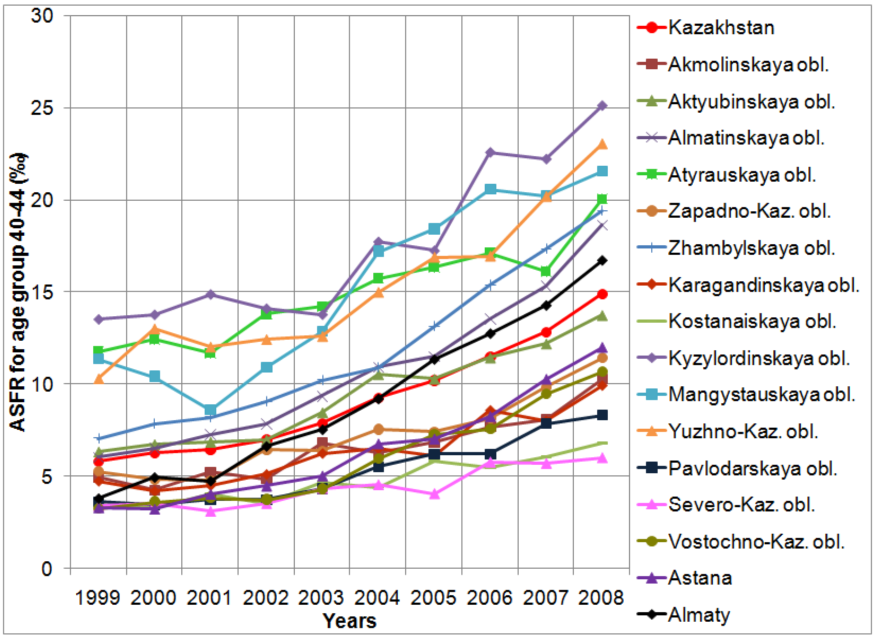
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 23– Age-specific fertility rates by regions of Kazakhstan (‰), age group 35-39, period 1999 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

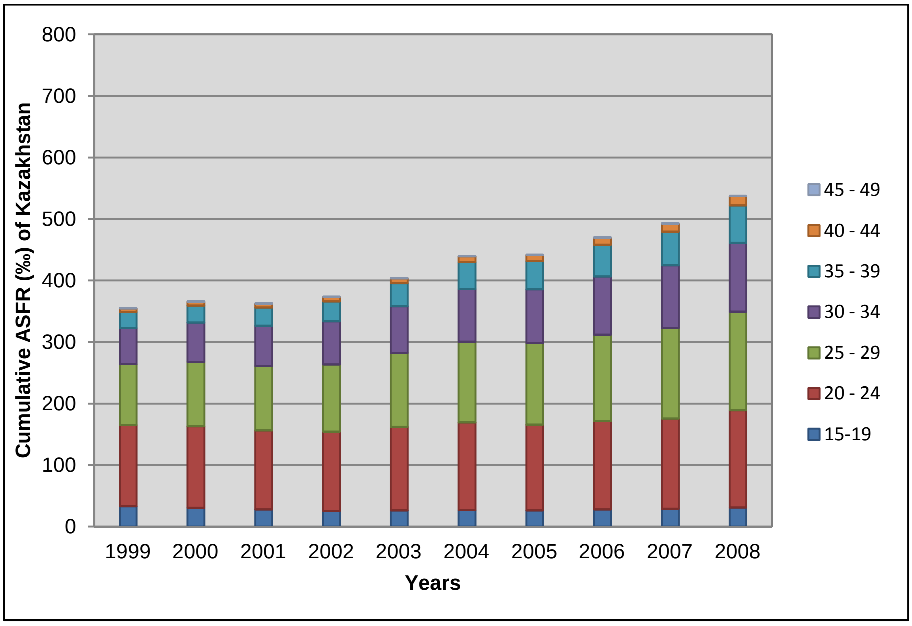
Fig.24– Age-specific fertility rates by regions of Kazakhstan (‰), age group 40-44, period 1999 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

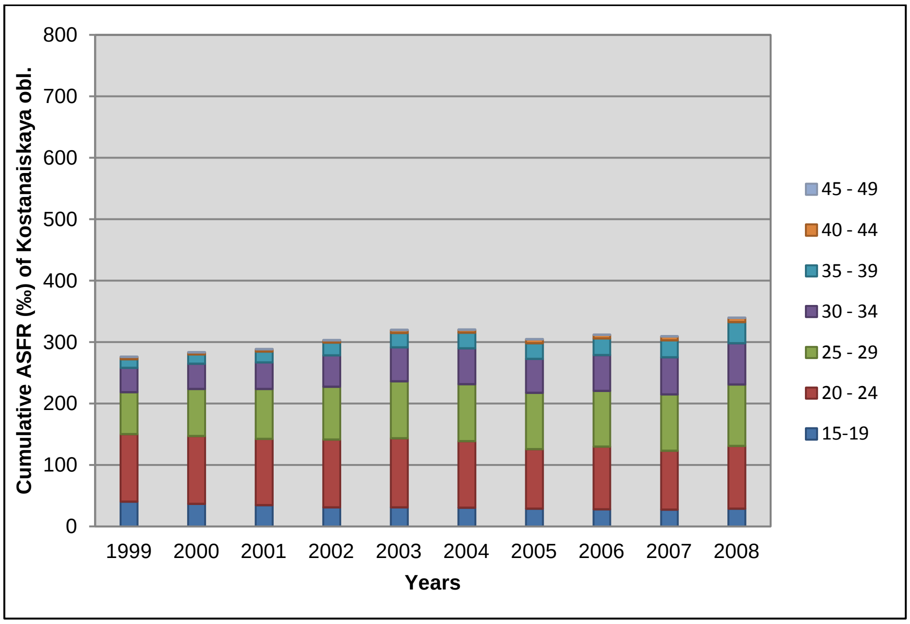
The following pictures from Fig.25 to Fig. 27 present cumulative age-specific fertility rates developing through whole the period separately for each region. This is important since they indicate shares of age groups and how they changes in the specified regions. Fig. 25 gives picture about Kazakhstan as a baseline for two contrasting regions: Kostanaiskaya oblast and Mangystauskaya oblast. The first province is characterized by low fertility without significant changes along time, when Mangaustauskaya oblast has shown more distinctive growths of fertility.

Fig. 25–Cumulative Age-specific fertility rates of Kazakhstan (‰), period 1999 - 2008



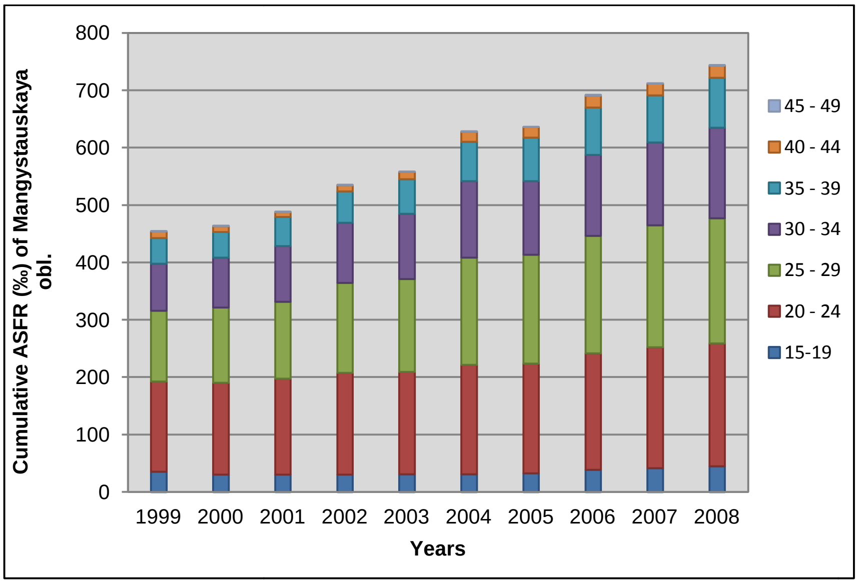
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 26–Cumulative Age-specific fertility rates of Kostanaiskaya obl. (‰), period 1999 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

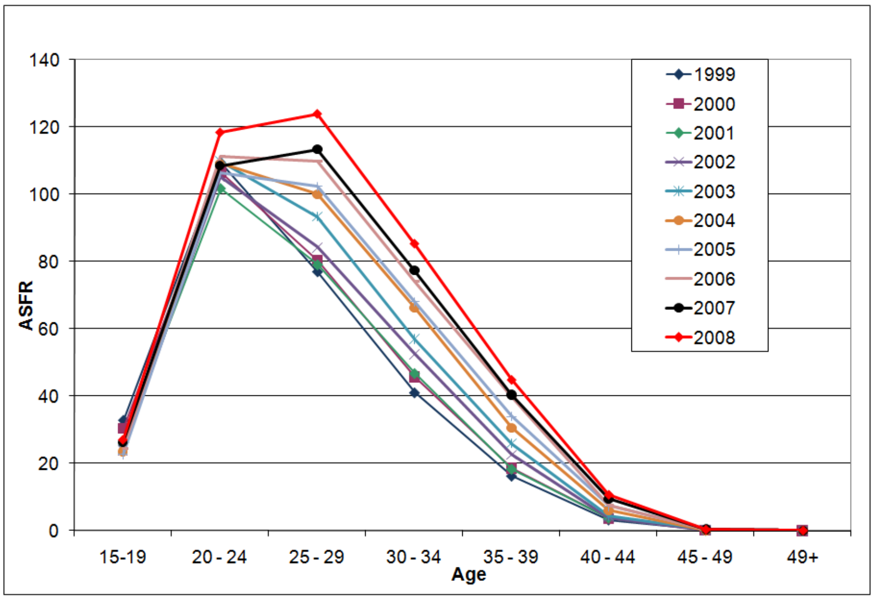
Fig. 27–Cumulative Age-specific fertility rates of Mangystauskaya obl. (‰), period 1999 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 28 shows transformations of the age specific fertility curves for Vostochno-Kazakhstanskaya as an example of changing patterns of age specific fertility rates. In this province fertility level is not high with moderate past fertlity increase along the time. Peak of fertility in 20-24 ages at the beginning of observation period has moved to the peak in ages 25-29 with following modifications in the shape of the fertility curves with increasing impact of older age groups. These changing pattern of fertility curves are true for all regions while some of them including Astana still have had peaks at the ages 20-24 which could be explained by influence of east mentality and attachment to the Muslim religion, as well as adaptation to the life in the capital, housing projects, etc.

Fig. 28–Age-specific fertility rates of Vostochno-Kazakhstanskaya obl. (‰), period 1999 – 2008

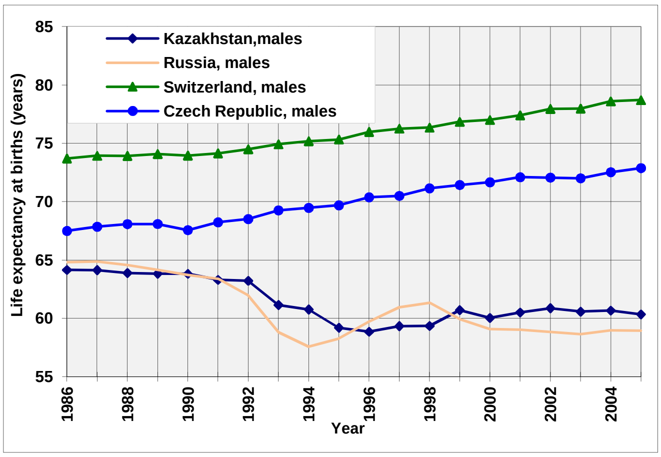


Source of data: The Agency on Statistics of the Republic of Kazakhstan

5.2 Mortality in the regions of Kazakhstan

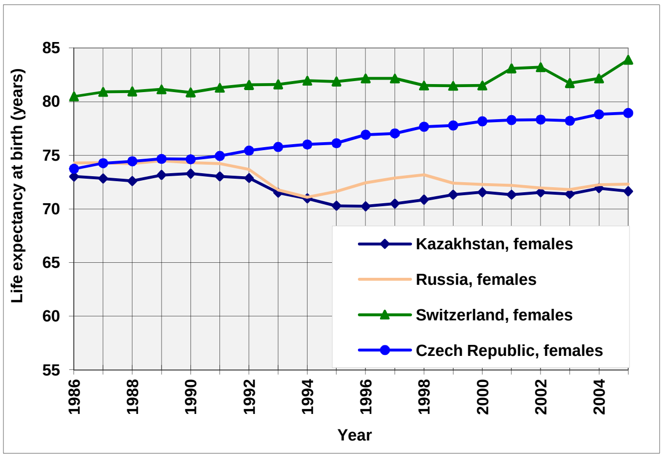
Consequences of political and economic changes and disorderliness in early 90s resulted in the excess of mortality in contradiction to general worldwide tendency of decreasing mortality. Fig.29 and Fig.39 demonstrate trends of life expectancies at birth for the period 1986-2005 for males and females. Underlying data are obtained from the site of the World Health Organization. Alarming trends are observed for Kazakhstan whereas Switzerland displayed very favorable situation from the beginning of interval, also steady progress can be seen in Czech Republic. Transition period, consequences of the reconstruction of health system, unemployment of medical staff and periods of shortage of medications, as well many other socio-economic reasons have triggered mortality increase in Kazakhstan.

Fig.29 – Life expectancy at birth in selected countries for males, 1986-2005



Source of data: World Health Organization

Fig. 30 – Life expectancy at birth in selected countries for females, 1986-2005

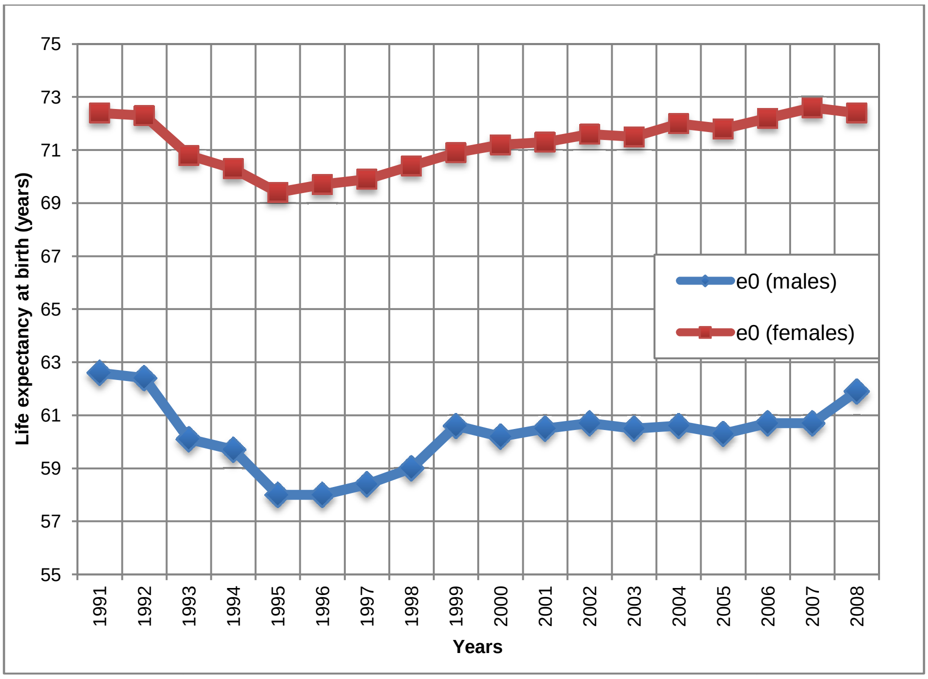


Source of data: World Health Organization

The towering growth of mortality was characterized by high mortality of men of working ages. For females mortality also increased but not so strong like for males, displaying quite big difference between genders in mortality. Fig. 31 shows the trends of life expectancies for males and females of Kazakhstan and displays how they fluctuate along the period 1991-2008.

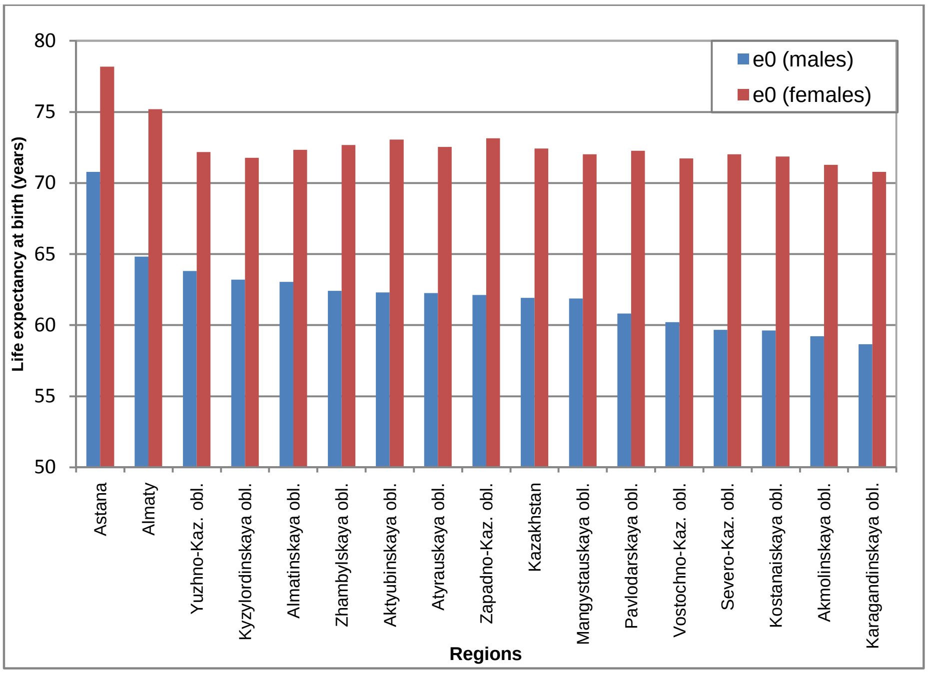
Fig. 32 displays life expectancies at birth by regions of Kazakhstan and emphases how big is difference between genders for mortality. The highest mortality corresponds to Karagandinskaya oblast, the lowest mortality level belongs to Astana.

Fig. 31– Life expectancy at birth of Kazakhstan for males and females, period 1991 - 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Fig. 32 – Life expectancy at birth by regions of Kazakhstan, 2008 year

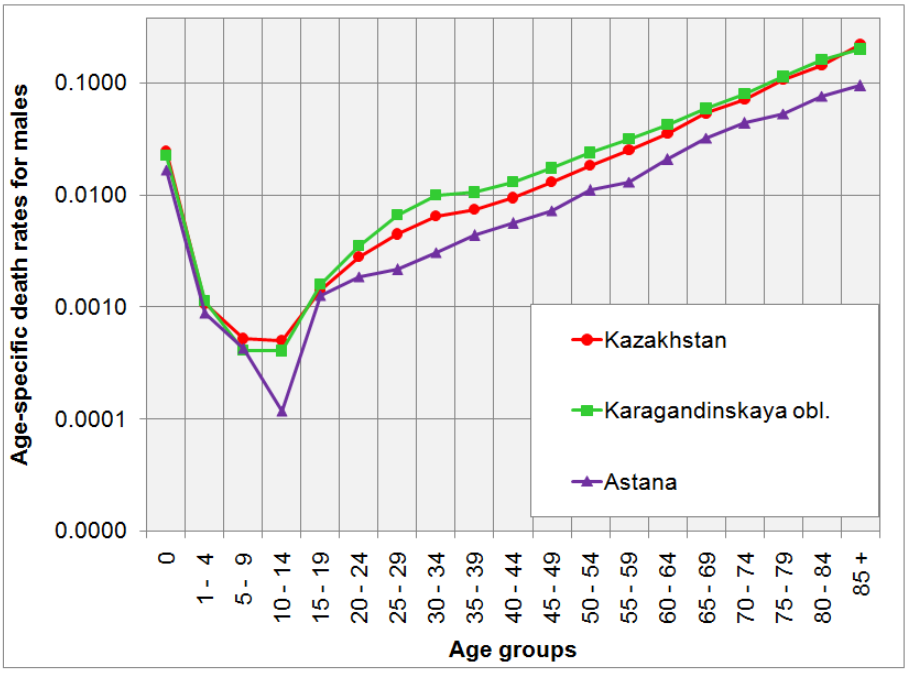


Notes: Sorted according to life expectancy at birth for males in descending order
Source of data: The Agency on Statistics of the Republic of Kazakhstan

Since it is important for regional population projections to know about shapes of age-specific death rates the following picture Fig. 33 demonstrates age-specific death rates for

Kazakhstan, as average level, and age-specific death rates for two administrative divisions with the highest and the lowest life expectancy at birth. There are obvious bumps in the ASDR curves for men of working ages. Such differences in shapes could be explained by the fact that Karagandinskaya oblast is industrial region with following consequences such as occupational diseases, ecological problems, alcohol and tobacco consumption, while Astana, as a capital, possess medical facilities and staff, support and customers able to defray costs.

Fig. 33– Age-specific death rates, males, Kazakhstan, Karagandinskaya oblast, Astana, 2008



Source of data: The Agency on Statistics of the Republic of Kazakhstan

Chapter 6

Methodological aspects of multiregional approach

6.1 Preparing interregional transition migration data

Using of multiregional model requires origin-destination migration matrix. Willekens, Por, Raquillet (1981) described the estimation procedures in case of lack of data for inferring migration patterns from marginal sums of two- or n-dimensional arrays. In two-dimensional case these marginal sums usually represent row or column sums.

Bi- and multi-proportional adjustment methods include the entropy maximization and the I-divergence problems. The entropy method tries to determine the "most probable" elements of array under the given marginal conditions when no initial array is known a priory. The entropy maximizing method was introduced in regional science by Wilson (1967, 1970). The I-divergence minimizing problem tries to estimate an "a posteriori" array which is as close as possible to an "a priori" array, and which satisfies some given constraints (row and column sums).The distance function used to measure the "closeness" of the arrays is the I-divergence or Kullback-Leibler information number (Kullback 1959), also called information for discrimination, information gain, or entropy of a posteriori distribution relative to an a priori distribution (Renyi 1970). The entropy method can be considered a special case of the more general I-divergence method if the initial estimates of the elements of the array are set equal to any scalar value, for example, unity. The entropy method is more suited for estimating detailed migration flows on the basis of an total information only, the I-divergence method is preferable for updating migration-flow tables (Willekens, Por, Raquillet 1981).

In case of Kazakhstan deriving process of age-sex specific flow matrix is divided into two phases. On the first phase the Three face algorithm is applied to disaggregate flow matrix by gender. On the second phase disaggregation by age group could be done by using One-face, One-edge algorithm using age structure of interregional migration at the national level, which is a very simple and fast method. Alternative method for inferring age profiles is again through of applying the Three face algorithm by using three matrices which gives more precise results.

One-face, One-edge algorithm was applied for inferring age profiles for 19 age groups, these origin-destination data disaggregated by sex and age groups were further applied for population projection of 16 administrative regions.

Since the Three face algorithm requires more data and produces better results this algorithm was applied for disaggregation by 101 age groups. Obtained age-sex specific origin-destination data were furthers applied to project populations in four macroregions of Kazakhstan.

6.1.1 Disaggregation by sex

In the first step, marginal flow matrix is disaggregated separately for every gender using the vital migration statistic data for the year 2008. The three face algorithm applied for this purpose uses data from three available matrices:

C_{ij} – marginal flow matrix for both genders combined.

A_{jk} – matrix of arrivals by region, j- destination region from 1 to 16, k – from 1 to 2 (males or females);

B_{ik} – matrix of departures by region, I – sending region from 1 to 16, k – gender.

Tab. 4 presents values of matrix of arrivals A_{jk} and matrix of departures B_{ik}.

