WP. 4
26 September 2007

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(Bucharest, 10-12 October 2007)

Agenda item 6: Population projections

REPLICATING THE OFFICIAL POPULATION PROJECTION FOR SWEDEN USING A TIME SERIES APPROACH

Invited Paper

Submitted by Statistics Sweden[1]

I. INTRODUCTION

1. The official population projection for Sweden is produced at Statistics Sweden every third year and the result is presented in a publication; see for example Statistics Sweden, 2006. The official population projection for Sweden follows the traditional cohort projection method and is thus an example of what we in this paper will call a deterministic population projection. It is however widely recognized that there is a need for statistically assessing the precision in a population projection and several approaches have been suggested for estimating the future error boundaries for a population projection. In the past, as well as in the present, in Sweden, as well as in many other counties, the method in use has been the scenario approach in which the demographic components are assumed to enfold according to high or low constant levels.

2. In response to current demographic discussions concerned with estimating the precision of population projections a Statistics Sweden Development Project was undertaken. The aim of the development project was to study the relatively new branch of this area which has brought theory and application of stochastic processes into main focus in demographic contexts, (see e.g. Alho, 1990, Goldstein, 2004 and Keilman, Pham and Hetland, 2002) and to apply these techniques to Swedish population data. The development project was documented in the Statistic Sweden publication *Stochastic Population Projections for Sweden*, Hartmann and Strandell, 2006.

3. During the course of the development project a number of stochastic population projections for Sweden were produced, based on different assumptions and techniques, albeit not in direct connection with the official population projection. It is the belief of the author of this

[1] Prepared by Gustaf Strandell.

paper that stochastic population projections are seldom produced in direct connection with official population projections produced around the world. The general understanding at Statistics Sweden has been that the deterministic and the stochastic techniques constitute altogether different approaches which can not be coordinated and which necessarily gives incoherent results.

4. The main conclusion of this paper is that the results of the Swedish official population projection from 2006 could have been attained using a stochastic time series approach. Specifically we illustrate how the 2006 Swedish official population projection emerges as a mean of a large number of extrapolations of past demographic time series behaviour. In the terminology of this paper we thus replicate the official population projection for Sweden with a stochastic population projection. In order to maintain simplicity and transparency the stochastic population projection has been produced using standard one-dimensional time series models, which are easily explained and applied.

5. Our time series approach also leads to assessment of error boundaries for the official projection, error boundaries, which we believe are more precise, then those, which are attained using the classical scenario approach. These error boundaries also demonstrate, for example, the fact that it is much more difficult to forecast the young population than the older population of Sweden.

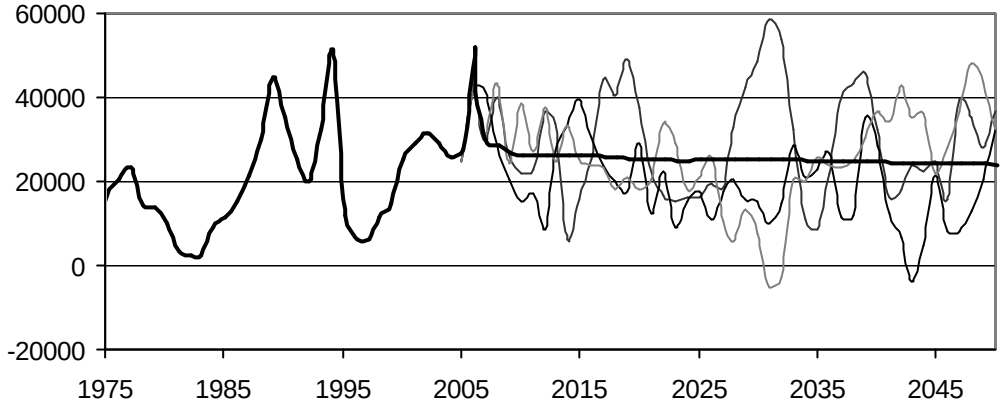
II. STOCHASTIC AND DETERMINISTIC POPULATION PROJECTIONS AT STATISTICS SWEDEN

6. The purpose of this section is to give a description of the difference between deterministic and stochastic population projections, as these concepts are understood at Statistics Sweden. The description given here is deliberately non technical and elementary.

7. The progress over time of a population obeys some rather strong deterministic rules. All individuals in the population will die at some point, and those who doesn't die gets one year older for each year, only women can get can give birth to new individuals and only at certain ages, etc. In all population projections that are made at Statistics Sweden these deterministic facts are incorporated at the simulation stage. The simulation process proceeds one year at the time and from the population in the previous step the current number of births, deaths and in- and out migrants are calculated and everyone who survives gets one year older. The simulated demographic components for the current year and the population from the previous year are then used to calculate the population for the current year.

8. For the simulation process to be feasible we need projections of the future mortality, the future fertility and the level of the future migration from which we can determine rates, risks and probabilities. In producing these projections the stochastic and the deterministic techniques follow different routes. Diagram 2.1 contains observed net migration in Sweden 1975-2005 and projections of the net migration as they can look in a deterministic and a stochastic population projection.

Diagram 2.1. Stochastic and deterministic projections of net migration in Sweden



9. The net migration in Sweden during 1975-2005 is characterized by heavy fluctuations with high tops and deep valleys, and it is not hard to believe that the future net migration will also fluctuate. However, we cannot say with certainty exactly when the future tops and valleys will occur or how high or deep they will be. The probability of guessing the right future net migration also gets smaller as we try to look further into the future.