Tab. 4- Matrix of arrivals Ajk, and matrix of departures Bik, 2008

	Aj1	Aj2	Bi1	Bi2
Akmolinskaya obl.	3780	4472	6102	7638
Aktyubinskaya obl.	1303	1714	2158	2742
Almatinskaya obl.	9363	11281	9675	13251
Atyrauskaya obl.	1449	1802	1330	1795
Zapadno-Kaz. obl.	870	1116	1440	1847
Zhambylskaya obl.	2602	3261	5927	7310
Karagandinskaya obl.	4578	4719	3716	4488
Kostanaiskaya obl.	1680	1817	2818	3446
Kyzylordinskaya obl.	1686	2291	3347	4293
Mangystauskaya obl.	2985	2991	1232	1451
Yuzhno-Kaz. obl.	3233	3739	8743	10171
Pavlodarskaya obl.	1640	1734	2062	2491
Severo-Kaz. obl.	1378	1526	2996	3695
Vostochno-Kaz. obl.	2609	2603	5272	6296
Astana	15035	18661	3884	4135
Almaty	16900	23156	10389	11834

The purpose of the algorithm is to disaggregate marginal flow matrix into two marginal flow matrices for males and females, which means to find M_{ijk}, where k changes from 1 to 2.

The Three face algorithm consists of the following steps:

At the beginning of algorithm it is assumed that s=0 and $m_{ijk}^{(0)} = 1$ for any i, j, k. That means initial elements of flow matrix equal to unity.

Step 1. Adjusting M_{ijk} using knowledge of matrix C_{ij}

$$m_{ijk}^{(3s+1)} = \frac{m_{ijk}^{(3s)} \cdot c_{ij}}{\sum_{k=1}^{l} m_{ijk}^{(3s)}} \quad \text{for any } i, j$$

Step 2. Next adjustment of M_{ijk} is based on the data from matrix of arrivals A_{jk}

$$m_{ijk}^{(3s+2)} = \frac{m_{ijk}^{(3s+1)} \cdot a_{jk}}{\sum_{i=1}^{n} m_{ijk}^{(3s+1)}}$$

Step 3. The last corrections is made through of matrix of departures B_{ik}

$$m_{ijk}^{(3s+3)} = \frac{m_{ijk}^{(3s+2)} \cdot b_{ik}}{\sum_{j=1}^{m} m_{ijk}^{(3s+2)}}$$

If $\left|\frac{m_{ijk}^{(3s+3)}}{m_{ijk}^{(3s+2)}} - 1\right| \leq \varepsilon$ fulfilled then Stop, otherwise s=s+1 and return to Step 1 (Willekens, Por, Raquillet 1981).

The algorithm is repeated 5 times in our case and as a condition for continuation is used: max $(m_{ijk}^{(3s+3)} - m_{ijk}^{(3s+2)}) \leq 0.0001$. As a result two gender specific flow matrices are received and displayed in Tab. 5 and Tab. 6.

6.1.2 Disaggregation by age groups using One-Face One-Edge algorithm

On the second phase One-face, One-edge algorithm (1F1E) is applied for disaggregation by age, where marginal flow matrix is available as well as age structure of interregional migration at the national level. The disaggregation is applied separately for males and females flow matrices using the following formulas (Willekens, Por, Raquillet 1981):

$m_{ijk} = c_{ij} \times \frac{m_{..k}}{m_{...}}$,

where C_{ij} is marginal flow matrix, and here index k denotes age group.

The ratio $\frac{m_{..k}}{m_{...}}$ is the national age composition of the interregional migrants (Tab.7) and it is applied to all values of the flow matrices $m_{ij.}$ hence the age structure is assumed to be uniform for all flows, but gender specific. After the conducting the first and the second phases the flow matrix was disaggregated by sex and age, and can be used in multiregional projections.

Tab. 5 - Disaggregated flow matrix for females, 2008

	Destination regions																
Origin regions	Akmolinskaya obl.	Aktyubinskaya obl.	Almatinskaya obl.	Atyrauskaya obl.	Zapadno-Kaz. obl.	Zhambylskaya obl.	Karagandinskaya obl.	Kostanaiskaya obl.	Kyzylordinskaya obl.	Mangystauskaya obl.	Yuzhno-Kaz. obl.	Pavlodarskaya obl.	Severo-Kaz. obl.	Vostochno-Kaz. obl.	Astana	Almaty	Total
Akmolinskaya obl.	0	28	235	24	13	65	544	392	28	45	172	184	585	72	4997	253	7638
Aktyubinskaya obl.	48	0	227	335	208	66	65	54	276	465	123	24	8	32	345	465	2742
Almatinskaya obl.	178	97	0	110	95	383	419	82	370	185	207	92	64	534	513	9922	13251
Atyrauskaya obl.	16	237	139	0	235	56	16	10	40	380	85	8	6	20	174	374	1795
Zapadno-Kaz. obl.	22	283	148	287	0	39	59	22	24	177	90	18	13	25	291	350	1847
Zhambylskaya obl.	134	60	1138	95	31	0	396	59	177	267	495	54	31	113	1441	2820	7310
Karagandinskaya obl	390	18	365	34	43	183	0	170	98	63	255	197	154	187	1695	637	4488
Kostanaiskaya obl.	481	47	228	18	16	80	359	0	15	35	66	80	122	44	1550	305	3446
Kyzylordinskaya obl	59	320	551	77	44	213	202	46	0	123	542	22	10	58	819	1208	4293
Mangystauskaya obl.	17	162	148	269	93	57	18	24	36	0	115	11	8	28	137	327	1451
Yuzhno-Kaz. obl.	148	114	1064	144	73	711	645	84	480	730	0	102	87	107	2311	3371	10171
Pavlodarskaya obl.	180	23	169	17	25	38	380	88	23	26	115	0	77	247	725	358	2491
Severo-Kaz. obl.	1480	7	129	12	12	45	344	196	6	21	108	98	0	44	1002	189	3695
Vostochno-Kaz. obl.	144	40	1506	44	26	76	420	59	43	74	80	467	41	0	1395	1884	6296
Astana	990	34	188	40	38	211	445	362	177	59	274	190	214	218	0	695	4135
Almaty	185	245	5045	295	164	1037	407	168	497	341	1013	187	108	874	1267	0	11834
Total	4472	1714	11281	1802	1116	3261	4719	1817	2291	2991	3739	1734	1526	2603	18661	23157	86883

Tab. 6 - Disaggregated flow matrix for males, 2008

		Destination regions																
		Akmolinskaya obl.	Aktyubinskaya obl.	Almatinskaya obl.	Atyrauskaya obl.	Zapadno-Kaz. obl.	Zhambylskaya obl.	Karagandinskaya obl.	Kostanaiskaya obl.	Kyzylordinskaya obl.	Mangystauskaya obl.	Yuzhno-Kaz. obl.	Pavlodarskaya obl.	Severo-Kaz. obl.	Vostochno-Kaz. obl.	Astana	Almaty	Total
Origin regions	Akmolinskaya obl.	0	20	180	19	9	49	500	343	19	44	140	163	504	68	3866	179	6102
	Aktyubinskaya obl.	38	0	173	255	156	49	60	47	189	442	100	21	7	30	266	326	2158
	Almatinskaya obl.	141	71	0	83	71	283	383	71	253	176	168	81	55	501	395	6943	9675
	Atyrauskaya obl.	12	162	99	0	166	38	14	8	26	339	65	7	4	18	125	246	1330
	Zapadno-Kaz. obl.	17	208	113	219	0	29	54	19	17	169	73	15	11	23	225	247	1440
	Zhambylskaya obl.	113	47	924	76	25	0	384	54	129	270	428	51	28	113	1181	2103	5927
	Karagandinskaya obl	328	14	295	27	34	143	0	156	72	64	219	183	140	186	1382	472	3716
	Kostanaiskaya obl.	395	35	180	15	13	61	339	0	10	35	56	73	108	42	1235	221	2818
	Kyzylordinskaya obl	48	237	428	60	34	159	188	41	0	119	448	20	8	55	641	860	3347
	Mangystauskaya obl.	15	133	128	229	79	48	19	24	28	0	106	11	7	30	119	257	1232
	Yuzhno-Kaz. obl.	132	93	912	122	62	589	660	81	370	780	0	101	84	113	1995	2649	8743
	Pavlodarskaya obl.	147	17	133	14	19	29	355	79	16	25	96	0	67	237	571	257	2062
	Severo-Kaz. obl.	1190	6	100	10	10	34	319	173	5	20	88	87	0	42	780	134	2996
	Vostochno-Kaz. obl.	124	31	1252	36	22	61	417	56	32	76	70	448	38	0	1169	1437	5272
	Astana	916	29	167	36	33	181	473	367	142	66	259	196	213	239	0	567	3884
	Almaty	164	199	4279	249	137	850	413	162	379	360	916	183	102	912	1084	0	10389
	Total	3780	1303	9363	1449	870	2602	4578	1680	1686	2985	3233	1640	1378	2609	15035	16899	71091

Tab. 7- National age composition of the interregional migrants by gender, 2008

	$\frac{m_{..k}}{m_{...}}$ **(males)**	$\frac{m_{..k}}{m_{...}}$ **(females)**
0	0.016	0.012
1 - 4	0.047	0.034
5 - 9	0.043	0.033
10 - 14	0.039	0.030
15 - 19	0.111	0.108
20 - 24	0.202	0.262
25 - 29	0.147	0.166
30 - 34	0.103	0.100
35 - 39	0.077	0.069
40 - 44	0.062	0.049
45 - 49	0.056	0.043
50 - 54	0.034	0.028
55 - 59	0.023	0.023
60 - 64	0.013	0.012
65 - 69	0.011	0.012
70 - 74	0.008	0.009
75 - 79	0.004	0.006
80 - 84	0.003	0.004
85+	0.001	0.002
Total	1	1

6.1.3 Disaggregation by age groups using the Three Face algorithm

Firstly, from regional age-specific data on arrivals it is necessary to subtract immigration data, as a result internal migration data on arrivals are obtained, then they are converted to age-specific proportions of internal regional arrivals, and these proportions is assumed to be similar to interregional age schedule on arrivals. Every vector of obtained regional age-specific proportions is multiplied by corresponding total interregional arrivals, and 16 vectors of age-specific interregional arrivals are received, these 16 vectors comprise matrix of arrivals A_{jk} where index j denotes destination region, k means age group from 1 to 101.

Secondly, for obtaining departure matrix B_{ik}, where index i is a sending region, and k specifies age group, it is necessary from regional age-specific data on departures subtract emigration data, and internal age-specific data on departures are received. Again, they are converted to age-specific proportions, multiplied by appropriate total interregional departures. Obtained 16 vectors of interregional departures constitute departure matrix B_{ik}.

Thirdly, for inferring age- specific origin-destination flow data using marginal flow matrix, arrival matrix and departure matrix Three Face algorithm is applied. This algorithm is described by Willekens, Por, Raquillet (1981).

There are three available matrices:

C_{ij} – marginal flow matrix. the early presented Tab. 5 and Tab 6 display origin-destination data separately for females and males.

A_{jk} – arriving matrix, where j – destination region from 1 to 16, k – age group from 1 to 101

B_{ik} – departure matrix, where i – sending region from 1 to 16, k – age group from 1 to 101.

The purpose of the algorithm is to find m_{ijk}.

The whole algorithm is based on the three formulas, namely these formulas are trying to adjust migration by using three available matrices C_{ij}, A_{jk}, and B_{ik}. Algorithm is initialized (s=0) by assuming mijk = 1, i.e. migration events for every direction and every age are assigned to 1.

The first adjustment of migration flows takes into account knowledge of marginal flow matrix C_{ij}:

$$m_{ijk}^{(3s+1)} = \frac{m_{ijk}^{(3s)} \cdot c_{ij}}{\sum_{k=1}^{l} m_{ijk}^{(3s)}} \quad \text{for any i, j}$$

On the second stage migration flows are corrected with taken into consideration knowledge of matrix of arrivals A_{jk}:

$$m_{ijk}^{(3s+2)} = \frac{m_{ijk}^{(3s+1)} \cdot a_{jk}}{\sum_{i=1}^{n} m_{ijk}^{(3s+1)}}$$

Since age profiles of inflows and outflows do not correspond to each other the last adjustment is taken with regard to known matrix of departures B_{ik}:

$$m_{ijk}^{(3s+3)} = \frac{m_{ijk}^{(3s+2)} \cdot b_{ik}}{\sum_{j=1}^{m} m_{ijk}^{(3s+2)}}$$

Whether it worth to stop algorithm or again iteratively repeat three previous adjustments of migration, the next checking formula answers:

$$\left| \frac{m_{ijk}^{(3s+3)}}{m_{ijk}^{(3s+2)}} - 1 \right| \leq \varepsilon.$$

In the end algorithm gives age-specific origin-destination matrix M_{ijk} separately for every gender for the year 2008.

6.2 Imposing consistency in multistate population projections

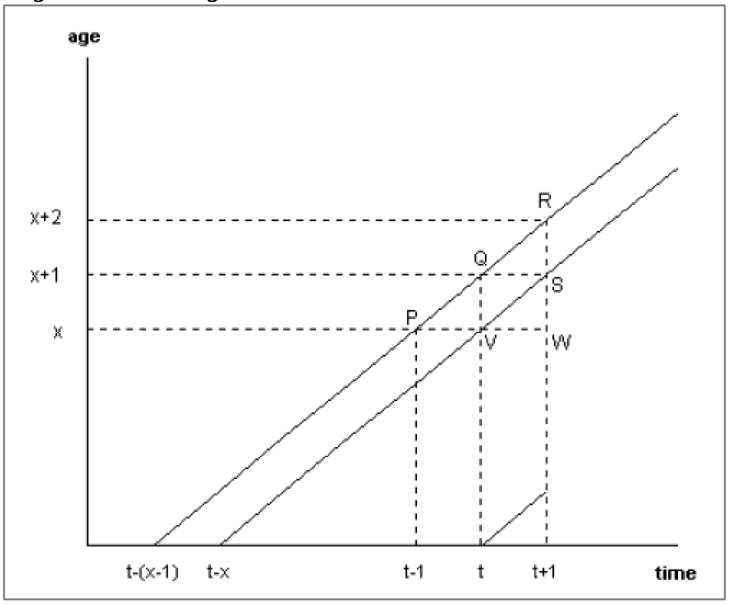
Multistate population projection model uses stock and flow data. Stock data are defined by numbers of persons in each population category, whereas flow data define number of demographic events that occur in the unit projection interval, between t and t+1. Demographic events can be divided into two broad category: internal events and external events. Internal events are characterized by internal position of the origin and internal position of the destination. External events are caused by mortality, fertility and external migration forces and composed of exits events, birth events and entry events.

6.2.1 Period-cohort observational plan

Flow data are counted according to the second set of events which is specified by parallelogram QRSV in the Lexis diagram depicted in Fig. 34. Period-cohort observational model follows people belonging to a certain cohort, passing two age groups and experiencing events between two successive points in time. In other words events data are counted by age at the beginning of the time interval, or at the end of the time interval (Keilman 1985).

Demographic rates which are necessary for producing projection are calculated through of observation of events experienced by persons of certain age at the beginning of the year divided by the population exposed to the risk of experiencing such events.

Fig. 34 –Lexis diagram



Source: Keilman 1985

Number of age groups for events data is conditioned by the number of age groups in initial population plus one, i.e. events are counted for each age group in the initial population plus observation on events for those who are not alive at the beginning of the interval but will be born during observation period. The last age category is denoted as' Births' or '-5-1' age group (Van Imhoff 1999).

Preliminary preparations data for inputting into Lipro include aggregation of single year age groups flow data into 5 years age groups and 5 year observation intervals. The formulas below and notations are taken from Lipro tutorial (Van Imhoff 1999). If E(x) denotes events data of people aged x at the beginning of the 5-year interval, E(x+1) means number of events for the second year, then aggregation of events occurred during 5 year period (x..x+4) is estimated by following formula:

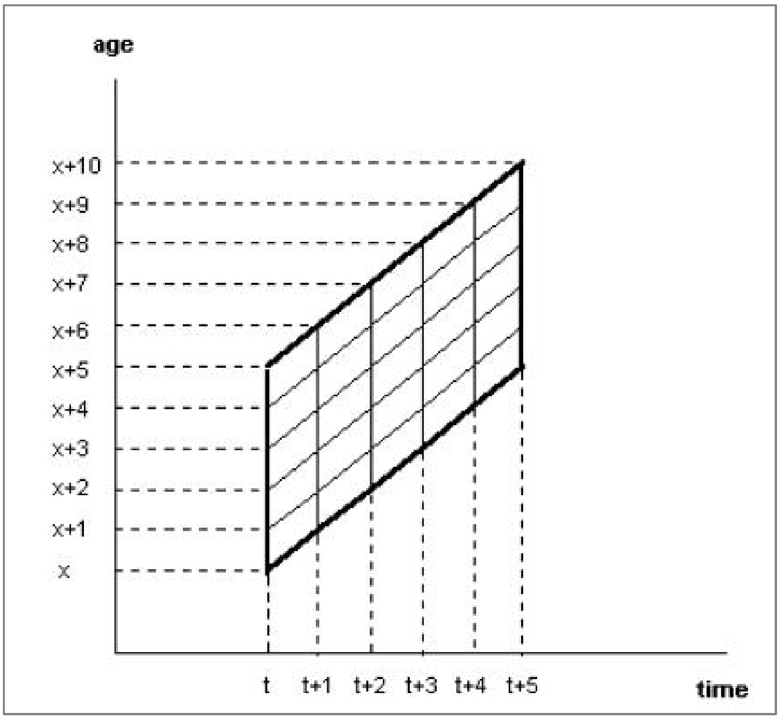
$$E(x..x+4) = 1\times E(x) + 2\times E(x+1) + 3\times E(x+2) + 4\times E(x+3) + 5\times E(x+4) + 4\times E(x+5) + 3\times E(x+6) + 2\times E(x+7) + 1\times E(x+8)$$

Events for people who not yet but will be born during 5-year interval is aggregated according to formula:

$$E(-5..-1) = 5\times E(-1) + 4\times E(0) + 3\times E(1) + 2\times E(2) + 1\times E(3)$$

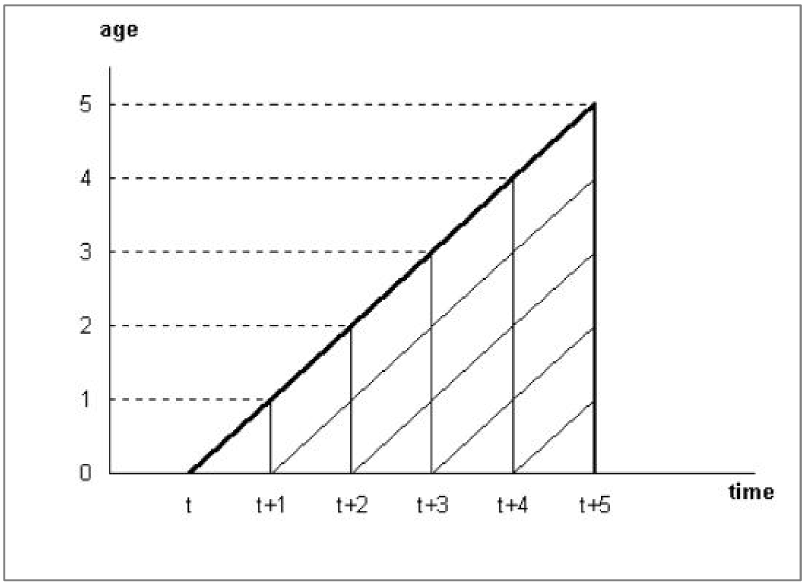
Fig. 35 and Fig. 36 graphically illustrated both rules of aggregation.

Fig. 35 - Aggregation of single-year events into 5-year interval



Source: Van Imhoff 1999

Fig. 36 - Aggregation of single-year events into 5-year interval for newborns



Source: Van Imhoff 1999

6.2.2 Consistency in multistate population projections

Consistency is a situation in which a set of model variables satisfy a certain constraint. When exogenous variables play a role in this constraint it is a matter of external consistency. Internal consistency arises when the constraint applies only to endogenous variables. Necessity of setting constraints is descended from inconsistencies within framework of aggregation and decomposition, interrelations and inadequate modelling (Keilman 1985).

In general, consistency relations are imposed on numbers of events. Adjustments to the number of events affect the number of person-years lived within each state, therefore, all rates change, even though the number of only some of the events are adjusted (Van Imhoff, E. 1992).

Consistency algorithm uses iterative approach which is capable of leaving the rates not directly related to the consistency relations intact, without losing the correspondence between rates and number of events and consists of the following steps:

1. Compute for initial exposure rates number of events;
2. Adjust number of events by the consistency relations;
3. Compute rates corresponding to adjusted number of events;
4. Replace the unrelated rates by their initial values;
5. Compute number of events for rates obtained in (4);
6. Replace those number of events which enter the consistency relations by their adjusted values as obtained in (2);
7. Repeat Steps 3-6 until convergence has been reached (Van Imhoff 1992).

Van Imhoff and Keiman (1991) provided example of imposing internal constraints using the passive and dominant relations implemented in Lipro for nuptiality model where entries into widowhood should follow the projected number of deaths. Imposing passive-dominant consistency relations means that linear combination of passive events after adjustment should be equal dominant events, when dominant events are left unchanged. These mortality-dominant widowhood consistency relations serve as a model for imposing constraints in chapter 8 on projected number of events for four macroregions.

Chapter 7

Multiregional population projections for 16 administrative divisions of Kazakhstan

Regional population projections are very important for a country with low population density, like Kazakhstan, with taking into consideration past population decline and successful efforts made by authorities to reimburse this loss of population. Twenty years ago after the breakdown of Soviet Union in 1991 Kazakhstan declared independence. Economic, social and ecological problems caused increase of mortality, especially for working ages and infant mortality, as well as fertility drop and significant outflow of emigrants, and resulted in population decrease. In 1997 the existing at that moment 19 administrative regions were transformed into 14 regions plus two cities, Astana and Almaty, old and new capitals. In the same year, Kazakhstan announced profamily policy with baby bonus system that in combination with ethnic migration policy helped to increase fertility, while economic growth affected increase of life expectancy at birth and attracted inflow of immigrants from neighboring Asian countries with higher fertility evoking subsequent changes in ethnic composition of regional population.

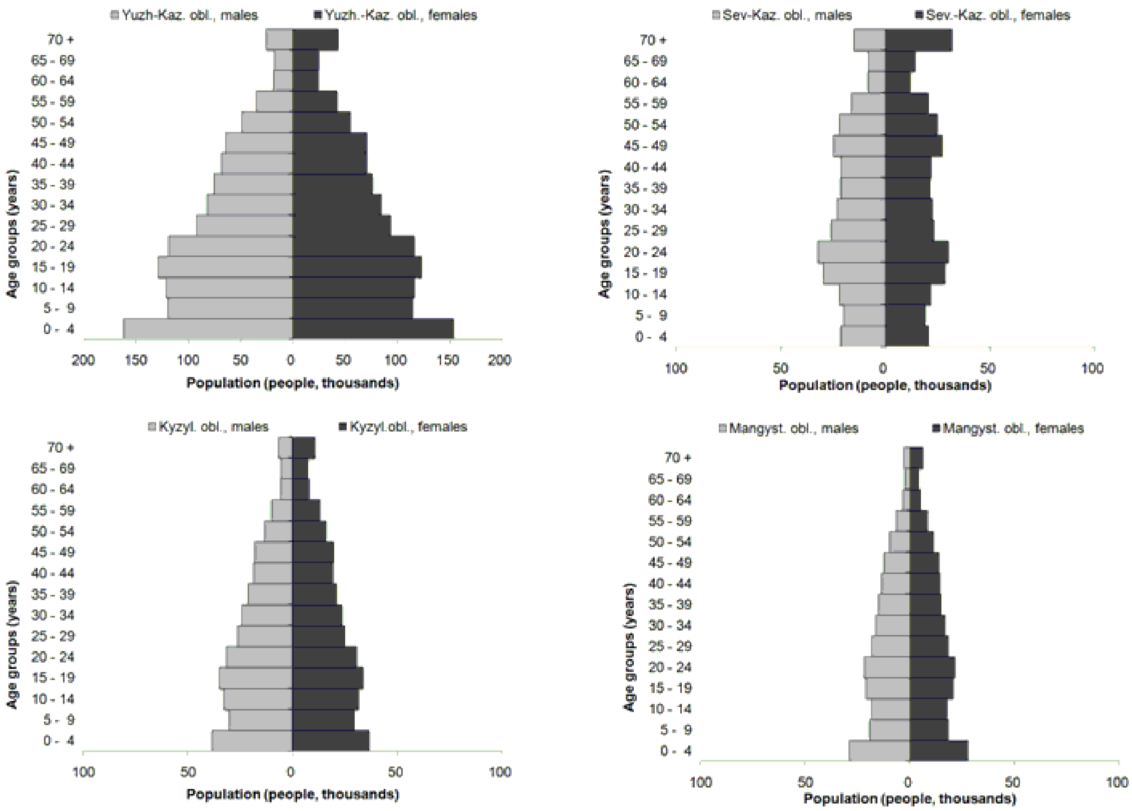
This chapter presents example of implementing regional population projections for 16 administrative divisions of Kazakhstan with projection horizon 2009-2029. Period projection consists of four projection intervals, each interval is five years long. Initial population is population of 16 administrative divisions on 1 January 2009. Interregional migration data are obtained by applying method the Three Face algorithm to disaggregate by gender and One Face One Edge algorithm to disaggregate by 19 age groups. Constant variant has constant fertility, mortality, transition rates and external migration numbers corresponding to the year 2008. In medium variant fertility and mortality rates of the divisions will be changed according to clusters to which every division belongs.

7.1 Implementation of population projections for 16 administrative divisions

Multiregional population projections of 16 administrative divisions of Kazakhstan are conducted using the software Population Development Environment Analysis (PDE) designed by the International Institute for Applied Systems Analysis (IIASA) as population projection model with interacting states. The input data for running projections consist of regional jump-off populations on 1January 2009 and age-specific fertility, mortality and transition rates, and net migration numbers corresponding to the 2008. Fig. 37 shows age pyramids of initial population of four selected provinces, they are Yuzhno-Kazakhstanskaya oblast, Severo-Kazakhstanskaya oblast, Mangystauskaya oblast, Kyzylordinskaya oblast.

The multistate approach used in the PDE software considers international migration as net migration and pays more attention to interregional movements, where out-flow interregional movements distinguishes from in-flow interregional movements with taking into consideration how regions influence each other by migration flows, and what every region gains from specific regions or how many people region loses for benefit of other regions. Interregional migration data are obtained by applying method the Three Face algorithm to disaggregate by sex and One Face One Edge algorithm to disaggregate by 19 age groups.

Fig. 37 – Age pyramids of Yuzhno-Kazakhstanskaya oblast, Severo-Kazakhstanskaya oblast, Mangystauskaya oblast, Kyzylordinskaya oblast, 01.01.2009



Source of data: The Agency on Statistics of the Republic of Kazakhstan

7.2 Constant variant and its results

The population projections for 16 administrative divisions are made using variants: constant and medium variants. Constant variant is based on mortality and fertility rates corresponding to the year 2008 and are kept constant during the whole projection period, 20 years.

Fig. 38 – Projected population by administrative divisions, 2009-2029, constant variant

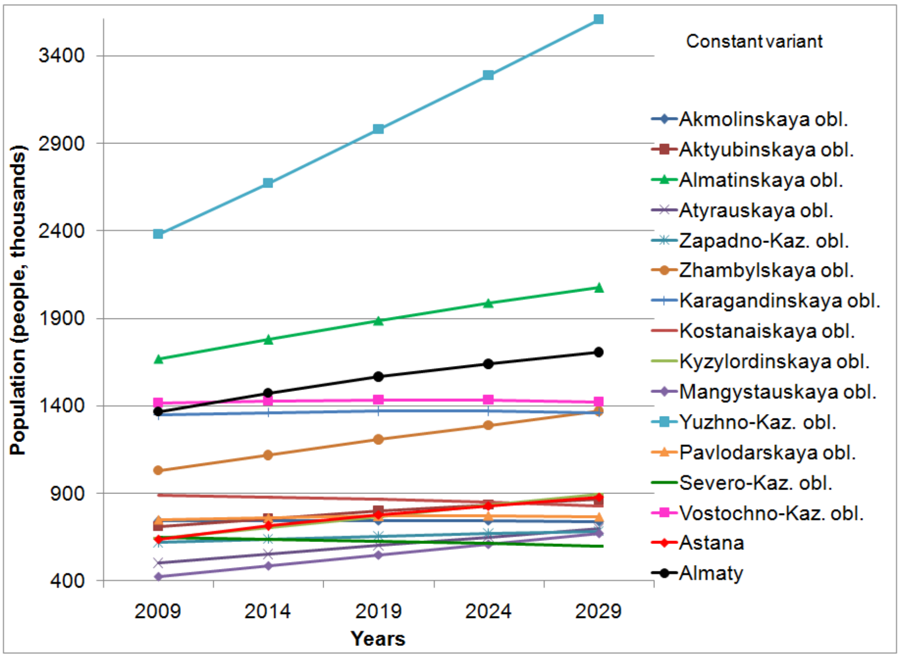


Fig. 38 displays population projection for 16 administrative for four intervals using constant variant. As it obvious from picture Yuzhno-Kazakhstanskaya oblast, outsider from the start of projection, has the biggest population and the highest growth. Almatinskaya oblast, Almaty, Zhambylskaya oblast sufficiently increase their population under constant variant assumptions.

Fig. 39 presents age pyramids of final population for four administrative divisions. Shape of three of them, Yuzhno-Kazakhstanskaya oblast, Severo-Kazakhstanskaya oblast, and Kyzylordinskaya oblast, has triangle form of Christmas tree. Pyramid's shape of Severo-Kazakhstanskaya oblast shows relatively older and small population in comparison with other regions.

Fig. 40 shows that for all regions proportion of age group 0-14 prevail over age group 65+, and wide proportion of working population is bigger than the sum of proportions 0-14 and 65+. The biggest proportions of age group 0-14 are in Yuzhno-Kazakhstanskaya oblast, Mangystauskaya oblast and Kyzylordinskaya oblast. The biggest proportions of working population are in Almaty, Karagandinskaya oblast and Kostanaiskaya oblast. The biggest proportions of elderly population are in Kostanaiskaya oblast, Severo-Kazakhstanskaya oblast and Vostochno-Kazakhstanskaya oblast.

Next picture confirms that after 20 years of population projection under constant variant assumptions youth dependency is bigger than old dependency for all regions without exception.

According to Fig. 41 the highest youth dependency and total dependency belongs to Yuzhno-Kazakhstanskaya oblast, whereas the highest old dependency belongs to Kostanaiskaya oblast.

Fig. 39 – Age pyramids of Yuzhno-Kazakhstanskaya oblast, Severo-Kazakhstanskaya oblast, Mangystauskaya oblast, Kyzylordinskaya oblast, 01.01.2029, constant variant

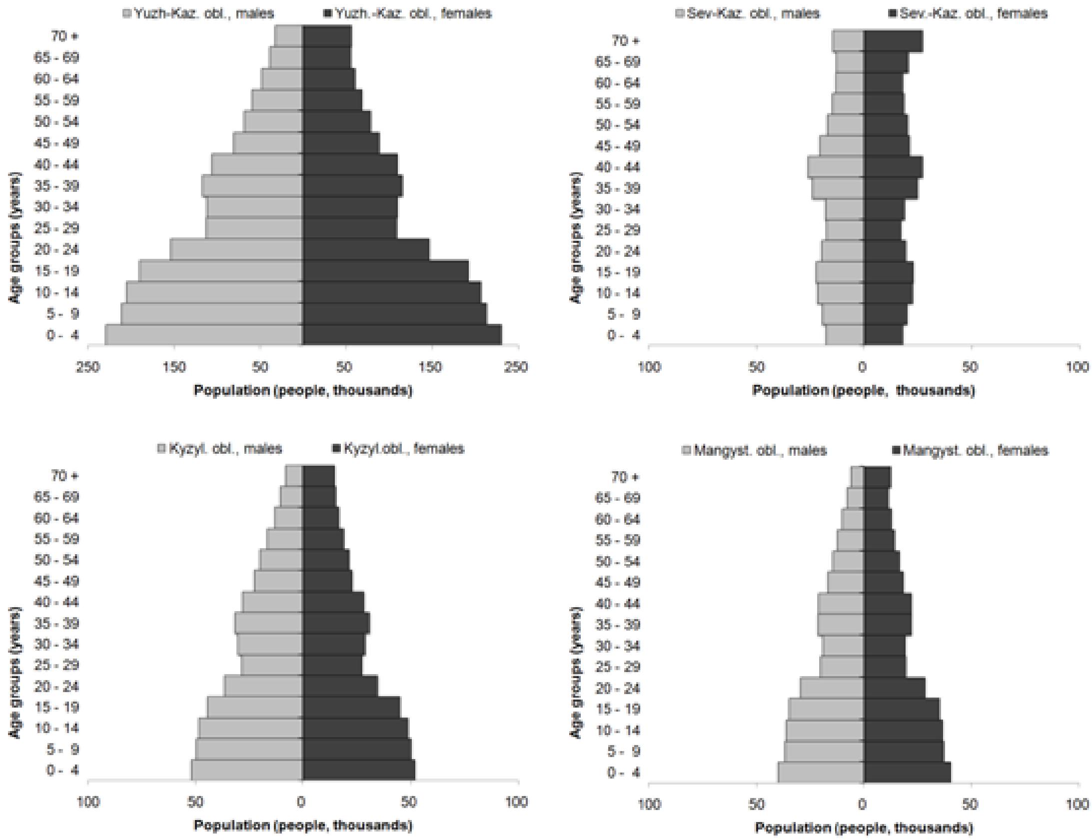


Fig. 42 shows comparison of mean age for starting and final populations. After 20 years of development with constant variant assumptions mean age is slightly increased in regions. The biggest increase of mean age belongs to Astana, Kostanaiskaya oblast and Pavlodarskaya oblast.

Next picture presents comparison of median age for starting and final populations. The regions with high fertility decrease their median age, as in case of Mangystauskaya oblast, Zhambylskaya oblast, Yuzhno-Kazakhstanskaya oblast, Atyrauskaya oblast and Kyzylordinskaya oblast. In other regions median age is increased.

In general, diversity of regional median age is bigger than difference in regional mean age. Nevertheless, the lowest mean and median age in starting and final populations belongs to Yuzhno-Kazakhstanskaya oblast. In starting population the highest mean age belongs to Kostanaiskaya oblast, the highest median age belongs to Severo-Kazakhstanskaya oblast. In final population of constant variant the highest mean and median ages belongs to Kostanaiskaya oblast.

Fig. 40 – Three age groups of final population by regions, constant variant, 01.01.2029

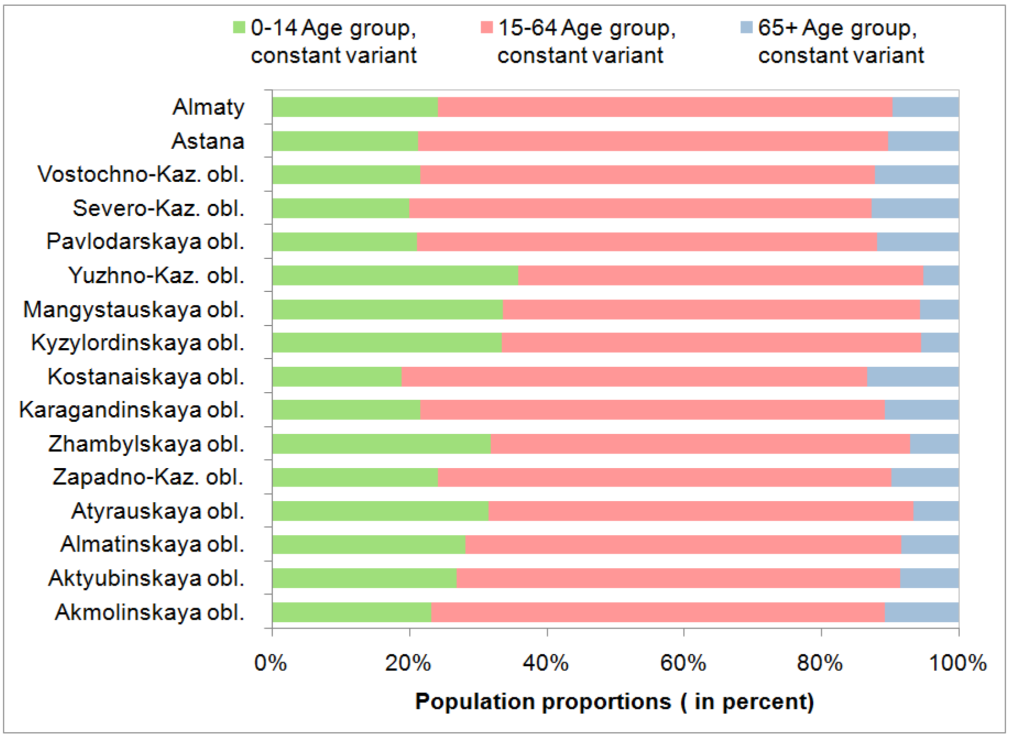


Fig. 41 –Youth dependency, old dependency, and total dependency by administrative divisions, constant variant, 01.01.2029

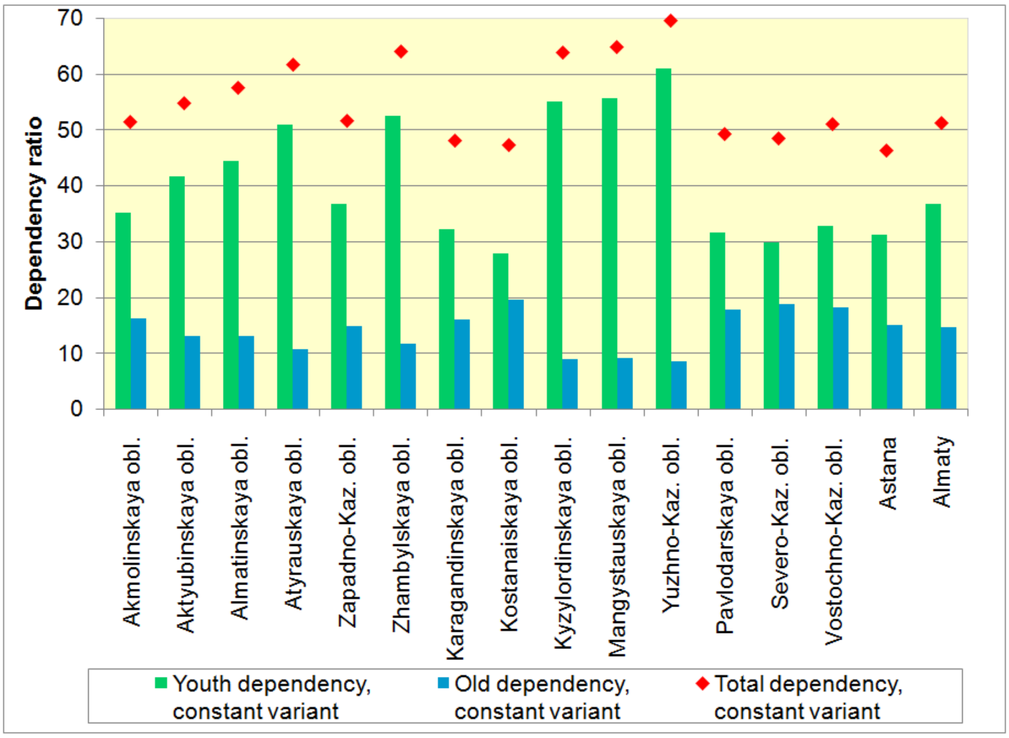


Fig. 42 – Mean age by administrative divisions, constant variant, 01.01.2029

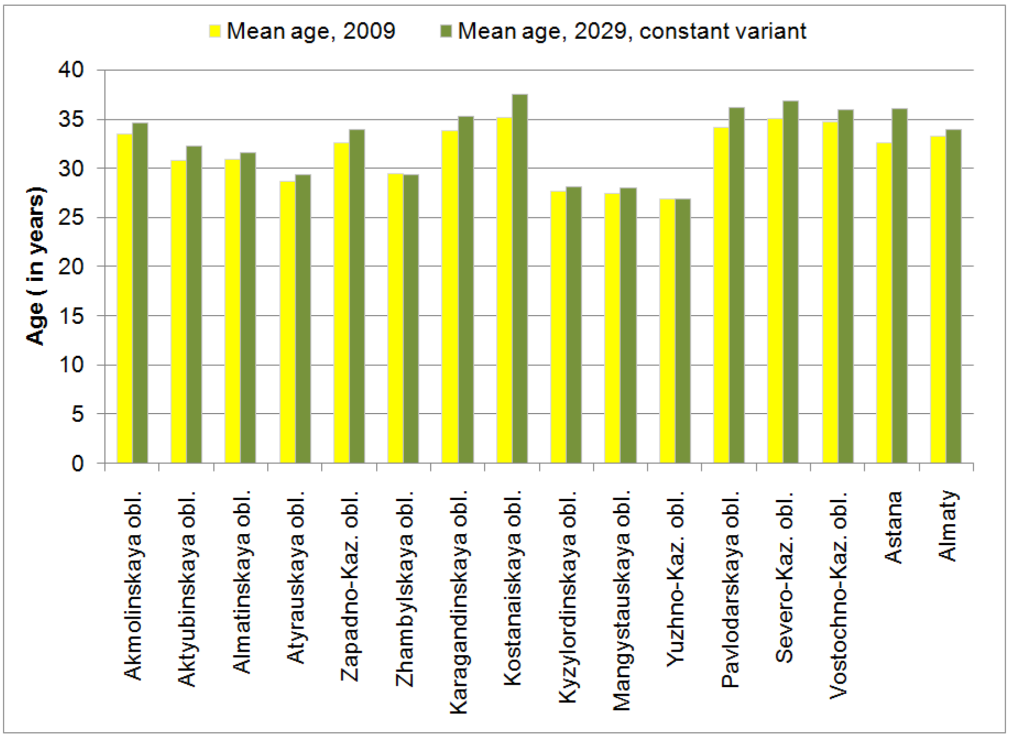
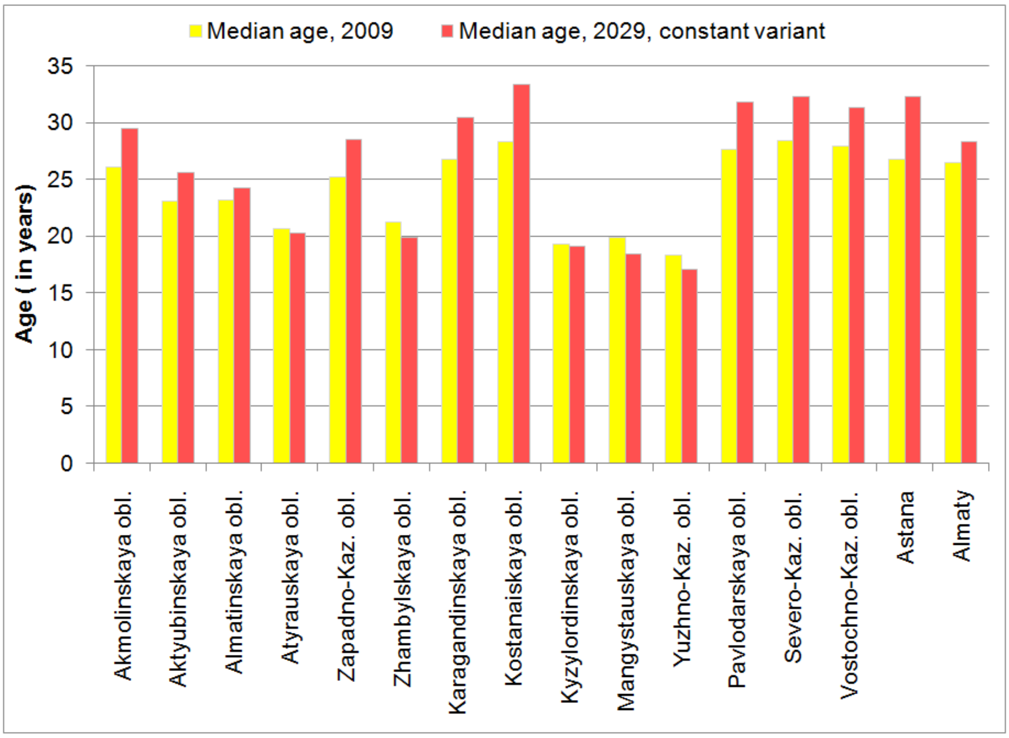


Fig. 43 – Median age by administrative divisions, constant variant, 01.01.2029



7.3 Medium variant, assumptions and results

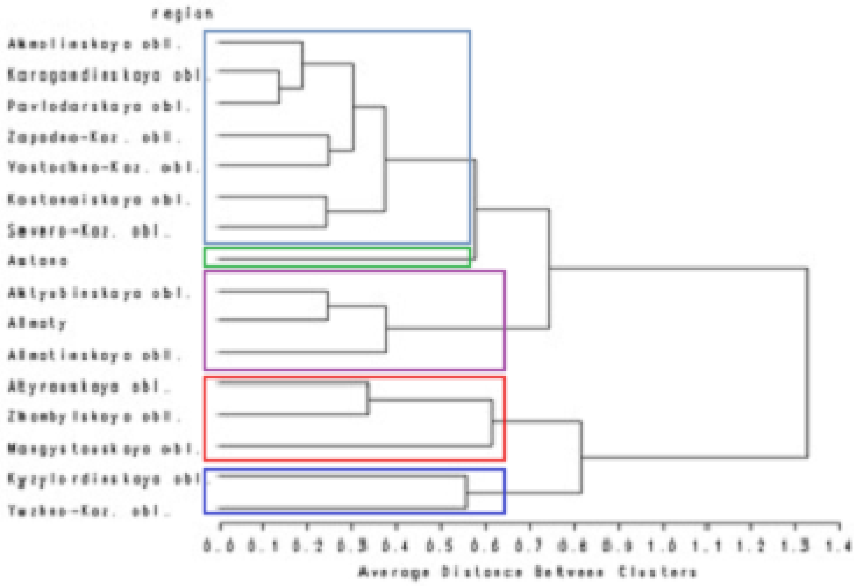
Medium variant is grounded on changing fertility and mortality, whereas migration and transitions due to their high uncertainty are assumed to be fixed corresponding to the year 2008. Regional age-specific fertility and mortality rates will be modified for both projection intervals. And age-specific fertility curve will change shape of curve with shift fertility to older ages, while in mortality assumptions shape will not be changed only mortality level will be decreased.

In order to derive region specific assumptions clustering procedures were applied for building fertility and mortality regional assumptions.

7.3.1 Fertility assumptions of medium variant

Initial 16 administrative divisions are grouped into clusters according to similarity of age-specific fertility rates in 2008 year. In other words, age-specific fertility rates of 16 administrative divisions are served as input for cluster procedure. Average link clustering produces four cluster. Additionally, Astana is defined as separate cluster for projection purposes. It is a city in the north of the country, there are some similarities are found with northern provinces during clustering, but due to status of capital and previous fast growth, it is assumed that future population development of Astana will differ from population development of northern provinces. Summarizing, 16 administrative divisions are grouped into five fertility clusters (Fig. 44).

Fig. 44 - Average link clustering of regional age-specific fertility rates, year 2008



Similarities of age-specific rates within clusters are clearly seen in next pictures. Fig. 45 displays age-specific fertility rates, the highest and high, of two neighboring regions in the south of the country, Yuzhno-Kazakhstanskaya oblast and Kyzylordinskaya oblast. In contrast the next group of regions in Fig. 46 is characterized by low age-specific fertility rates, in this cluster shape of the curve for Karagandinskaya oblast reminds curve for Astana due to vicinity of Karaganda city to Astana. Obvious likeness of age-specific rates for three regions Mangystauskaya oblast, Atyrauskaya oblast, and Zhambylskaya oblast is seen in Fig. 47. Resemblance of fertility rates for Almaty, Almatinskaya oblast and Aktuabinskaya oblast is contrasted with Astana fertility rates isolated as independent cluster in Fig.48.

Fertility rates for medium variant inside each cluster are modified according to the schema depicted in Fig. 49. Size of these modifications is specified in Tab. 8. Period projection consists of four projection intervals. Fig 49 shows that fertility rates for every projection interval are received by modifying fertility rates corresponding to every administrative divisions in preceding projection interval. In other words, clusters are created to specify expected changes in fertility pattern in relative terms.

Tab. 8 shows in the first projection interval the age-specific rates corresponding to the year 2008 are modified, and this percentage of modification are different for every cluster. In other words, assumptions about future development of fertility made separately for each cluster. The process of changing role of women in society including women education and labor force participation will result in slight fertility decline for regions with total fertility rate above replacement level, whereas regions with low fertility can try to converge their fertility to replacement level. In general, it is assumed shift fertility to older ages for the first two projection intervals, and slight fertility decline in all clusters for the last two intervals especially for clusters with the highest total fertility rates.

Fig. 45 – Age-specific fertility rates, cluster 1, year 2008

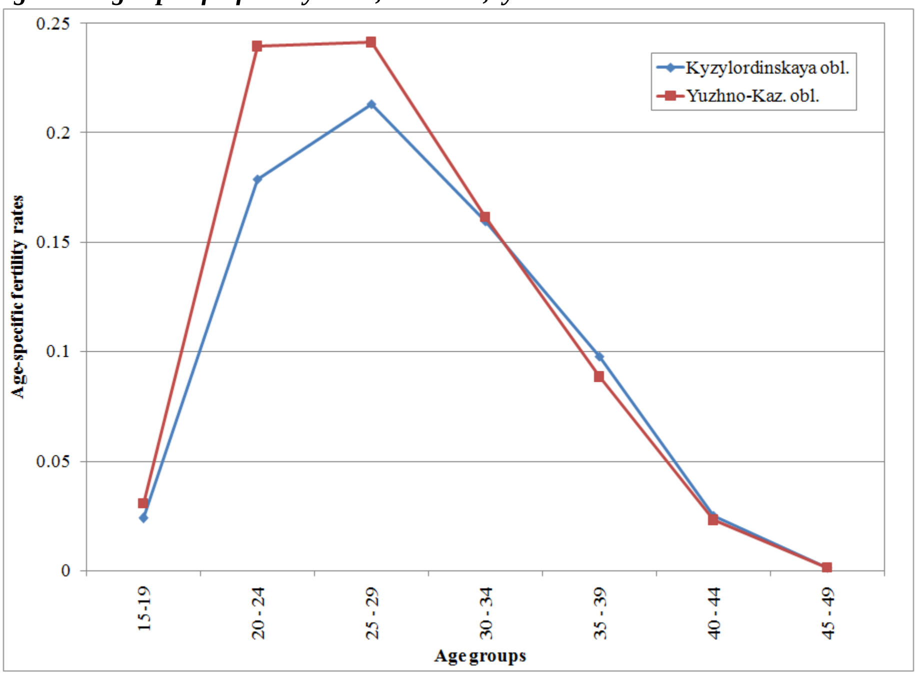


Fig. 46– Age-specific fertility rates, cluster2, year 2008

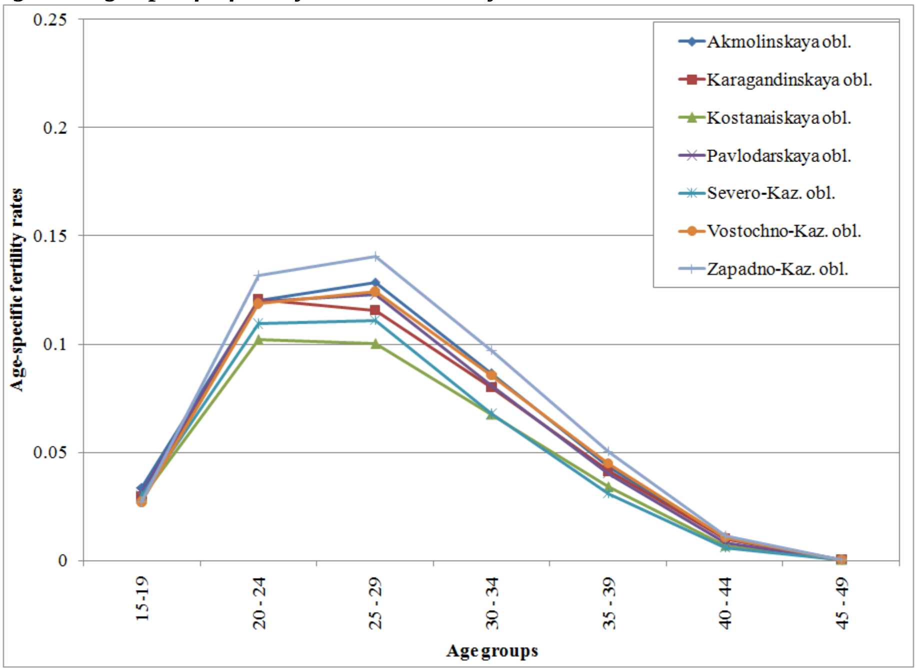


Fig. 47 – Age-specific fertility rates, cluster 3, year 2008

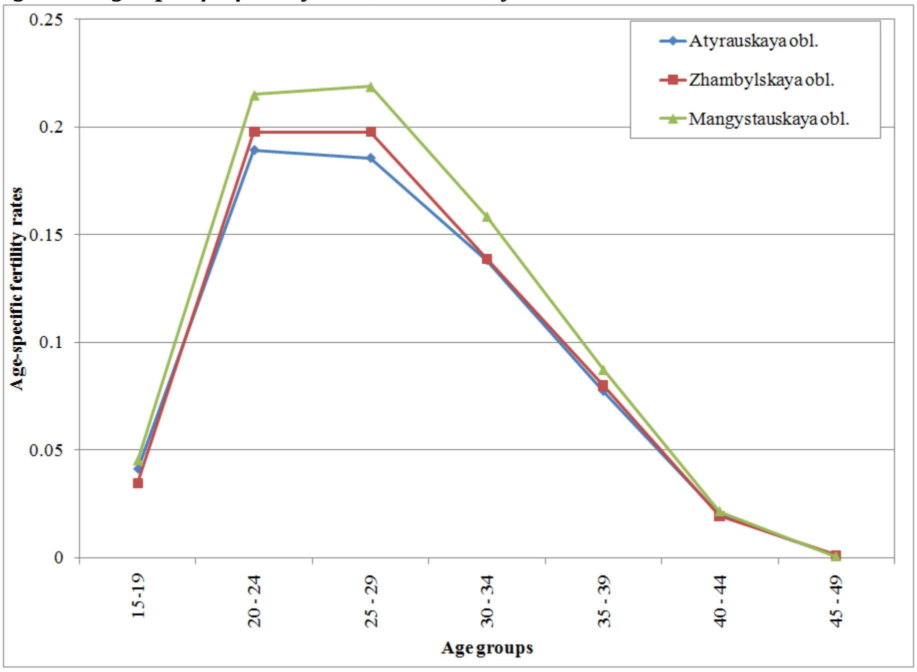


Fig. 48 – Age-specific fertility rates, cluster 4 and 5, year 2008

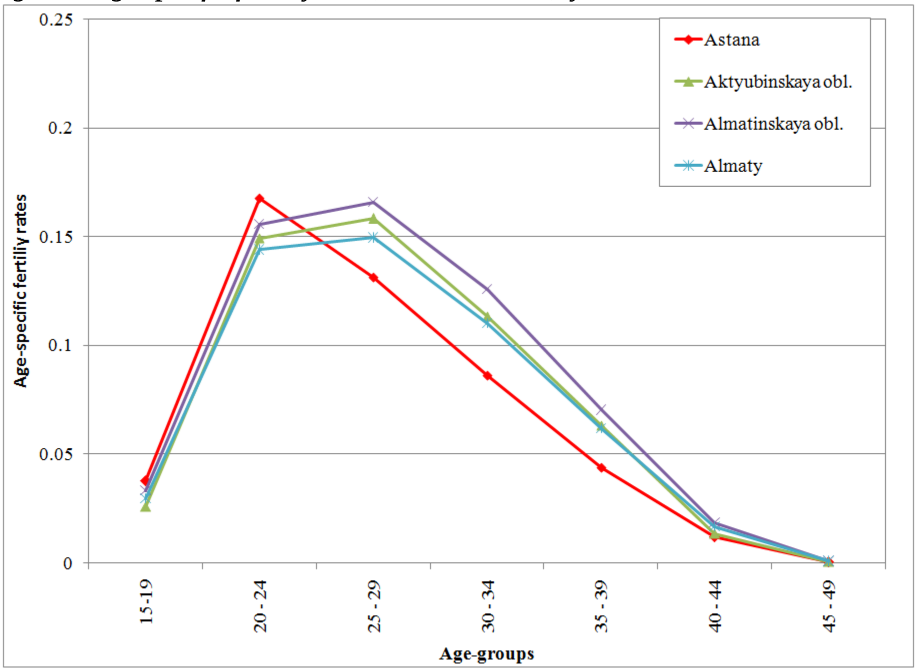
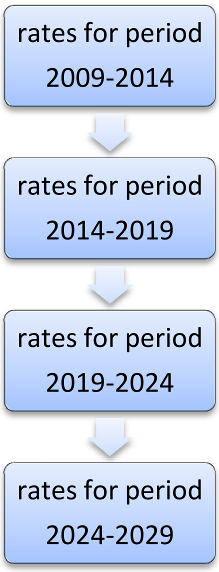


Fig. 49 – Schema of fertility and mortality rates modification, medium variant



Tab. 8 – Fertility modifications by clusters for medium variant

	2009-2014	**2014-2019**	**2019-2024**	**2024-2029**
	Modified age-specific fertility rates for the year 2008:	Age-specific fertility rates are modified in comparison with previous projection interval		
Astana	decrease for age group 20-24 for 5 %, increase for age group 25-29 for 5 %.	decrease for age group 20-24 for 5 %, increase for age group 25-29 for 5 %.	decrease for all age groups for 5 %,	without changes
Akmolinskaya, Karagandinskaya, Pavlodarskaya, Zapadno-Kazakhstanskaya, Vostochno-Kazakhstanskaya, Kostanaiskaya, Severo-Kazakhstanskaya	increase for age group 25-29 for 5 %.	increase for age group 25-29 for 5 %.	decrease for all age groups for 5 %,	without changes
Aktyabinskaya, Almaty, Almatinskaya	decrease for age group 20-24 for 5%, increase for 25-29 for 5 %.	decrease for age group 20-24 for 5%, increase for 25-29 for 5 %.	decrease for all age groups for 5 %,	without changes
Atyrauskaya, Zhambylskaya, Mangystauskaya	decrease for age group 20-24 for 5%.	decrease for age group 20-24 for 5%.	decrease for all age groups for 10 %,	without changes
Kyzylordinskaya, Yuzhno-Kazakhstanskaya	decrease for age groups 20-24, 25-29 for 5%.	decrease for age groups 20-24, 25-29 for 5%.	decrease for all age groups for 10 %,	without changes

7.3.2 Mortality assumptions of medium variant

Distinctive characteristic of mortality in Kazakhstan is that male mortality much higher than female mortality. Although spatial pattern of female mortality repeats in general spatial pattern of male mortality, nevertheless regional disparities in mortality levels by regions are bigger for males than for females. Regional age-specific mortality rates for males corresponding to the year 2008 were chosen as input for clustering initial 16 administrative regions of Kazakhstan.

The principal component analysis is applied to regional age-specific mortality rates corresponding to the year 2008 with purpose to reduce dimensionality. Fig. 50 reveals three large eigenvalues, 10.905, 3.305 and 1.342. These first three principal components are retained on the basis of the eigenvalues greater-than-one rule. The first component has large positive loadings for all five variables. The correlation with age-specific rates for age groups 55-60, 45-50, 60-65, 40-45, 30-35, 65-70 are especially high. The second component has highest loading for age groups 75-80, 0, 80-85, 85+. The third component has the highest loading for the age groups 5-10, 10-15, 15-20, and negative for age group 1-4. In order to enhance factor interpretability an orthogonal Varimax rotation is applied. Fig. displays the results of rotated factor pattern. After computing scores of three principal components, procedure Cluster produces five clusters based on output from Distance procedure. Fig. illustrates tree diagram obtained by using average linkage clustering.

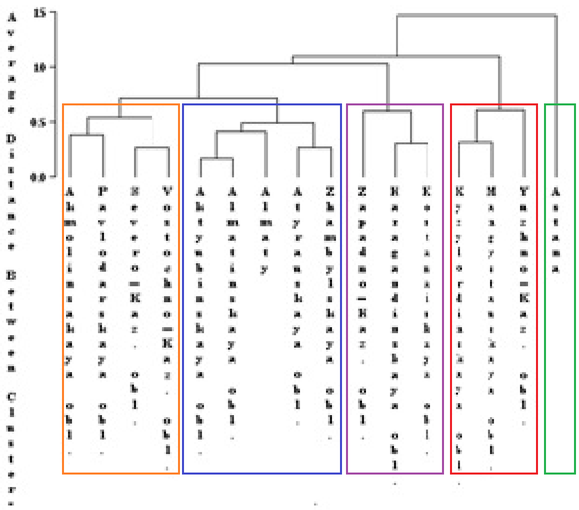
Fig. 50 –Output of principal component analysis

Eigenvalues of the Correlation Matrix: Total = 19 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	10.9049480	7.6001285	0.5739	0.5739
2	3.3048195	1.9627678	0.1739	0.7479
3	1.3420517	0.3469885	0.0706	0.8185
4	0.9950632	0.2337547	0.0524	0.8709
5	0.7613085	0.2640701	0.0401	0.9110

Factor Pattern				
		Factor1	Factor2	Factor3
M55	M55	0.94791	-0.17217	-0.08090
M45	M45	0.94166	-0.25335	-0.03646
M60	M60	0.91998	0.10743	-0.14669
M40	M40	0.90954	-0.25524	-0.12358
M30	M30	0.90476	-0.35363	-0.07371
M65	M65	0.90278	0.36361	0.05004
M35	M35	0.89768	-0.21017	-0.16136
M70	M70	0.88918	0.36174	-0.05916
M20	M20	0.87499	-0.26202	-0.01135
M25	M25	0.85542	-0.33080	-0.17732
M50	M50	0.85482	-0.35464	-0.13054
M15	M15	0.67537	-0.45085	0.33980
M75	M75	0.61858	0.74322	-0.00165
M0	M0	0.26573	0.68374	-0.00065
M80	M80	0.65840	0.66717	0.02564
M85	M85	0.51664	0.63997	0.10737
M5	M5	0.41658	0.06417	0.72409
M10	M10	0.44512	0.25982	0.44990
M1	M1	-0.08031	0.49872	-0.59744

Rotated Factor Pattern				
		Factor1	Factor2	Factor3
M30	M30	0.95368	0.10142	0.17112
M45	M45	0.93241	0.20938	0.19751
M40	M40	0.92786	0.18775	0.10732
M25	M25	0.92776	0.09272	0.05735
M50	M50	0.92559	0.07408	0.10640
M55	M55	0.91394	0.28148	0.14216
M35	M35	0.90773	0.21991	0.06084
M20	M20	0.87218	0.17234	0.20977
M60	M60	0.78522	0.51204	0.02572
M15	M15	0.69177	-0.06648	0.54027
M75	M75	0.21154	0.94354	-0.00274
M80	M80	0.27196	0.89628	0.04449
M85	M85	0.14042	0.81143	0.09937
M65	M65	0.60858	0.74223	0.16868
M70	M70	0.62567	0.72789	0.06091
M0	M0	-0.06750	0.72775	-0.06282
M5	M5	0.14595	0.29190	0.77166
M10	M10	0.15595	0.46212	0.47976
M1	M1	-0.13272	0.36902	-0.67699

Fig. 51 – Average link clustering of regional mortality, 2008 year



Assumptions were derived for every cluster using past 10 years development of mortality in regions. In contrast to fertility mortality assumptions did not changed the shape of curves of age-specific mortality rates and resulted in lowering of level mortality according to common worldwide trend of declining mortality.

Tab. 9 presents mortality assumptions separately for every cluster for every projection interval. The first projection interval adopts regional age-specific mortality rates corresponding to every administrative divisions in year 2008 without changes. The following projection intervals modify regional mortality rates using regional rates of preceding interval. Every cluster changes mortality rates according to its own pace, percentages of these changes in comparison with previous interval are shown in Tab. 9.

Tab. 9 – Mortality modifications by clusters for medium variant

	2009-2014	**2014-2019**	**2019-2024**	**2024-2029**
		Age-specific mortality rates decline in comparison with previous projection interval		
Astana	Age-specific rates of the year 2008	Males: 5% Females: 3%	Males: 5% Females: 3%	Males: 5% Females: 3%
Yuzhno-Kazakhstanskaya, Mangystauskaya, Kyzylordinskaya	Age-specific rates of the year 2008	Males: 5% Females: 5%	Males: 5% Females: 5%	Males: 5% Females: 5%
Kostanaiskaya, Karagandinskaya, Zapadno-Kazakhstanskaya	Age-specific rates of the year 2008	Males:3% Females:5%	Males:3% Females:5%	Males:3% Females:5%
Zhambylskaya, Atyrauskaya, Almaty, Almatinskaya, Aktyabinskaya	Age-specific rates of the year 2008	Males:5% Females:5%	Males:5% Females:5%	Males:5% Females:5%
Vostochno-Kazakhstanskaya, Severo-Kazakhstanskaya, Pavlodarskaya, Akmolinskaya,	Age-specific rates of the year 2008	Males:3% Females:3%	Males:3% Females:3%	Males:3% Females:3%

7.3.3 Results

Estimation of future population development depends on the initial age structure of regions: how old or how young generations prevail in regions, besides in assumptions of medium variant it is suggested fertility shift to older ages, while it is clear from initial population structure that these generations are less numerous.

Future population development in regions of Kazakhstan with medium variant assumptions is depicted in Fig. 52. Regions in the south of Kazakhstan with relatively big populations at the beginning of projection and high past fertility shows steady population increase even under medium variant assumptions, with anticipation of slight fertility and mortality changes.

Fig. 53 presents age pyramids of final population for four administrative divisions. The bottom bands related to new-born generations are not so wide like in constant variant.

Subsequently, Fig. 54 shows that population divided into three big age groups displays almost the same proportions of youth, working and elderly population like in constant variant.

Next picture also confirms that in final population of medium variant old dependency is not bigger than youth dependence. Nevertheless, it is striking how regions differ in the youth dependence, namely contrasting combination of high youth dependence with low old dependence opposite to low youth dependence with relatively high old dependence.

In the last two pictures of this chapter regional mean and median ages of final population are compared with starting values at the beginning of projection. Twenty years of medium variant increase the mean and median ages in regions of Kazakhstan and population become older. The biggest increase is related to Astana, already possessor of the highest life expectancy, whereas Yuzhno-Kazakhstanskaya oblast has the lowest mean and median ages, Kostanaiskaya oblast obtains the highest mean and median ages at the end of projection.

Summarizing, results of constant and medium variant with 20 years population horizon show that efforts spent to recovery fertility result in population bonus, such as increased proportion of population of working ages.

Fig. 52 – Projected population by 16 administrative divisions, 2009-2029, medium variant

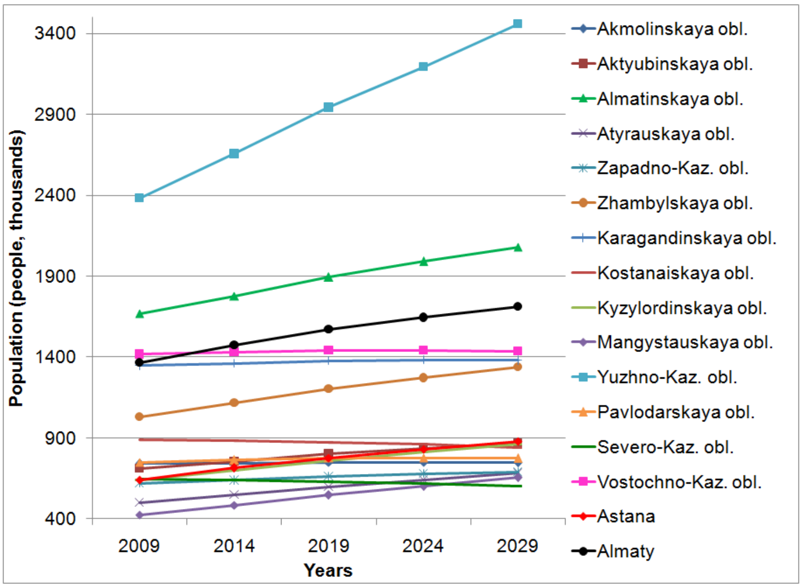


Fig. 53 – Age pyramids of Yuzhno-Kazakhstanskaya oblast, Severo-Kazakhstanskaya oblast, Mangystauskaya oblast, Kyzylordinskaya oblast, 01.01.2029, medium variant

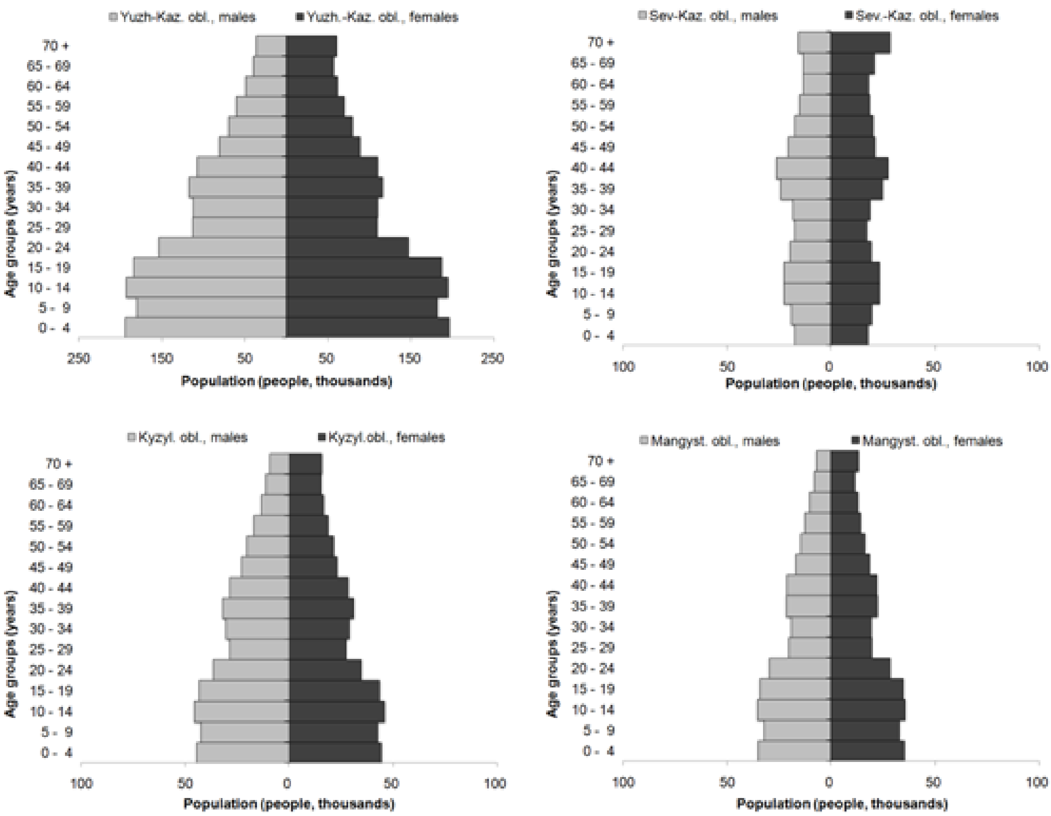


Fig. 54 – Three age groups of final population by regions, medium variant, 01.01.2029

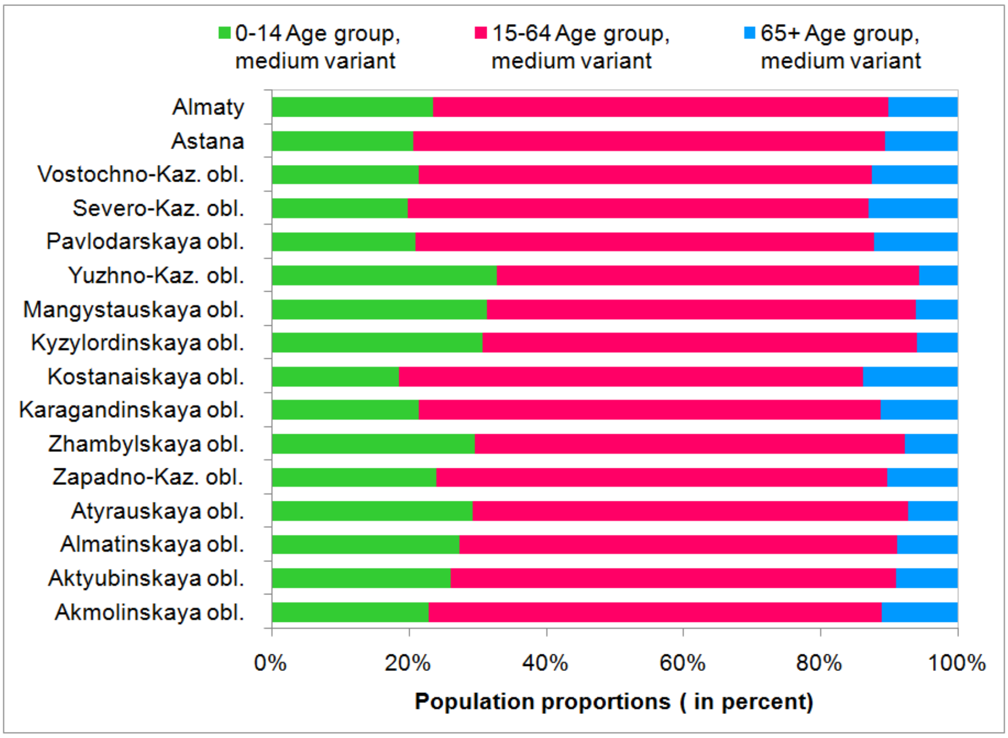


Fig. 55 –Youth dependency, old dependency, and total dependency by regions, medium variant, 01.01.2029

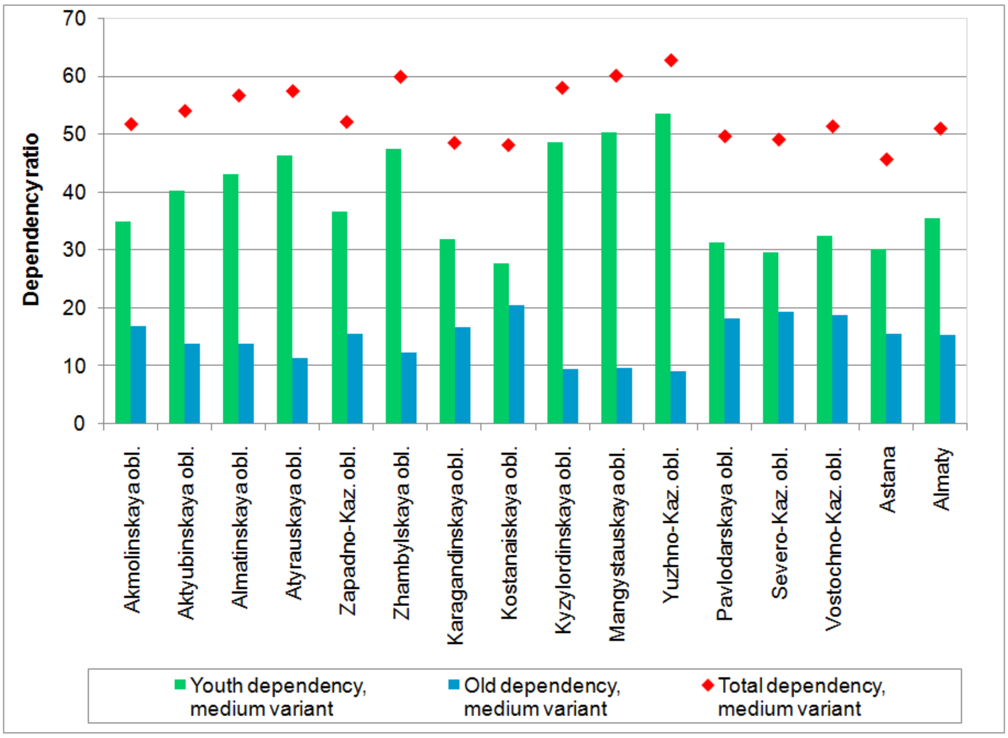


Fig. 56 – Mean age by administrative divisions, medium variant, 01.01.2029

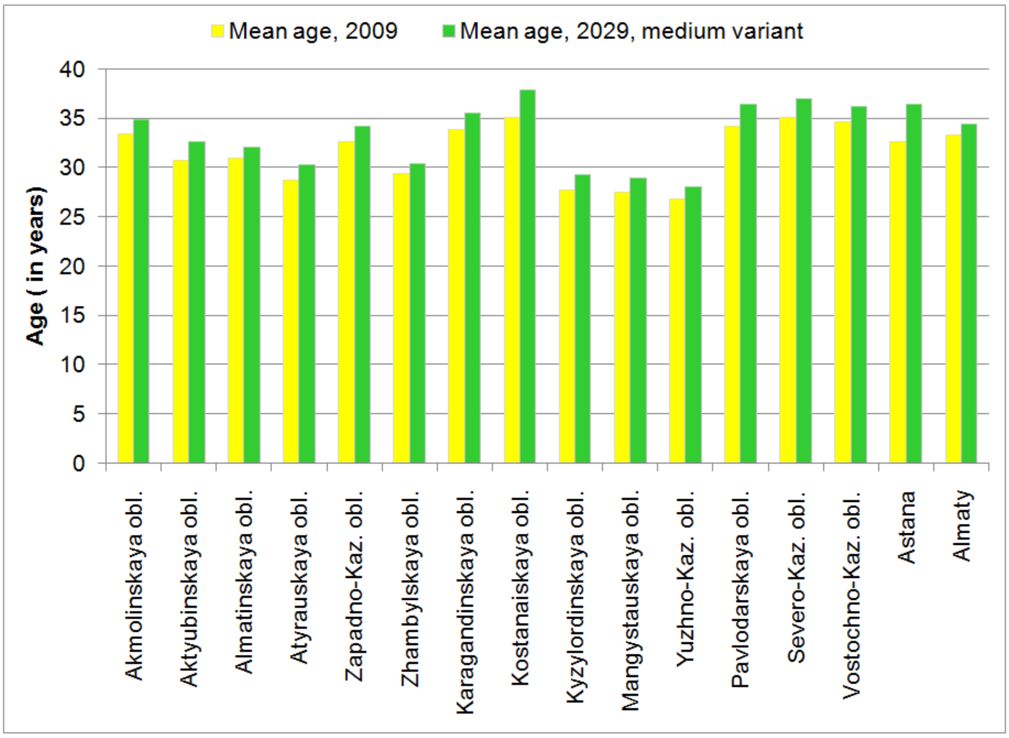
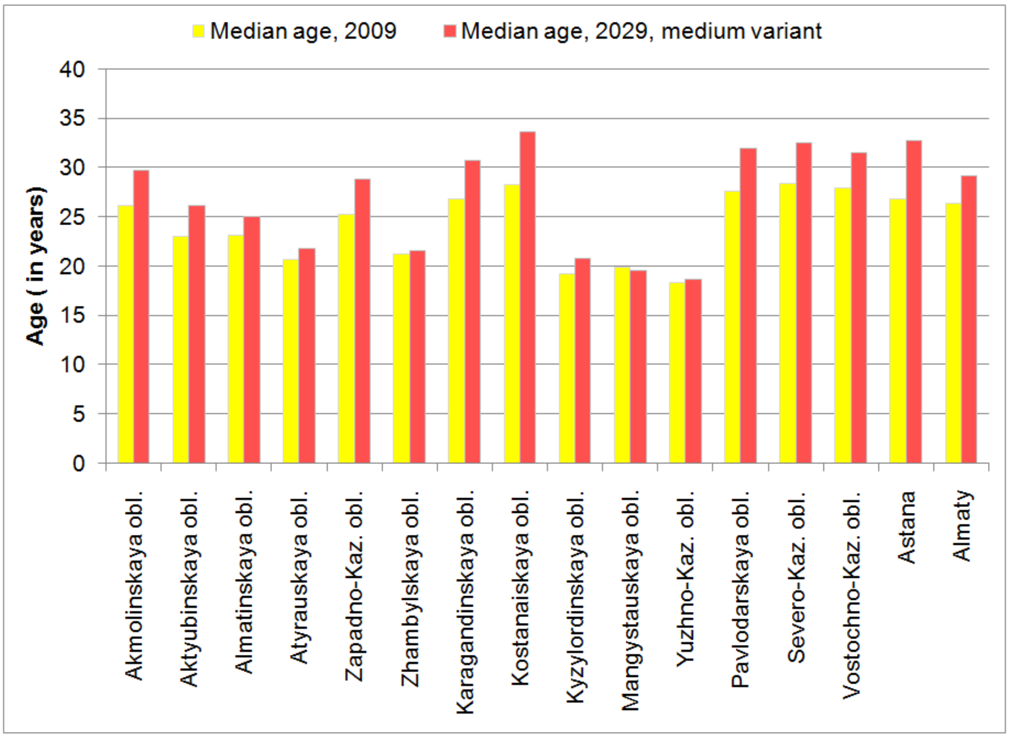


Fig. 57 – Median age by administrative divisions, medium variant, 01.01.2029



Chapter 8

Multiregional population projections of four macroregions of Kazakhstan

This chapter presents a population projection for four macroregions using period-observational approach and imposed consistency relations. Initial population is population in for macroregions on 01.01.2004. Projection period is 2004-2059. Period-cohort observational plan covers period 2004-2009. Interregional migration data for the year 2008 are obtained by applying method the Three Face algorithm to disaggregate by gender and once again Three Face algorithm to disaggregate by 101 age groups. Interregional migration data for the 2008 after applying corrective factors produce interregional migration data for the years 2004, 2005, 2006, 2007.

Single year events data are aggregated into 5 years age groups and 5 year observation intervals according period-cohort observational approach, and then 16 administrative regions will be transformed into four macroregions according to interregional migration flows. Constant variant is based on data collected during period-observational period 2004-2008 and kept them constant during 11 projection intervals. Principal variant assumes fertility decrease and life expectancy at birth increase, and as a result gives lower estimation of future population in four macroregions of Kazakhstan. This method allows evaluating how population size of macroregions is influenced by interregional migration impact or by fertility decrease. Since age profiles of interregional migration data are obtained through of disaggregation by sex and by age groups, internal consistency constraints are applied to put restrictions on births and deaths events in four macroregions of Kazakhstan. These bounds are implemented by using passive-dominant relations where dominant level is represented by country level of births and death events.

8.1 Implementation of population projections of four macroregions

These multiregional projections are conducted using LIPRO application. LIPRO (Lifestyle PROjections) is software developed by Netherlands Interdisciplinary Demographic Institute (NIDI) for multistate demographic analysis and projection. LIPRO 2.0 appeared in 1991 as an application for household projections. Lipro 3 was created in 1994. The latest version is LIPRO 4.0 was developed in 1998/1999.

The software describes development of population over time when population is broken-down by certain demographic characteristics, such as region, age, sex, marital or household status etc.

Events data for projection purposes are divided into 20 age groups (Births, 0-4, 5-9, .. , 90+). Observation period covers 2004-2008 years, projection period consists of 11 projections intervals. Every projection interval is 5 years long. Initial population is population on 01.01.2004, final population is population on 01.01.2059.

Preliminary preparations data for inputting into Lipro include aggregation of single year age groups flow data into 5 years age groups and 5 year observation intervals.

8.2 Aggregation by age

Single year initial population per 1 January 2004 is aggregated into 5 year age groups using following rules:

P_{Births} (2004) = 0

P_{0-4} (2004) = P_0 (2004) + P_1 (2004) + P_2 (2004) + P_3 (2004) + P_4 (2004) + P_5 (2004)

P_{90} (2004) = P_{91} (2004) + P_{92} (2004) + P_{93} (2004) + P_{94} (2004) + P_{95} (2004) + P_{96} (2004) + P_{97} (2004) + P_{98} (2004) + P_{99} (2004)

Aggregated births, deaths and net migration counts are calculated according to general above given formulas. As an example the following rows describes aggregating of deaths for the period from 1 January 2004 to 31 December 2008:

D_{Births} = D_0(2004) + D_{0-1}(2005) + D_{0-2} (2006) + D_{0-3} (2007) + D_{0-4} (2008)

D_{0-4} = D_{1-5}(2004) + D_{2-6}(2005) + D_{3-7}(2006) + D_{4-8}(2007) + D_{5-9}(2008)

D_{90+} = D_{91-99}(2004) + D_{92-99}(2005) + D_{93-99}(2006) + D_{94-99}(2007) + D_{95-99}(2008)

8.3 Interregional migration data for period-cohort observational plan 2004-2008

In the chapter 6 interregional age-sex specific flow matrices are derived for the year 2008 after applying the method Three Face algorithm. These matrices have data for 256 origin-destination directions, disaggregated by two sexes and 101 age groups.

In the first step, abovementioned interregional age-sex specific profiles for the year 2008 are extended for the years 2004, 2005, 2006, 2007 based on the knowledge of totals. It is assumed that interregional migration profiles for the year 2008 could be valid for the years 2004, 2005, 2006, 2007 after applying adjustment. Regional totals of arrivals and departures for these years are divided by regional arrivals and departures for the year 2008, these obtained ratios (Tab. 11 and Tab. 12) are used as corrective factors. In other words, the age-sex-direction specific profiles for the year 2008 are multiplied by corrective factors, as the result interregional flows for years 2004, 2005, 2006, 2007 are received.

The second step is devoted to aggregation single-year data into 5-year data. Obtained age-sex specific profiles for 256 directions for the years 2004, 2005, 2006, 2007, 2008 are aggregated into larger age groups using the same formulas adopted for aggregation births, deaths, and net migration events during period 2004-2008.

M_{Births} = M_0(2004) + M_{0-1}(2005) + M_{0-2} (2006) + M_{0-3} (2007) + M_{0-4} (2008)

$M_{0\text{-}4} = M_{1\text{-}5}(2004) + M_{2\text{-}6}(2005) + M_{3\text{-}7}(2006) + M_{4\text{-}8}(2007) + M_{5\text{-}9}(2008)$

$M_{90+} = M_{91\text{-}99}(2004) + M_{92\text{-}99}(2005) + M_{93\text{-}99}(2006) + M_{94\text{-}99}(2007) + M_{95\text{-}99}(2008)$

Tab. 11 - Multiplicative factors for In-Interregional movements, 2004, 2005, 2006, 2007

	2004	2005	2006	2007	2008
Akmolinskaya obl.	1.056	1.178	1.174	1.073	1
Aktyubinskaya obl.	1.012	1.074	1.065	1.080	1
Almatinskaya obl.	1.036	0.859	0.810	0.898	1
Atyrauskaya obl.	0.493	0.788	0.786	0.867	1
Zapadno-Kaz. obl.	0.898	0.965	0.936	0.918	1
Zhambylskaya obl.	0.679	0.678	0.671	0.827	1
Karagandinskaya obl.	0.795	0.763	0.791	0.869	1
Kostanaiskaya obl.	0.986	1.004	0.961	0.910	1
Kyzylordinskaya obl.	0.356	0.477	0.555	0.726	1
Mangystauskaya obl.	0.344	0.576	0.548	0.686	1
Yuzhno-Kaz. obl.	0.631	0.687	1.048	1.062	1
Pavlodarskaya obl.	1.099	0.981	0.894	0.982	1
Severo-Kaz. obl.	0.886	1.079	0.967	0.937	1
Vostochno-Kaz. obl.	1.193	1.138	1.023	1.002	1
Astana	0.587	0.637	0.702	0.775	1
Almaty	0.949	0.980	0.918	0.951	1

Tab. 12 - Multiplicative factors for Out-Interregional movements, 2004, 2005, 2006, 2007

	2004	2005	2006	2007	2008
Akmolinskaya obl.	0.764	0.781	0.798	0.819	1
Aktyubinskaya obl.	0.715	0.699	0.784	0.892	1
Almatinskaya obl.	0.923	0.947	0.868	0.828	1
Atyrauskaya obl.	0.670	0.844	0.926	0.914	1
Zapadno-Kaz. obl.	0.795	0.935	0.924	0.970	1
Zhambylskaya obl.	0.909	0.898	0.876	0.931	1
Karagandinskaya obl.	0.859	0.848	0.860	0.896	1
Kostanaiskaya obl.	0.831	0.853	0.918	0.952	1
Kyzylordinskaya obl.	0.895	0.887	0.845	0.971	1
Mangystauskaya obl.	0.895	0.951	1.070	1.056	1
Yuzhno-Kaz. obl.	0.663	0.759	0.761	0.885	1
Pavlodarskaya obl.	0.851	0.922	0.908	0.948	1
Severo-Kaz. obl.	0.900	1.025	1.146	1.082	1
Vostochno-Kaz. obl.	1.034	0.917	0.884	0.871	1
Astana	0.619	0.663	0.711	0.749	1
Almaty	0.662	0.657	0.667	0.838	1

8.4 Aggregation by regions

Interregional flows are influenced and supported by the following factors:

•Economic development of regions,

•Proportion of leading industries in regions; planned industrial development and infrastructure;

•Job market, demand and competition, expected salaries;

•Financial and educational institutions;

•Urban planning strategies, vicinity of other strong attractors;

•Saturation, demand and competition in housing market ; housing projects;

•Ethnic composition of regions, neighboring countries, social and cultural ties.

Based on the data of interregional flows collected during the observation period 2004-2008 and with taken into consideration above mentioned factors 16 administrative regions are grouped into four macroregions according to previous and expected similarity in migration flows.

The first macroregion is characterized by strong interregional migration in-flows and contains two administrative divisions: Astana, Almaty;

The second macroregion is specified by prevailing interregional migration in-flows, and four administrative divisions belong to this macroregion: Aktyubinskaya oblast, Almatinskaya oblast, Atyrauskaya oblast, Mangystauskaya oblast;

The third macroregion is influenced by steady interregional migration out-flows and includes five divisions: East-Kazakhstan oblast, Zhambulskaya oblast, North-Kazakhstan oblast, Kostanayskaya oblast, Kyzylordinskaya oblast;

In the forth macroregion interregional out-flows are prevailing over in-flows, and five divisions comprise this group: Akmolinskaya oblast, West-Kazakhstan oblast, Karagandinskaya oblast, Pavlodarskaya oblast, South-Kazakhstan oblast.

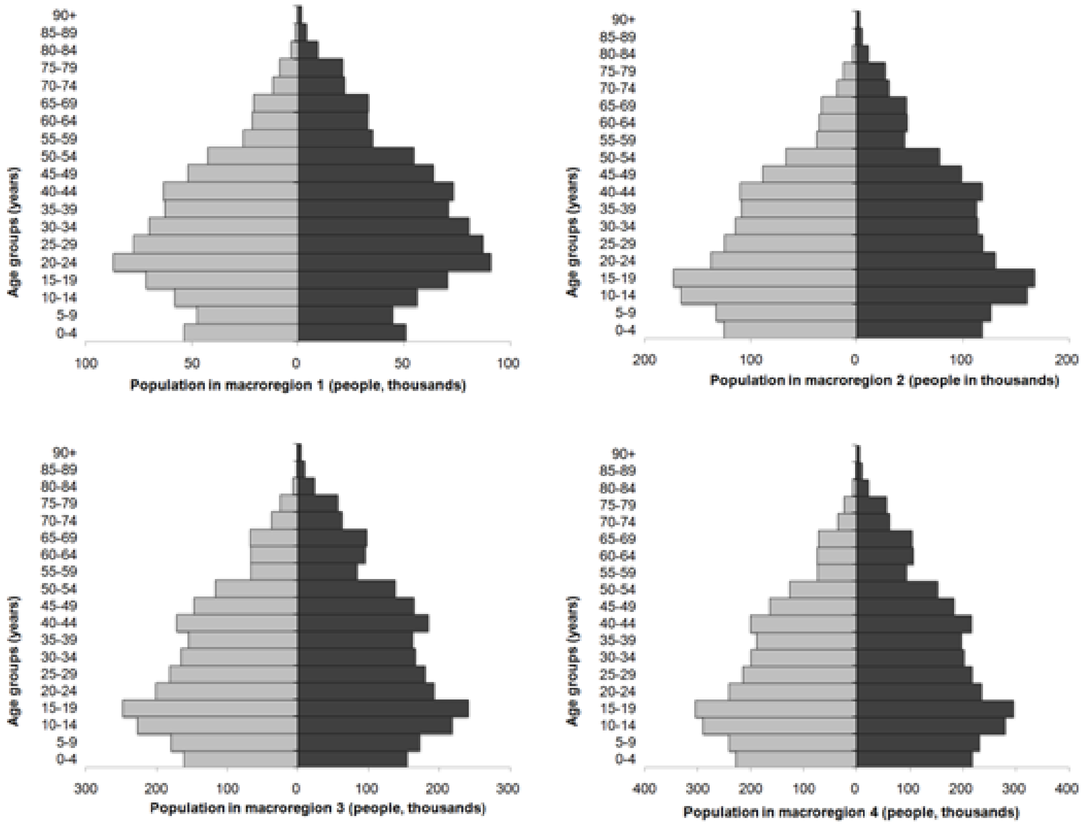
The number of regions after aggregation reduced from 16 to 4, and flow data for macroregions are computed by summing events of corresponding administrative regions belonging to the same macroregion, while the number of interregional migration directions is reduced from 16*16 to 4*4.

All events after aggregation are divided by the appropriate middle populations times five in order to receive rates.

Initial populations for four macroregions along with obtained mortality, fertility, interregional migration rates are implemented into Lipro as input for projection purposes using Import property from Excel file supported by the program. At the same time net migration for the macroregions are imported as numbers.

Age pyramids of initial population of four macroregions are depicted in Fig. 58.

Fig. 58 - Age pyramids of initial population by macroregions, 01.01.2004



Source of data: The Agency on Statistics of the Republic of Kazakhstan

8.5 Constant variant

Rates of fertility, mortality and interregional migration and net numbers of external migration are kept constant on the level corresponding to the observation period 2004-2009.

Population projections for constant and principal variants are produced with and without imposing consistency relations with recalculation of adjusted demographic rates to meet restrictions on births and deaths counts.

8.6 Principal variant

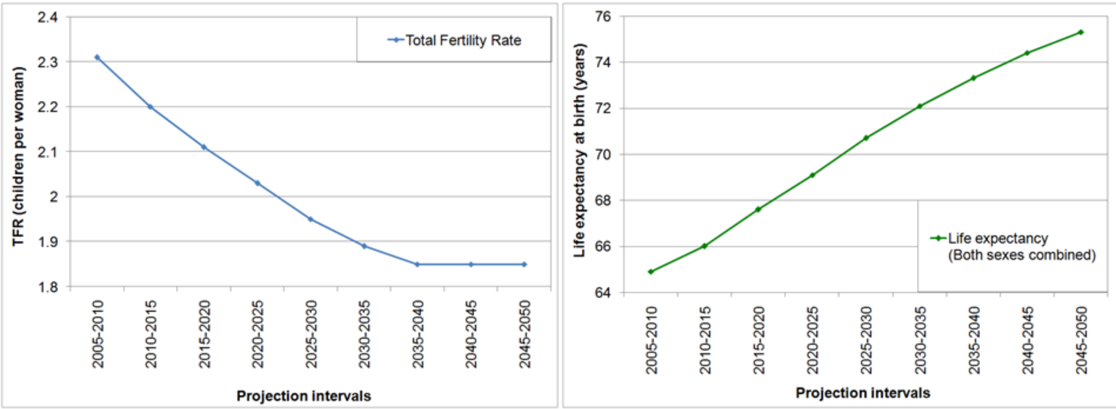
In setting assumptions for principal variant the United Nations population projections assumptions are taken as a benchmark, namely pace of change of demographic components according to World Population Prospects, the 2008 Population Revision Database served as an example in suggesting future development of fertility, mortality and migration in Kazakhstan.

With regard to medium variant of UN population projections Fig. 59 presents anticipated steady decline of future development of fertility for Kazakhstan during the period 2005-2035 and remaining constant for the period 2035-2050.

According to UN medium variant for mortality assumptions, life expectancy will increase constantly in future and this increase will slowdown from 2.5 percent to 1.4 percent as expected improvement of people longevity (Fig. 59).

Under normal migration assumptions, the future net migration for Kazakhstan is kept constant over most of the projected period.

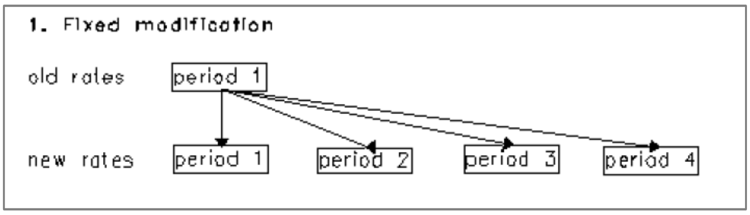
Fig. 59 - Projected Total Fertility Rate and Life expectancy for Kazakhstan, medium variant of UN population projections, 2005-2050



Source of data: World Population Prospects: the 2008 Population Revision Database

For implementing changing rates for population projections fixed scenario modification is chosen, in which initial rates for the first projection interval are served for obtaining new rates of the output rates file. Fig. 60 shows how this mechanism works in Lipro.

Fig. 60 - Schema of fixed rate modification in Lipro



Source: Van Imhoff 1999

In principal variant mortality will decline by 20 % over the period 2004-2059. The decline is uniform over age groups, and it will occur linearly over time. Thus, in 2004-2008 the mortality rates are 100% of their 2004-2008 levels, in 2009-2013 98%, 2014-2018 96%,2019-2023 94%, 2024-2028 92%, 2029-2033 90%, and so on, in the last interval 2054-2059 mortality are 80% of the 2004-2009 levels.

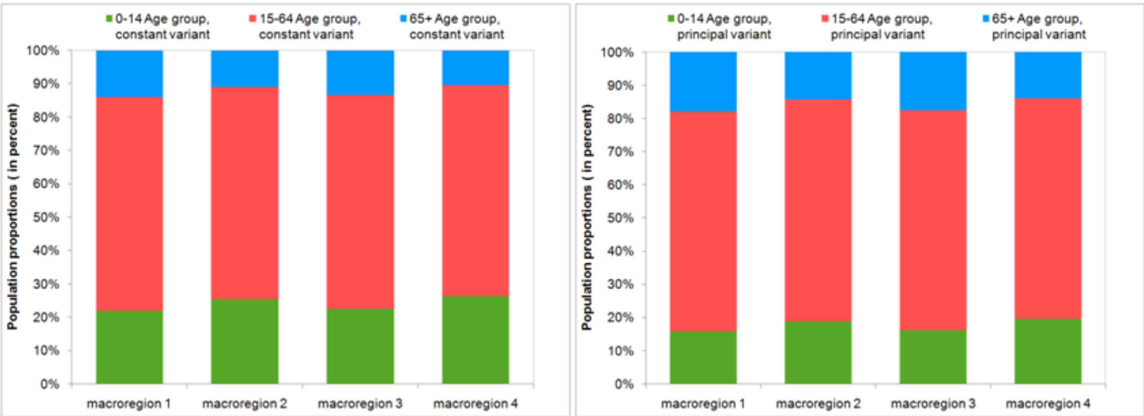
Opposite to mortality improvement fertility rates will decrease over time. Age pattern of fertility will be kept without changes, but values for period TFR will be specified with regard to eleven projection intervals: 2.4, 2.4, 2.3, 2.2, 2.1, 2.0, 1.9, 1.8, 1.7, 1.7, 1.7.

Interregional migration rates as well as net migration numbers will be constant corresponding to the level of the observation period.

8.7 Results

Fig. 61 compares final populations of constant variant with final population of principal variant. This picture shows that populations are divided into three age groups, 0-14, 15-64 and 65+. Proportions of youth generations are bigger in constant variant, whereas principal variant has bigger proportions of elderly populations.

Fig. 61 – Three age groups of final population by macroregions, constant and principal variants, 01.01.2059



Distribution of final population over age groups separately for constant and principal variants are displayed in Fig. 62 and Fig. 63.

Constant variant in Fig. 62 shows that highest population proportions are concentrated in younger ages, four regions totally agree in this conclusion, which means, that if fertility and mortality are left unchanged, then anticipated high proportions of younger generations means future financial expenses and efforts spent by country and regions to raise these generations.

Final population of principal variant in Fig. 63 reveals high concentration of middle-aged population and small proportions of youth, which means that after additional 10 years it could bring difficulties for pension system and challenge to social security system.

Next picture shows that youth dependency are higher for constant variant, and macroregion 4 is a possessor of the highest youth dependency for constant and principal variants. Old dependency is higher for principal variant, and in macroregion 1 and macroregion 3 old dependency prevails over youth dependency. Total dependency is lower in principal variant.

According to Fig. 65 and Fig. 66 mean and median ages of final population are higher than mean and median ages of starting population. Macroregion 1 has the highest mean and median ages in starting population, the highest median age for constant variant. The highest median age for principal variant belongs to macroregion 3.

Fig. 62 - Population by macroregions, constant variant, 01.01.2059

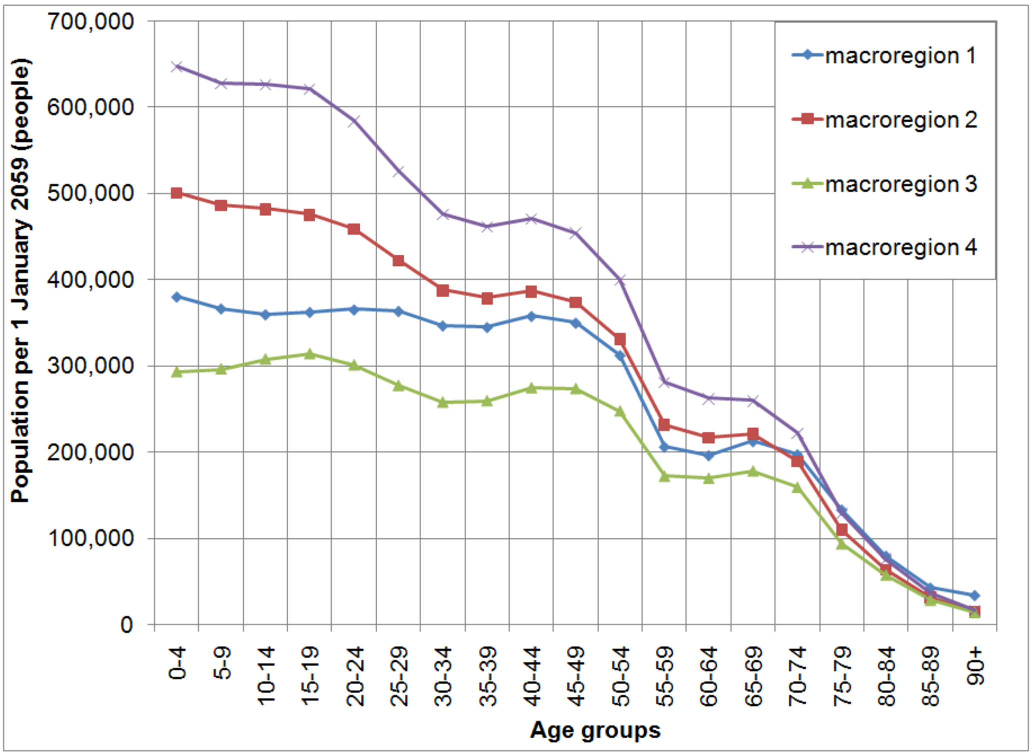


Fig. 63 - Population by macroregions, principal variant, 01.01.2059

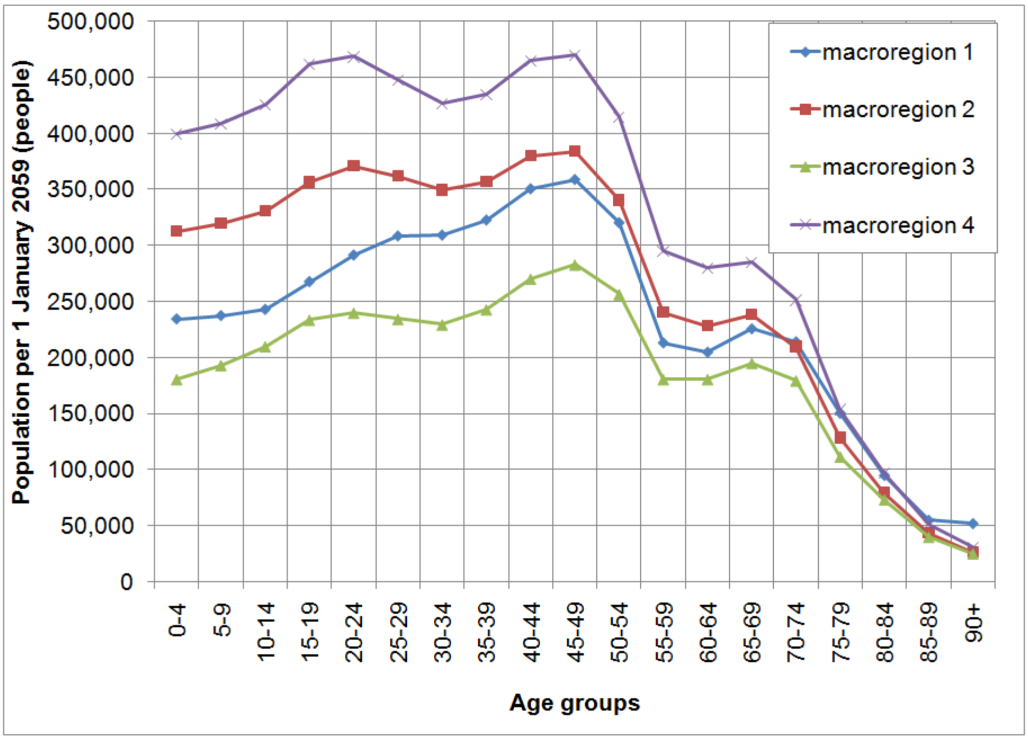


Fig. 64 –Youth dependency, old dependency, and total dependency by macroregions, constant and principal variants, 01.01.2059

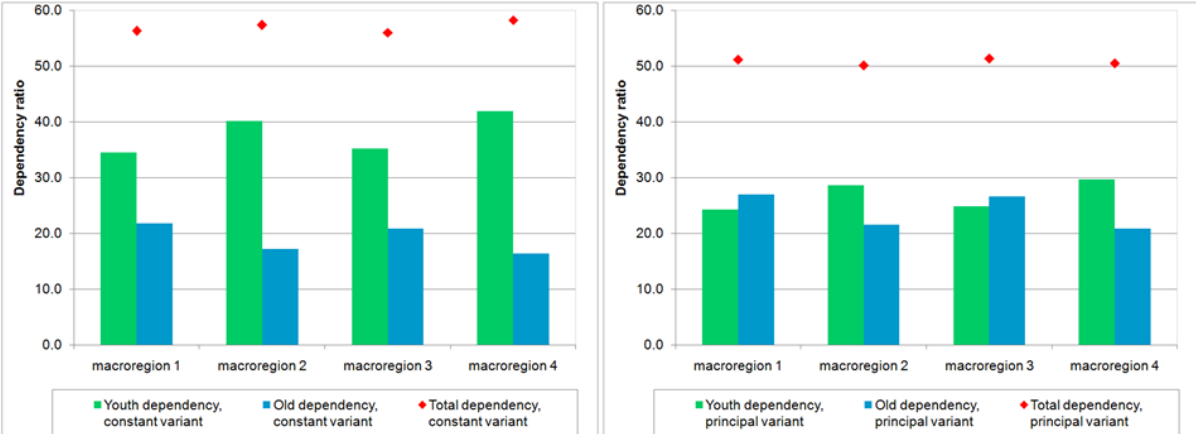


Fig. 65 – Mean age by macroregions, constant and principal variants, 2004 and 2059 years

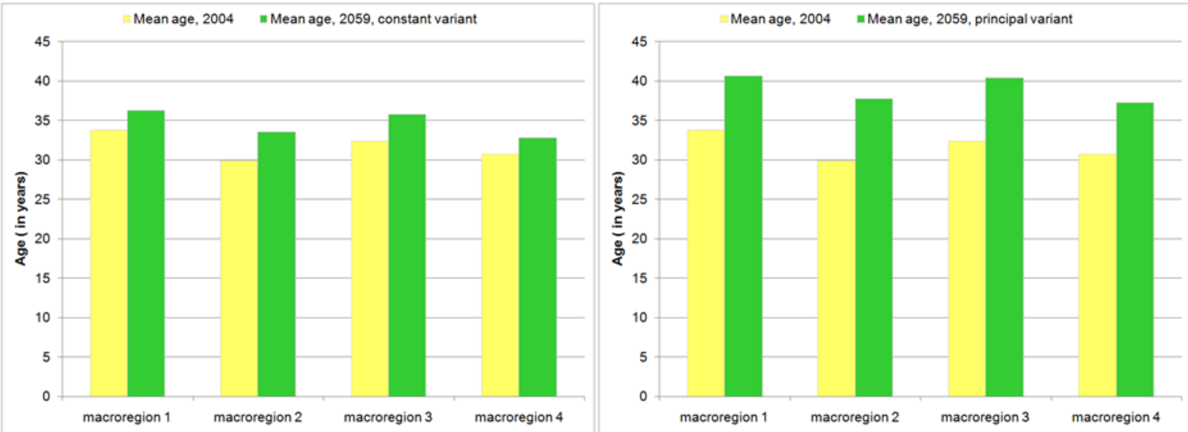
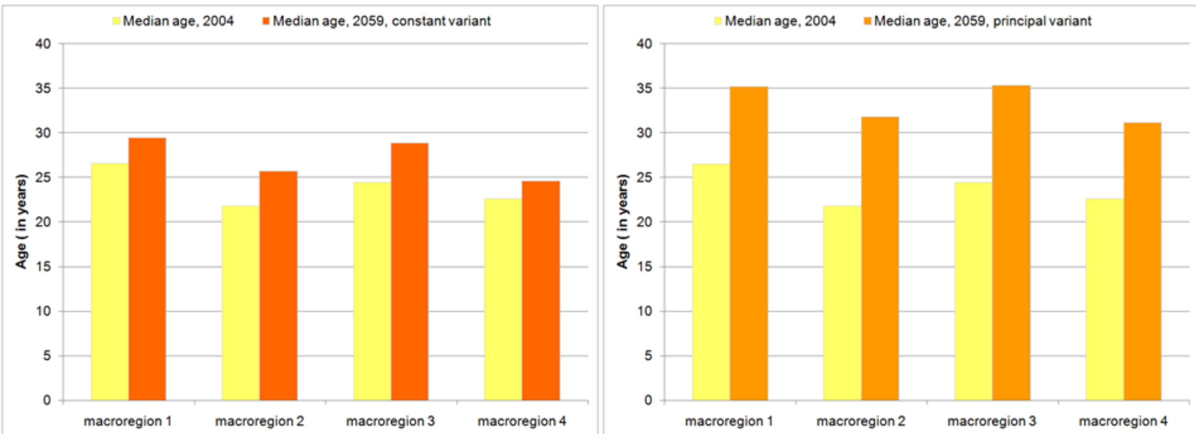


Fig. 66 – Median age by macroregions, constant and principal variants, 2004 and 2059 years



8.8 Passive - dominant consistency relations

For constant and principal variants the same set of passive and dominant consistency relations are implemented on births and deaths. Assuming that interregional migration data always contain some inaccuracy, as well complex way of deriving and inferring final interregional rates which included rounding error of small numbers, projected numbers of births and deaths for the country as a whole are chosen as dominant constraints which estimate future numbers of birth and death events ignoring effect of interregional migration as well as regional inequality of demographic components. Imposed consistency relations imply that sum of projected number of births for four macroregions have to be equal to projected number of births for country as a whole, while sum of the projected total number of deaths in four macroregions is restricted to be equal to the country level of number of deaths.

Tab. 13 and Tab. 14 show initial and adjusted births and deaths in four macroregions for females and males of constant variant along 11 projection intervals. Tab. 15 and Tab. 16 present projected births and deaths events before and after setting consistency for principal variant. Imposing passive-dominant consistency relations means that linear combination of passive events after adjustment should be equal dominant events, when dominant events are left unchanged. For example, in Tab. 13 during interval 2004-2009 initial number of deaths in four macroregions are 36056, 59755, 114365 and 124200, after adjustment they are changed to 35986, 59639, 114141 and 123958, and sum of adjusted numbers gives dominant event, 333723 deaths in country population projections. Summarizing adjustments made due consistency bounds it is clear that along the time births of four macroregions after adjustment are decreased, while deaths events are increased to meet consistency constraints.

8.8.1 Results of imposing consistency constraints

Imposed consistency restraints on flow data influence in result stock data. How constrained and unconstrained projections for four macroregions of constant variant relate to each other along 11 projections intervals is shown in Fig. 67. Unconstrained projections predict higher estimation of expected populations, whereas constrained on births and deaths projections show slightly lower results. Nevertheless, the impact of imposing constraints on births and death is insignificant in comparison with the size of the populations, so the projected populations mostly coincide over most of the period 2004-2059.

The fact, that principal variant gives smaller population projections than constant not only for last interval, but for the whole period is depicted in Fig. 68.

The third macroregion started as the second large by the size of population, in the final population has become the smallest by the size of population. These results are shown by both variants, which means that observed migration outflows will reduce population size of this macroregion even if fertility will remain at the level of the year 2008.

The first macroregion started as the smallest at the beginning will grow almost linearly in constant variant and slightly modest in principal variant. In final population this macroregion has won bronze medal by the size due to migration inflows.

At the start of projection the third macroregion was second small by the size of the population and gradually this macroregions has grown and has become the silver prize winner. But for this macroregion there are bigger differences between constant and principal variants, which means growth of population is conditioned by positive net migration and high fertility.

The fourth macroregion, the largest at the beginning according to the size of population, will remain the largest in the final population. The constant variant has demonstrated comparatively slow growth of population. But the principal variant reveals decline of population, in other words, even if net migration will remain at the level of the year 2008 and mortality will decrease, changes in fertility will cause decline of population in this macroregion.

Fig. 69 and Fig. 70 shows age pyramids of four macroregions after imposing constraints on births and deaths. All age pyramids for constant variant in Fig 69 have wide bottom bands. This variant overestimates expected growth of population in macroregions, namely increasing proportions of younger generations, since it is based on demographic rates computed using data before economic crisis with flourishing and stimulated fertility and decreasing mortality.

In case of principal variant, the shapes of pyramids of final population depicted in Fig. 70 reveal signs of ageing population with impact of decreasing mortality and falling fertility, and show that big generations born during period of increasing fertility approach their preretirement age 50-54.

The last two pictures Fig. 71 and Fig. 72 present differences between constrained and unconstrained populations for constant and principal variants separately at the end of projections. Four macroregions in both pictures unanimously demonstrate higher results for population projections without consistency relations. Population estimations for principal variants are lower in comparison with constant variant.

In summary, both variants suggest constant level of interregional migration and external net migration, since it is very difficult to predict future flows, directions, preferences, intensity, duration, age profile and gender differences of future migration development. Demographic components like fertility, mortality and migration affect future population development, and at the same time depend on many factors. Implications of population projections for sustainable regional population policy can be seen in diminishing pressure of pushing factors of internal migration, improving living conditions of people in remote settlements, supporting single and working mothers, improving quality of health care system, building educational facilities and supporting elderly population.

Fig. 67 - Projected populations by macroregions with and without imposed consistency, constant variant, 2004-2059

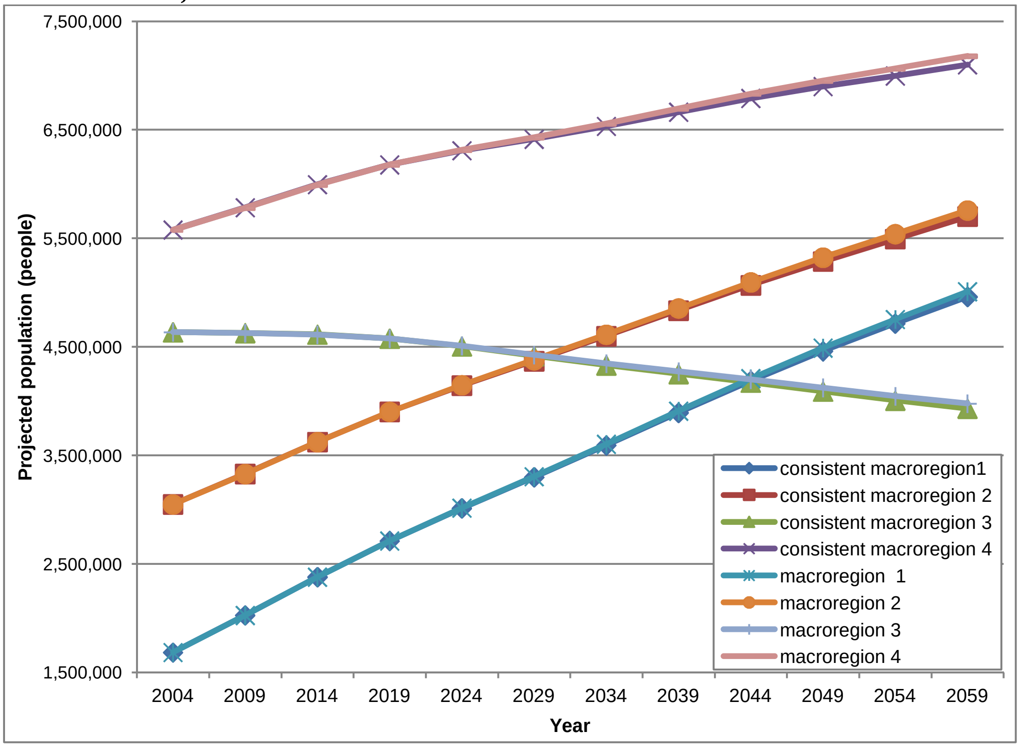


Fig. 68 - Projected constrained populations by macroregions for principal and constant variants, 2004-2059

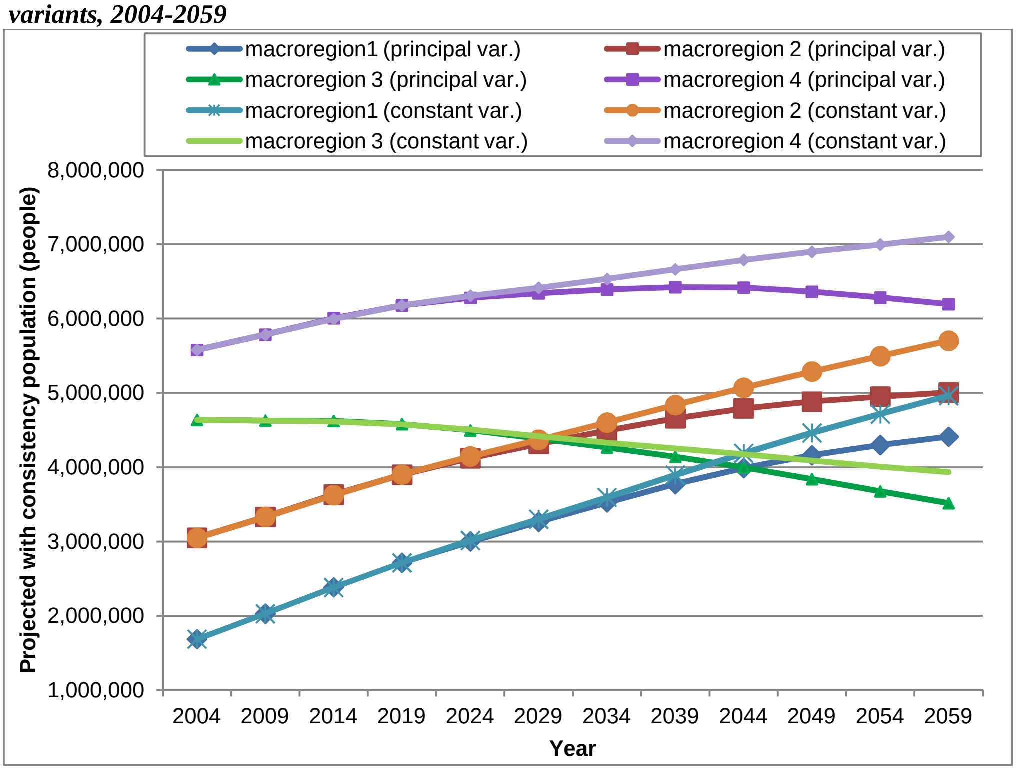


Fig. 69 - Age pyramid of final population by macroregions, constant variant with imposed constraints, 01.01.2059

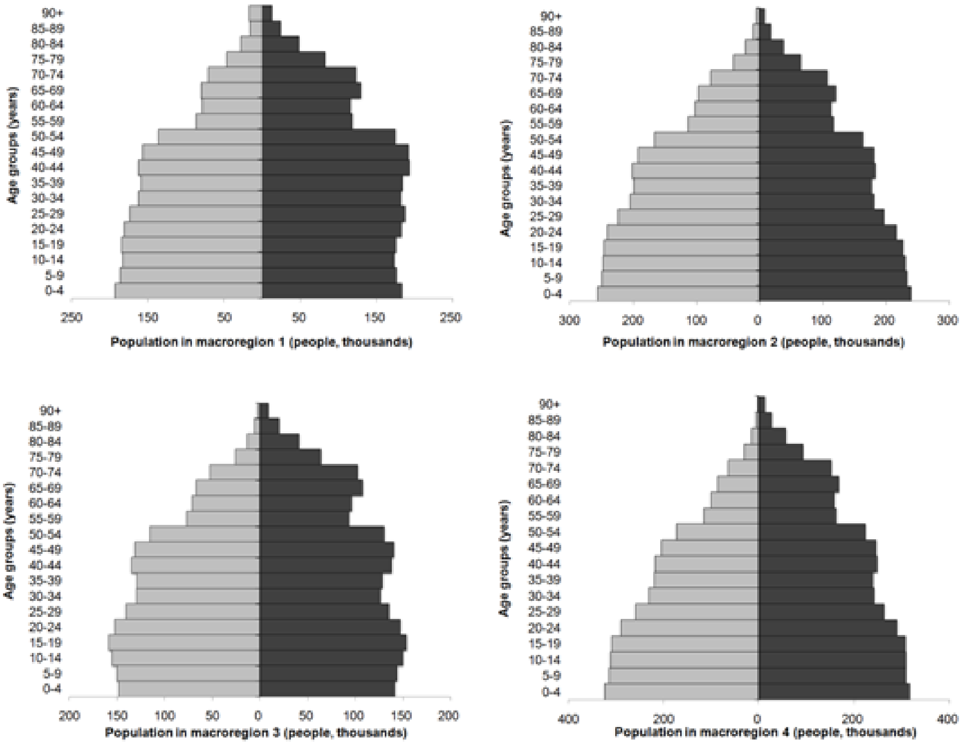


Fig. 70 - Age pyramid of final population by macroregions, principal variant with imposed constraints, 01.01.2059

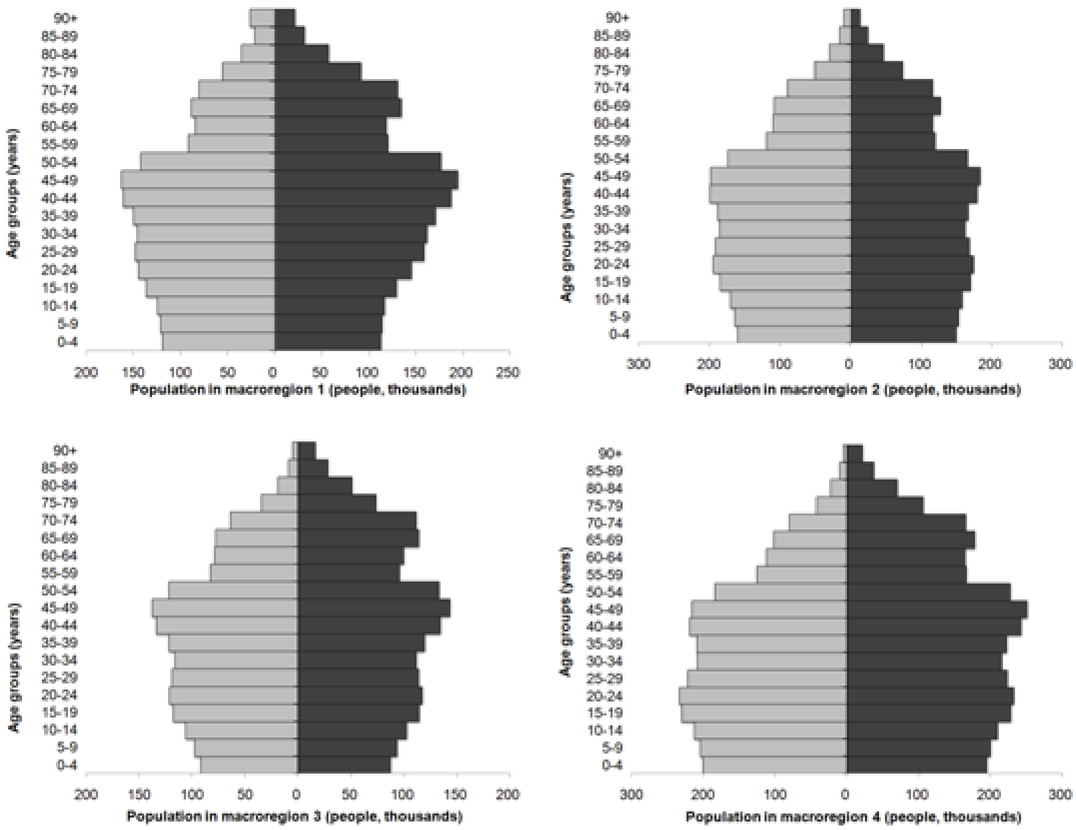


Fig. 71 - Final population by macroregions before and after consistency, constant variant, 01.01.2059

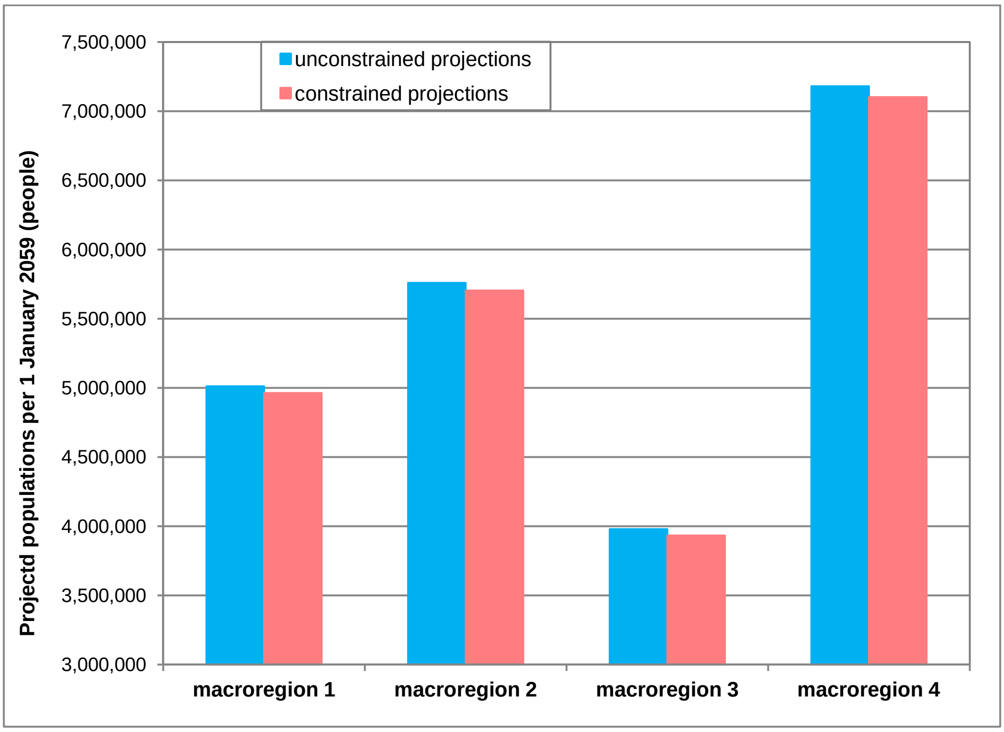
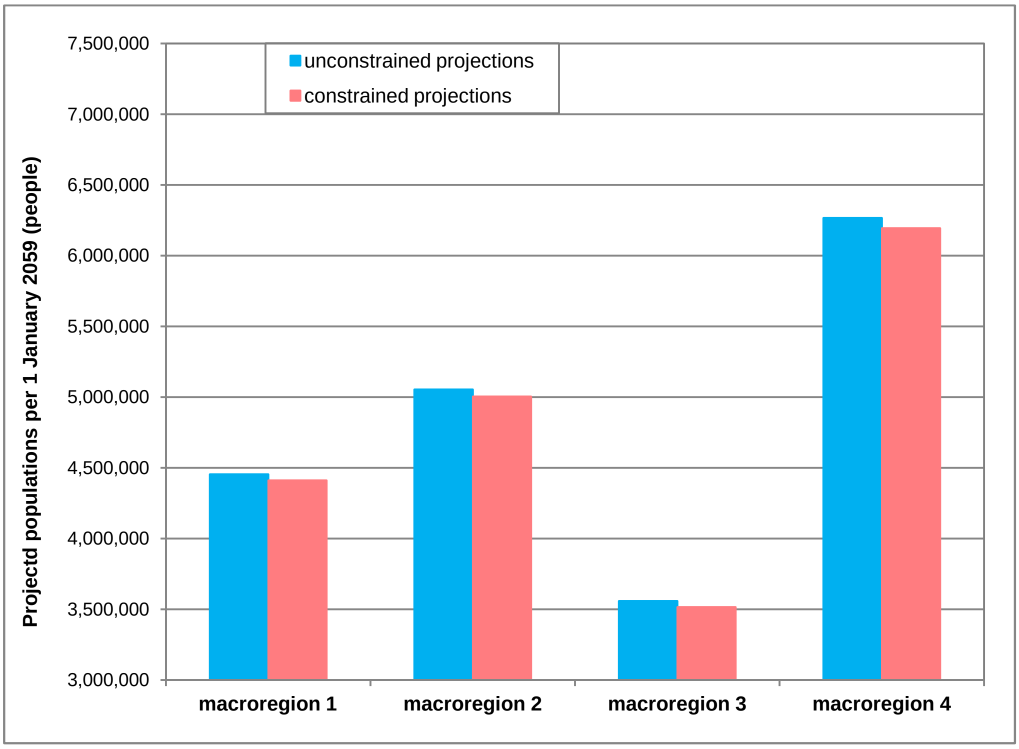


Fig. 72 - Final population by macroregions before and after consistency, principal variant, 01.01.2059



Tab. 13 - Regional births and deaths for all projected periods before and after imposing consistency, constant variant, females

		Before consistency				KZ	After consistency			
		macro region 1	macro region 2	macro region 3	macro region 4		macro region 1	macro region 2	macro region 3	macro region 4
2004-2008	Deaths	36,056	59,755	114,365	124,200	333,723	35,986	59,639	114,141	123,958
	Births	95,543	166,143	195,715	286,243	744,338	95,632	166,298	195,898	286,510
2009-2013	Deaths	42,762	67,085	122,078	136,565	368,846	42,804	67,149	122,196	136,697
	Births	108,810	183,872	196,670	303,439	792,469	108,766	183,797	196,590	303,316
2014-2018	Deaths	49,494	74,175	127,578	147,328	400,122	49,686	74,463	128,073	147,900
	Births	113,002	187,204	183,993	298,586	781,264	112,783	186,840	183,635	298,006
2019-2023	Deaths	56,298	81,189	130,468	156,046	427,228	56,726	81,807	131,460	157,234
	Births	112,389	180,825	165,069	280,844	736,451	111,982	180,171	164,471	279,827
2024-2028	Deaths	65,108	90,099	135,083	166,593	461,989	65,836	91,106	136,592	168,455
	Births	120,076	182,940	156,193	278,761	734,855	119,569	182,168	155,533	277,584
2029-2033	Deaths	75,350	100,333	140,014	177,790	500,954	76,490	101,851	142,132	180,480
	Births	136,241	197,436	158,157	295,594	784,763	135,780	196,768	157,621	294,593
2034-2038	Deaths	87,361	111,972	145,811	190,711	545,166	88,879	113,918	148,345	194,025
	Births	153,262	214,344	160,572	312,533	837,909	152,752	213,630	160,036	311,491
2039-2043	Deaths	99,736	122,665	149,429	201,489	584,281	101,643	125,010	152,286	205,341
	Births	164,674	225,816	158,383	318,958	864,614	164,064	224,979	157,796	317,775
2044-2048	Deaths	111,790	131,468	149,938	209,198	616,083	114,330	134,456	153,345	213,952
	Births	170,249	230,245	151,819	316,428	864,906	169,498	229,228	151,149	315,032
2049-2053	Deaths	123,605	138,562	148,129	214,317	641,040	126,856	142,206	152,025	219,954
	Births	175,521	234,313	146,172	316,020	867,704	174,651	233,152	145,448	314,453
2054-2058	Deaths	135,161	145,019	145,129	218,197	662,064	139,059	149,202	149,314	224,490
	Births	184,851	243,169	144,400	323,363	891,162	183,897	241,914	143,655	321,695

Tab. 14 - Regional births and deaths for all projected periods before and after imposing consistency, constant variant, males

		Before consistency				KZ	After consistency			
		macro region 1	macro region 2	macro region 3	macro region 4		macro region 1	macro region 2	macro region 3	macro region 4
2004-2008	Deaths	38,156	67,520	150,518	186,329	437,773	37,746	66,795	148,903	184,329
	Births	102,009	176,259	205,955	303,003	787,950	102,103	176,422	206,144	303,282
2009-2013	Deaths	46,790	78,572	156,126	197,017	476,982	46,641	78,322	155,629	196,390
	Births	116,175	195,210	206,954	321,106	838,872	116,095	195,077	206,813	320,887
2014-2018	Deaths	55,696	89,823	159,997	206,114	513,771	55,929	90,199	160,667	206,976
	Births	120,679	198,601	193,816	315,710	826,863	120,396	198,135	193,362	314,970
2019-2023	Deaths	64,578	100,582	162,057	212,425	546,034	65,343	101,774	163,976	214,941
	Births	120,078	191,530	173,929	297,046	779,371	119,585	190,744	173,215	295,827
2024-2028	Deaths	74,532	112,515	165,273	220,035	583,517	75,986	114,709	168,496	224,326
	Births	128,323	193,823	164,367	295,151	777,830	127,693	192,873	163,561	293,704
2029-2033	Deaths	84,712	124,112	166,498	226,449	617,266	86,894	127,308	170,785	232,280
	Births	145,542	209,513	166,328	313,015	830,828	144,919	208,617	165,616	311,676
2034-2038	Deaths	95,502	136,118	167,349	233,290	652,201	98,514	140,411	172,627	240,649
	Births	163,648	227,538	169,026	330,671	887,021	162,939	226,552	168,293	329,237
2039-2043	Deaths	105,749	147,248	166,173	238,014	681,246	109,620	152,640	172,257	246,729
	Births	175,813	239,539	166,802	337,386	915,142	174,972	238,393	166,004	335,773
2044-2048	Deaths	115,330	157,228	164,015	241,139	706,152	120,170	163,826	170,898	251,258
	Births	181,831	244,121	159,867	334,779	915,429	180,810	242,750	158,969	332,899
2049-2053	Deaths	124,602	166,745	161,853	244,371	730,093	130,411	174,519	169,399	255,764
	Births	187,503	248,416	153,874	334,447	918,431	186,324	246,855	152,907	332,345
2054-2058	Deaths	133,400	174,951	158,593	246,422	749,143	140,090	183,726	166,547	258,781
	Births	197,453	257,905	151,978	342,257	943,337	196,152	256,206	150,977	340,002

Tab. 15 - Regional births and deaths for all projected periods before and after imposing consistency, principal variant, females

		Before consistency				KZ	After consistency			
		macro region 1	macro region 2	macro region 3	macro region 4		macro region 1	macro region 2	macro region 3	macro region 4
2004-2008	Deaths	36,056	59,755	114,365	124,200	333,723	35,986	59,639	114,141	123,958
	Births	95,543	166,143	195,715	286,243	744,338	95,632	166,298	195,898	286,510
2009-2013	Deaths	42,027	65,937	120,014	134,236	362,549	42,066	65,998	120,125	134,360
	Births	109,505	185,050	197,930	305,383	797,545	109,461	184,975	197,850	305,259
2014-2018	Deaths	47,921	71,814	123,720	142,706	387,599	48,099	72,081	124,181	143,238
	Births	109,002	180,591	177,493	288,038	753,657	108,791	180,240	177,148	287,478
2019-2023	Deaths	53,830	77,617	125,222	149,408	409,058	54,225	78,187	126,141	150,505
	Births	103,721	166,905	152,360	259,221	679,737	103,346	166,300	151,809	258,282
2024-2028	Deaths	61,539	85,141	128,509	157,896	437,712	62,196	86,050	129,882	159,583
	Births	105,824	161,264	137,688	245,724	647,753	105,377	160,583	137,106	244,687
2029-2033	Deaths	70,445	93,798	132,167	166,949	470,111	71,471	95,164	134,093	169,382
	Births	114,438	165,878	132,894	248,402	659,376	114,051	165,318	132,445	247,562
2034-2038	Deaths	80,884	103,731	136,766	177,632	507,290	82,226	105,451	139,034	180,579
	Births	121,391	169,864	127,198	247,608	663,874	120,992	169,306	126,780	246,795
2039-2043	Deaths	91,737	113,003	139,864	186,751	540,773	93,363	115,006	142,343	190,061
	Births	121,208	166,333	116,355	234,336	635,873	120,760	165,718	115,925	233,470
2044-2048	Deaths	102,348	120,735	140,510	193,423	568,657	104,487	123,259	143,446	197,465
	Births	114,733	155,575	101,967	212,561	582,221	114,220	154,880	101,511	211,611
2049-2053	Deaths	112,683	126,769	138,816	197,515	589,740	115,414	129,842	142,181	202,303
	Births	113,419	152,307	94,021	203,556	560,415	112,837	151,526	93,539	202,512
2054-2058	Deaths	122,592	131,822	135,592	199,908	605,808	125,895	135,374	139,245	205,294
	Births	113,622	150,934	88,246	198,294	548,094	113,003	150,112	87,765	197,214

Tab. 16 - Regional births and deaths for all projected periods before and after imposing consistency, principal variant, males

		Before consistency				KZ	After consistency			
		macro region 1	macro region 2	macro region 3	macro region 4		macro region 1	macro region 2	macro region 3	macro region 4
2004-2008	Deaths	38,156	67,520	150,518	186,329	437,773	37,746	66,795	148,902	184,329
	Births	102,009	176,259	205,955	303,003	787,950	102,103	176,422	206,144	303,282
2009-2013	Deaths	45,934	77,152	153,397	193,579	468,536	45,785	76,902	152,899	192,950
	Births	116,917	196,461	208,280	323,163	844,245	116,837	196,327	208,138	322,943
2014-2018	Deaths	53,657	86,743	154,825	199,125	496,550	53,896	87,129	155,514	200,011
	Births	116,408	191,585	186,969	304,557	797,644	116,135	191,136	186,531	303,843
2019-2023	Deaths	61,201	95,738	154,888	202,329	520,598	61,968	96,937	156,829	204,864
	Births	110,818	176,785	160,539	274,175	719,351	110,363	176,059	159,880	273,050
2024-2028	Deaths	69,615	105,768	156,408	207,070	550,028	71,058	107,960	159,650	211,361
	Births	113,092	170,857	144,893	260,173	685,636	112,537	170,019	144,183	258,897
2029-2033	Deaths	78,061	115,367	156,282	210,741	575,983	80,225	118,564	160,613	216,581
	Births	122,250	176,028	139,758	263,036	698,076	121,728	175,275	139,161	261,912
2034-2038	Deaths	86,911	125,254	156,006	214,833	603,059	89,901	129,563	161,372	222,223
	Births	129,616	180,324	133,903	261,951	702,771	129,061	179,551	133,329	260,829
2039-2043	Deaths	95,148	134,284	153,968	216,921	624,677	99,009	139,732	160,215	225,722
	Births	129,410	176,416	122,568	247,821	672,997	128,794	175,577	121,985	246,642
2044-2048	Deaths	102,643	142,121	151,050	217,363	642,248	107,509	148,859	158,211	227,669
	Births	122,549	164,900	107,396	224,844	616,184	121,856	163,967	106,788	223,572
2049-2053	Deaths	109,789	149,367	148,132	217,902	658,684	115,671	157,369	156,068	229,576
	Births	121,183	161,415	98,994	215,385	593,130	120,402	160,375	98,356	213,997
2054-2058	Deaths	116,197	155,081	144,043	216,780	669,325	123,040	164,214	152,525	229,546
	Births	121,394	160,017	92,898	209,833	580,134	120,561	158,919	92,261	208,393

Conclusion

This dissertation presents estimation of future population development in regions of Kazakhstan by analyzing previous regional dynamical evolvement of population and its demographic components. Despite some spatial differentiation there are some features and trends common for all regions. Like any other country of post soviet territory Kazakhstan is recovering after loss of population due to collapse of USSR, and demographic components have experienced fluctuations during past development. Mortality was increasing and it is decreasing now, fertility fell and raised and again according to the last data is slightly decreasing. Interregional migration flows have affected concentration and distribution inhabitants within country, whereas external migration flows are characterized by past exodus of emigrants and by implementation of ethnic migration policy program.

Both practical implementations of population projections for Kazakhstan are based on subjective judgment, nevertheless, chapter 2 describes diversity of methods and recent achievements in prediction of demographic components as well as population including stochastic population forecast. Population projection for 16 administrative regions and population projection for four macroregions apply multiregional approach. At the same time chapter 3 describes variety other models for regional population projections. Methods described in chapter 6 for inferring interregional migration flows in general way can be used for updating interregional migration data between censuses. Chapter 6 also describes consistency relations between regional and national levels of population projections, and chapter 8 presents results of implementing consistency constraints on birth and death events.

The first example of population projection for 16 administrative regions with 20 years prediction horizon shows example of benefits of demographic bonus of recovering fertility.

At the same time projections for four macroregions for period 2004-2059 show what will happen when population of people born during period of recovering fertility will be approaching retirement age. These generations are comparatively numerous, and old dependency will prevail over youth dependency in some regions.

Anticipation of population ageing allows to identify emerging problem than to simply deny it and be unprepared. Problem of population ageing raises many questions. What are demographic imperatives of increasing proportions of elders and decreasing proportion of children? Demographic and epidemiological transition will increase life expectancy at birth, what will be disability free life expectancy? Does it mean living longer, but better? What will be income of elderly population? How elderly population will sustain physical and mental health? How to provide long term care for people with chronic diseases, necessary drugs, surgical and technical procedures to treat age-related illnesses? And how elder women will survive? How to reduce poverty and poor health of widows? How frail older persons being left isolated in rural areas will have access to medical care?

Across the world developed countries are implementing active ageing policy, there are changes in governmental expectations and responsibilities, when states make efforts to reduce welfare expenditures by shifting them from public to private spendings. Other steps include increase of retirement age, sharing cost by the state, the employer and the employee, part-time work after official retirement. What lessons will be replicated in Kazakhstan? What solutions will be found on national and regional levels?

REFERENCES

Abishev, T. (2010). *Human Rights Commission, Kazakhstan.* Statement issued at the 2010 OSCE Review Conference. Working session: Humanitarian issues and other commitments, including: refugees and displaced persons; treatment of citizens of other participating States; human rights education. Warsaw, 5 October.

Alho J.M. and Spencer B.D. (2005). *Statistical demography and forecasting*. New York: Springer.

Assanbayev, M. (2010) *Analysis of internal migration processes: Conclusions, measures, recommendations.* Almaty. Soros Foundation.

Bapakova, S. (2010). Migration as a theme of round-table discussion. Press service of the Parliament of the Republic of Kazakhstan.http://www.nomad.su/?a=3-201004160035

Bell, W. (1992). ARIMA and principal component models in forecasting age-specific fertility. In N. Keilman, and H. Cruijsen (Eds.), National population forecasting in industrialized countries (pp. 178 – 200). Amsterdam: Swets and Zeitlinger.

Bell, W. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, 13, 279– 303.

Bell M., Blake M., Boyle P., Duke-Williams O., Rees P., Stillwell J. and Hugo G.(2002) Cross-national comparison of internal migration: issues and measures. *Royal Statistical Society* 165(3): 435-464.

Bell, W. R., & Monsell, B. C. (1991). *Using principal components in time series modeling and forecasting of age-specific mortality rates.* Paper presented at the American Statistical Association 1991 Proceedings of the Social Statistics Section.

Bijak, J. (2007). *Bayesian Model Selection in Forecasting International Migration: Simple Time Series Models and Their Extensions*. Paper presented at the Joint Eurostat-the United Nations Economic Commission for Europe (UNECE) Work Session on Demographic Projections, Bucharest, 10-12 October.

Biswas T., McHardy J. and Nolan M.A. (2009) Inter-regional migration: the UK experience. Department of economics University of Sheffield (Sheffield Economic Research Paper Series; SERP Number: 2008003.

Bleha, B. (2007). *Conception of spatial units appropriate for regional population forecasts*. Paper presented at the Joint Eurostat-the United Nations Economic Commission for Europe (UNECE) Work Session on Demographic Projections: 319-326.

Bleha, B. and Vaňo, B. (2010). *How to deal with sub-national forecasts in spatially very heterogeneous countries? Towards using some spatial theories and models.* Paper presented at the Joint Eurostat-the United Nations Economic Commission for Europe (UNECE) Work Session on Demographic Projections, Lisbon, 28-30 April: 309-317.

Bongaarts J. (2005). Long-range trends in adult mortality: models and projection methods. *Demography* 42(1): 23-49.

Bongaarts, J., & Feeney, G. (1998). On the quantum and tempo of fertility. *Population and Development Review*, 24, 271– 291.

Bongaarts, J., & Feeney, G. (2002). How long do we live? *Population and Development Review*, 28, 13– 29.

Bongaarts, J., & Feeney, G. (2003). Estimating mean lifetime. *Proceedings of the National Academy of Sciences*, 100, 13127– 13133.

Bongaarts, J., & Feeney, G. (2005). The quantum and tempo of lifecycle events (Working Paper No. 207). New York: Population Council.

Booth H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* 22: 547- 581.

Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, 56, 325–336.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), Biological *aspects of demography* (pp. 69 – 110). London: Taylor and Francis.

Brass, W. (1974). Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society Series A*, 137, 532– 583.

Brass, W. (1981). The use of the Gompertz relational model to estimate fertility. *International Population Conference,* Manila, vol. 3 (pp. 345–362).

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson logbilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31, 373–393.

Burcin, B. and Kučera, T. (2003). Changes in fertility and mortality in the Czech Republic: An attempt of regional demographic analysis. In: Kučera, T. Kučerová, O., Opara O., Schaich E. (eds.). New demographic faces of Europe. Springer: 371-417.

Caselli, G. (1996). Future longevity among the elderly. In G. Caselli & A. Lopez (Eds.), *Health and mortality among elderly populations*. Oxford: Clarendon Press.

Caselli, G. (2002). Age, period and cohort components in analyzing past and projecting future mortality trends. In G. Wunsch, M. Mouchart, and J. Duchene (Eds.), *The life table: Modelling survival and death* (pp. 117– 140). Dordrecht: Kluwer Academic Publishers.

Chandola, T., Coleman, D. A., & Hiorns, R. W. (1999). Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Population Studies*, 53, 317– 329.

Chandola, T., Coleman, D. A., & Hiorns, R. W. (2000). A reply to J. Antonio Ortega Osona and H. P. Kohler. *Population Studies*, 54, 351.

Coale, A. J., Demeny, P. G., & Vaughan, B. (1983). *Regional model life tables and stable populations* (2nd ed.). New York: Academic Press.

Coale, A. J., and Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index*, 55, 613- 643.

Coale, A. J., and Kisker, E. E. (1990). Defects in data on old-age mortality in the United States: New procedures for calculating schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4, 1 –31.

Coale, A. J., and McNeil, D. R. (1972). The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association*, 67, 743- 749.

Coale, A. J., and Trussell, T. J. (1974). Model fertility schedules: Variations in the age structure of childbearing in humanp opulations. *Population Index*, 40, 185- 258.

Coale, A. J., Demeny, P. G., and Vaughan, B. (1983). *Regional model life tables and stable populations* (2nd ed.). New York: Academic Press.

Congdon, P. (1990). Graduation of fertility schedules: An analysis of fertility patterns in London in the 1980s and an application to fertility forecasts. *Regional Studies*, 24, 311–326.

Crimmins, E. M. (1981). The changing pattern of American mortality decline, 1947–1977, and its implication for the future. *Population and Development Review*, 7, 229– 254.

De Beer, J. (1988). Predictability of demographic variables in the short run. *European Journal of Population*, 4, 283– 296.

De Beer, J. (1989). Projecting age-specific fertility rates by using time-series methods. *European Journal of Population*, 5, 315– 346.

De Beer, J. (1991). From birth expectations to birth forecasts: A partial adjustment approach. *Mathematical Population Studies*, 3, 1127–1144.

De Beer, J. (1997). The effect of uncertainty of migration on national population forecasts: The case of the Netherlands. *Journal of Official Statistics*, 13, 227–243.

De Beer, J. (2000). Dealing with uncertainty in population forecasting: Statistics Netherlands, Department of Population.

De Beer, J., & Alders, M. (1999). Probabilistic population and household forecasts for the Netherlands. Working Paper No. 45. Paper presented at the Joint ECE-Eurostat Work Session on Demographic Projections, 3–7 May 1999, Perugia, Italy.

De Beer, J., & Van Wissen, L. (Eds.) (1999). *Europe: One continent, different worlds; population scenarios for the 21st century*. Dordrecht: Kluwer.

Demographic Yearbook of the Republic of Kazakhstan for the year 2007 (2008). The Agency on Statistics of the Republic of Kazakhstan, Astana

Demographic Yearbook of the Republic of Kazakhstan for the year 2008 (2009). The Agency on Statistics of the Republic of Kazakhstan, Astana

Divino F., Egidi V. and Salvatore M.A. (2009). Geographical mortality patterns in Italy: A Bayesian analysis. Rostock: Max Planck Institute for Demographic Research. *Demographic research* 20(18):435-466. http://www.demographic-research.org/Volumes/Vol9/1/.

Denisenko M.B. and Kalmykova H.M. (2007) *Demography*. Moscow, Infra-M.

Easterlin, R. A. (1968). Population, labour force, and long swings in economic growth: The America experience. *General series*, vol. 86. .New York: National Bureau of Economic Research.

Ermisch, J. (1988). Econometric analysis of birth rate dynamics in Britain. *Journal of Human Resources*, 23, 563– 576.

Ermisch, J. (1992). Explanatory models for fertility projections and forecasts. In N. Keilman, & H. Cruijsen (Eds.), *National population forecasting in industrialized countries* (pp. 201–222). Amsterdam: Swets and Zeitlinger.

Ginsburg, D. (1972). Models of the Interaction of Language and Social Life. In: Gumperz, J. and Hymes, D. (eds.). *Directions in Sociolinguistics*. New York: Holt Rinehar: 35–71.

Girosi, F., and King, G. (2006). *Demographic forecasting*. Cambridge: Cambridge University Press.

Golulapati, R., De Ravin, J. W., & Trickett, P. J. (1984). Projections *of Australian mortality rates, 1981–2020* (occasional paper no.1983/2). Australian Bureau of Statistics.

Hanewald K. (2009) Mortality modeling: Lee-Carter and the macroeconomy. Berlin: Humboldt-Universität zu Berlin. (Discussion paper; SFB 649). http://sfb649.wiwi.hu-berlin.de

Hanika, A. (2010). *Austrian Regional Population Projections below NUTS-3*. Paper presented at the Joint Eurostat-the United Nations Economic Commission for Europe (UNECE) Work Session on Demographic Projections, Lisbon, 28-30 April.

Hyndman R.J. and Booth H. (2008). Stochastic population forecast using functional data models for mortality, fertility and migration. *International Journal of Forecasting* 24: 323-342. www.sciencedirect.com

Hyndman R.J. and Ullah M.S. (2006). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational statistics and data analysis* 51(4952-4956). www.sciencedirect.com

Jazybayeva A. (2011). *Future population development in Kazakhstan*, Paper presented at World Academy of Science, Engineering and Technology Conference, Paris.

Kazakhstan in numbers 1991-2008, The Agency on Statistics of the Republic of Kazakhstan, Astana

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

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

Keilman, N. (Ed.) (2005a). Perspectives on mortality forecasting. Probabilistic model. Stockholm: Swedish Social Insurance Agency. *Social Insurance Studies* vol.2

Keilman, N. Using Deterministic and Probabilistic Population Forecasts

Keilman, N., Pham, D. Q., and Hetland, A. (2002). Why population forecasts should be probabilistic - Illustrated by the case of Norway. *Demographic Researc*h, 6: 409-453.

Keilman, N.and Kučera, T. (1991). The impact of methodology on accuracy of population forecasts. Evidence from the Netherlands and Czechoslovakia. *Journal of Forecasting* 10. : 371-398.

Keyfitz N. and Caswell H.(2005). *Applied mathematical demography*. 3 rd edition.

Knudsen, C., McNown, R., & Rogers, A. (1993). Forecasting fertility: An application of time series methods to parameterized model schedules. *Social Science Research*, 22, 1 –23.

Kučera, T. (1998). Regionální populační prognózy: teorie a praxe prognózování vývoje lidských zdrojů v území. [Dizertační práce]. Praze:Přírodovědecká fakulta Univerzity Karlovy v Praze, Katedra demografie a geodemografie.

Kupiszewski M. and Rees P. (1999). *Lessons of internal migration from studies in ten European countries*. Paper presented at the Conference of European Statisticians, Perugia, 3-7 May.

Kupiszewski, M. and Kupiszewska, D. (2003). Internal migration component in subnational population projections in member states of the European Union. Warsaw: Central European Forum for Migration Research (CEFMR working paper; cefmr_wp_2003_02).

Lee, R. D. (1974). Forecasting births in post-transition populations: Stochastic renewal with serially correlated fertility. *Journal of the American Statistical Association*, 69, 607– 617.

Lee, R. D. (1980). Aiming at a moving target: Period fertility and changing reproductive goals. *Population Studies*, 34, 205– 226.

Lee, R. D. (1981). A model for forecasting fertility from birth expectations data. In G. E. Hendershot, & P. J. Placek (Eds.), *Predicting fertility: Demographic studies of birth expectations* (pp. 75– 99). Lexington, MA: Lexington.

Lee, R. D. (1990). Long-run global population forecasts: A critical appraisal. *Population and Development Review*, 16, 44– 71.

Lee, R. D. (1992). Stochastic demographic forecasting. *International Journal of Forecasting*, 8, 315–327.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9, 187– 202.

Lee, R. D. (1999). Probabilistic approaches to population forecasting. In W. Lutz, J. W. Vaupel, & D. A. Ahlburg (Eds.), *Frontiers of population forecasting. A supplement to Population and Development Review*, vol. 24 (pp. 156 –190). New York: Population Council.

Lee, R. D. (2000a). The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4, 80– 93.

Lee, R. D. (2000b). Long-term population projections and the US social security system. *Population and Development Review*, 26, 137–143.

Lee, R. D. (2005). Quantifying our ignorance: Stochastic forecasts of population and public budgets. In L. J. Waite (Ed.), *Aging, health, and public policy: Demographic and economic perspectives. Supplement to Population and Development Review*, vol. 30 (pp. 153– 175). New York: Population Council.

Lee, R. D., & Carter, L. R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87, 659–671.

Lee, R. D., Carter, L. R., & Tuljapurkar, S. (1995). Disaggregation in population forecasting: Do we need it? And how to do it simply? *Mathematical Population Studies*, 5, 217– 234.

Lee, R. D., & Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, 38, 537–549.

Lee, R. D., & Nault, F. (1993). Modelling and forecasting provincial mortality in Canada. Paper presented at the World Congress of the International Union of the Scientific Study of Population, Montreal. Unpublished.

Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, an low. *Journal of the American Statistical Association*, 89, 1175–1189.

Lee, R. D., & Tuljapurkar, S. (2000). Population forecasting for fiscal planning: issues and innovations. In A. Auerbach, & R. Lee (Eds.), *Demography and fiscal policy* (pp. 7 – 57). Cambridge: Cambridge University Press.

Li, N., & Wu, Z. (2003). Forecasting cohort incomplete fertility: A method and an application. *Population Studies*, 57, 303– 320.

Manton, K. G., Patrick, C. H., & Stallard, E. (1980). Mortality model based on delays in progression of chronic diseases: Alternative to cause elimination model. *Public Health Reports*, 95, 580– 588.

Manton, K., Stallard, E., & Tolley, H. D. (1991). Limits to human life expectancy: Evidence, prospects, and implications. *Population and Development Review*, 17, 603–637.

Manton, K. G., Stallard, E., & Singer, B. (1992). Projecting the future size and health status of the US elderly population. *International Journal of Forecasting*, 8, 433– 458.

Martin P. and Zürcher G. (2008). Managing migration: the global challenge. Population bulletin 63 (1).

McNown, R., & Rogers, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography*, 26, 645–660.

McNown, R., & Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8, 413–432.

McNown, R., Rogers, A., & Little, J. (1995). Simplicity and complexity in extrapolative population forecasting models. *Mathematical Population Studies,* 5, 235–257.

Muhidin S. (2002). *The population of Indonesia: Regional Demographic scenarios using a multiregional method and multiple data sources.* Amsterdam: Rozenberg Publishers.

Murphy, M. J. (1982). Gompertz and Gompertz relational models for forecasting fertility: An empirical exploration. *London School of Hygiene and Tropical Medicine*. London: Centre for Population Studies.

Murphy, M. J. (1984). Population forecasting: Alternative approaches (occasional papers). London: Office of Population Censuses and Surveys. Murphy,M. J. (1990). Methods of forecasting mortality for population projections. Population projections: Trends, methods and uses. *Occasional paper*, vol. 38 (pp. 87–102). London: OPCS.

Murphy, M. J. (1995). The prospect of mortality: England and Wales and the United States of America, 1962–1989. *British Actuarial Journal*, 1, 331–350.

Murphy, M. J. (2003). Bringing behavior back into micro-simulation: Feedback mechanisms in demographic models. *In Agent-based computational demography: Using simulation to improve our understanding of demographic behaviour* (pp. 159 – 174). Heidelberg: Physica-Verlag.

Ni, V. (2008). *Rights of migrant workers in Kazakhstan: National legislation, international standards and practices,* Almaty, International Labour Organization.

Oliveira, M. M., & Mexia, J. T. (2004). AIDS in Portugal: Endemic versus epidemic forecasting scenarios for mortality. *International Journal of Forecasting*, 20, 131–135.

Olshansky, S. J. (1987). Simultaneous/multiple cause–delay (SIMCAD): An epidemiological approach to projecting mortality. *Journal of Gerontology*, 42, 358– 365.

Olshansky, S. J. (1988). On forecasting mortality. *The Milbank Quarterly*, 66, 482–530.

Olshansky, S. J., & Carnes, B. A. (1997). Ever since Gompertz. *Demography*, 34, 1 –15.

Ortega, J. A., & Poncela, P. (2005). Joint forecasts of Southern European fertility rates with non-stationary dynamic factor models. *International Journal of Forecasting*, 21, 539–550.

Rees P.H., Wilson A.G. (1977). *Spatial population analysis.* London: Edward Arnold.

Regions of the Republic of Kazakhstan in 2008 (2009). The Agency on Statistics of the Republic of Kazakhstan, Astana

Renshaw A.E. and Haberman S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. Insurance: *Mathematics and economics* 33: 255-272. www.sciencedirect.com

Rogers, A. (1975). *Introduction to multiregional mathematical demography*. New York: Wiley.

Rogers, A. (1986). Parameterized multistate population dynamics and projections. *Journal of the American Statistical Association*, 81, 48–61.

Rogers, A. (1988). Age patterns of elderly migration: An international comparison. *Demography*, 25, 355– 370.

Rogers, A. (Ed.) (1995a). Population projections: Simple vs. complex models. Special issue of *Mathematical Population Studies*, Vol. 5(3).

Rogers, A. (1995b). Multiregional demography: Principles, methods, and extensions. London7 John Wiley.

Rogers, A. (1995c). Population forecasting: Do simple models outperform complex models? Mathematical Population Studies, 5, 187–202.

Rogers A. (1981). Advances in multiregional demography. Laxenburg: International Institute for Applied Systems Analysis (research reports RR-81-6).

Rogers A., Willekens F. and Raymer J. (2001). Modeling interregional migration flows: continuity and change. *Mathematical Population Studies* 9: 231-263.

Rogers, A., and Castro, L. J. (1981). Model migration schedules. Research report 81-30. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rychtaříková, J. (2006). Healthy life expectancy in the current Czech population, Demographie 48: 166-178. http://www.czso.org/csu/redakce.nsf/i/demographie.

Schoen R. and Jonsson S. (2003). Estimating multistate transition rates from population distributions. Rostock: Max Planck Institute for Demographic Research. *Demographic research* 9(1):1-24. http://www.demographic-research.org/Volumes/Vol9/1/.

Simpson, L. (2004), Integrating estimates and targets within a population projection, ESRC Research Methods Programme, Working Paper no.18.

Skorglund T. and Stokka A., (1988), Problems of linking single-region and multiregional economic models, CBS, Discussion Paper, No. 37.

Skorglund T. and Stokka A., (1988), Problems of linking single-region and multiregional economic models, CBS, Discussion Paper, No. 37.

Stillwell, J. (2005) *Inter-regional Migration Modelling: A Review and Assessment*. Paper prepared for the 45th Congress of the European Regional Science Association, Vrije Universiteit Amsterdam, The Netherlands, 23-27 August.

Sykes, Z. M. (1969). Some stochastic versions of the matrix model for population dynamics. *Journal of the American Statistical Association*, 44, 111 – 130.

Tabeau, E. (2001). A review of demographic forecasting models for mortality. In: Tabeau, E. Van Den Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*. Dordrecht: Kluwer Academic Publishers: 1 -32.

Tabeau, E., Ekamper, P., Huisman, C., and Bosch, A. (2001). Predicting mortality from period, cohort or cause-specific trends: A study of four European countries. In: Tabeau, E. Van Den Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective.* Dordrecht: Kluwer Academic Publishers: 159-187.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society*, Series A, Statistics in Society, 162, 5 -43.

Van Den Berg Jeths, A., Hoogenveen, R., De Hollander, G., and Tabeau, E. (2001). A review of epidemiological approaches to forecasting mortality and morbidity. In: Tabeau, E. Van Den

Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective.* Dordrecht: Kluwer Academic Publishers: 33-56.

Willekens, F. (1983). Multiregional population analysis for urban and regional planning. In: Batty M. and Hutchinson B. (eds.). *System analysis in urban policy making and planning*. New York: Plenum Press: 227-241.

Willekens, F. J. (2001). Gompertz in context: The Gompertz and related distributions. In: Tabeau, E. Van Den Berg Jeths, A., and Heathcote, C. (eds.). *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective.* Dordrecht: Kluwer Academic Publishers: 105-126.

Willekens, F., Por A. and Raquillet R. (1981).Entropy, multiproportional, and quadratic techniques for inferring patterns of migration from aggregate data. In: Rogers A. (1981). *Advances in multiregional demography*. Laxenburg: International Institute for Applied Systems Analysis (research reports RR-81-6): 83-124.

Willekens, F.J., De Beer, J., and Van der Gaag, N. (2005). *MicMac: from demographic to biographic forecasting*. Paper presented at the Joint Eurostat-UNECE Work Session on Demographic projections, Vienna. 21-23 September 2005.

Wilmoth, J.R. (1990) Variation in vital rates by age, period, and cohorts. In: Clogg C.C. *Sociological methodology*, Oxford: Basil Blackwell 20: 295-335.

Wilmoth, J. R. (1996). Mortality projections for Japan: A comparison of four methods. In G. Caselli, & A. Lopez (Eds.), *Health and mortality among elderly populations* (pp. 266– 287). New York: Oxford University Press.

Wilmoth, J. R. (2001). Age–period–cohort models in demography. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demographie: Analyse et synthese. La Dynamique des Populations*, vol. I (pp. 379– 397). Paris: INED.

Wilmoth, J. R. (2005). Some methodological issues in mortality projection, based on an analysis of the U.S. social security system. *Genus*, LXI, 179– 212.

Wilson T. and Bell M. (2004) Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research* 21(2):127-160.

Wilson, T. (2010). Model migration schedules incorporating student migration peaks. Rostock: Max Planck Institute for Demographic Research. *Demographic research* 23(8):191-222. http://www.demographic-research.org/Volumes/Vol23/8/.

Wissen L., van der Gaag N., Phil Rees and Stillwell J.(2002) In search of a modelling strategy for projecting internal migration in European countries. Demographic versus economic-geographical approaches. Report on behalf of Eurostat: Study of Past and Future Interregional Migration Trends and Patterns within European Union Countries.

Zaba, B. (1979). The four-parameter logit life table system. *Population Studies*, 33, 79–100.

Zaba, B. (1985). A parameterized procedure for projecting population. *International Population Conference*, Florence, vol. 4. (pp. 137–150). Lie`ge: International Union for the Scientific Study of Population.

Zaba, B. (1987). The indirect estimation of migration: A critical review. *International Migration Review*, 21, 1395– 1444.

Zeng, Y., Zhenglian, W., Zhongdong, M., & Chunjun, C. (2000). A simple method for projecting or estimating a and h: An extension of the Brass relational Gompertz fertility model. *Population Research and Policy Review*, 19, 525– 549.