10. The thick black curve in the diagram stretching from 2006 to 2050 shows how the net migration has been projected in a deterministic projection at Statistics Sweden (in the official projection made at Statistics Sweden the in migration and the out migration are projected separately, the net migration shown here has been calculated from those projections). Since the producers of this projection do not know (at least with certainty) the placement or the shape of the future tops and valleys they have put the net migration at a reliable even mean level. In the official population projection for Sweden even mean levels are projected for the mortality and the fertility in a similar way.

11. The fluctuating thinner curves which stretches from 2006 to 2050 in Diagram 2.1 shows some projections of the net migration from a stochastic projection. In producing the stochastic projection we want the projected time series to have the same statistic characteristics as the time series we wish to extrapolate into the future. We want the stochastic projections to reflect both the horizontal trend and the fluctuations around the trend which are present in the observed time series in the diagram. However, again, we can not know neither the shape nor the placement of the future tops or valleys. Therefore, instead of projecting just one possible future for the net migration we project a big number of possible futures which all follow the same trend and contain the same amount of variation, but in which the width, the height and the placement of the future tops and valleys are determined at random.

12. In a stochastic population projection we use tools from our time series toolbox to produce a big number of triples of time series according to the philosophy described above. Each triple consists of one projection of the future net migration, one projection of the future fertility and one projection of the future mortality. Each triple is then put into the population simulation machiner, which manufactures a possible future development of the population. The raw product

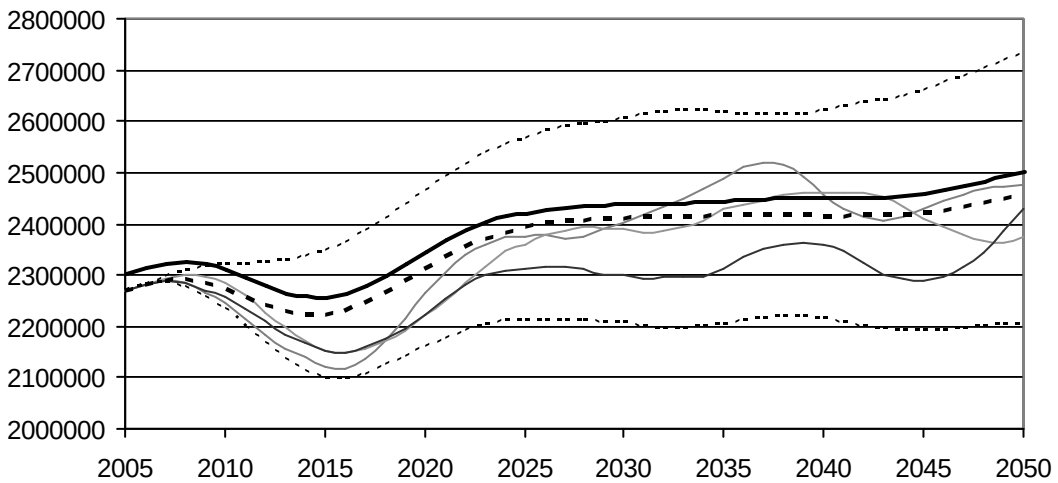
of a stochastic population projection consists therefore of an entire family of population projections.

13. To conclude, we note that the difference between the deterministic and the stochastic approach can be expressed in terms of when in the production process we turn to considering mean values. In a deterministic projection we start by projecting future mean values for the underlying demographic components. These mean values are then used in the simulation of the future development of the population. In a stochastic projection, in order to keep the simulation of the population's development as close to reality as possible, we want the projections of the future development of the demographic components to have the same characteristics as there observed counterparts. This results in a large number of population projections, from which we can calculate a mean projection at the end of the production process.

14. If the purpose of our projection activities only is to project a reliable future population progression in the mean, then this can be achieved just as well by any of the two methods described above. However, the stochastic approach results in a much larger set of data then the deterministic approach, and this extra data contains extra information which can be put to good use.

15. Diagram 2.2 shows the projected number of people aged 0-20 in Sweden for the years 2006-2050 and the corresponding observed number for 2005. The thick black line in the middle actually consists of two lines, one broken line which shows the result of the Swedish official population projection 2006 and one non broken which shows the mean result of the stochastic population projection to be presented in the next section. Needles to say, the two lines almost coincide. The area between the two broken lines constitutes 95 % prediction intervals for the stochastic projection. Since the results of the two projections coincides so closely we claim that the prediction intervals from the stochastic projection can be seen as 95 % prediction intervals for the official Swedish population projection.

Diagram 2.2. Stochastic and deterministic projections of population of Sweden age 0-20



16. The stochastic projection on display in Diagram 2.2 has been produced using time series models, which has been fitted to observed data for the demographic components for the last 30

years, i.e. 1976-2005. The prediction intervals could hence be understood as follows: Suppose that the trends which has been observed in data for the demographic components for the last 30 years will not radically change during the projection period, and suppose that the future random deviations from the trends in form of variation will be of the same type and the same size as those which has been observed. For each year during the projection period the probability is then 95 % that the indicated prediction interval will contain the real observed population that year. The prediction intervals are thus reliable as long as the statistic properties of the demographic components will not change drastically during the prediction period.

III. METHODS, ASSUMPTIONS AND RESULTS

17. In this section we present the stochastic population projection which according to our claim replicates the Swedish 2006 official population projection. We first discuss the methods we have used to project the fertility, the net migration and the mortality and we then present the results of the projection and compare these to those in the official projection.

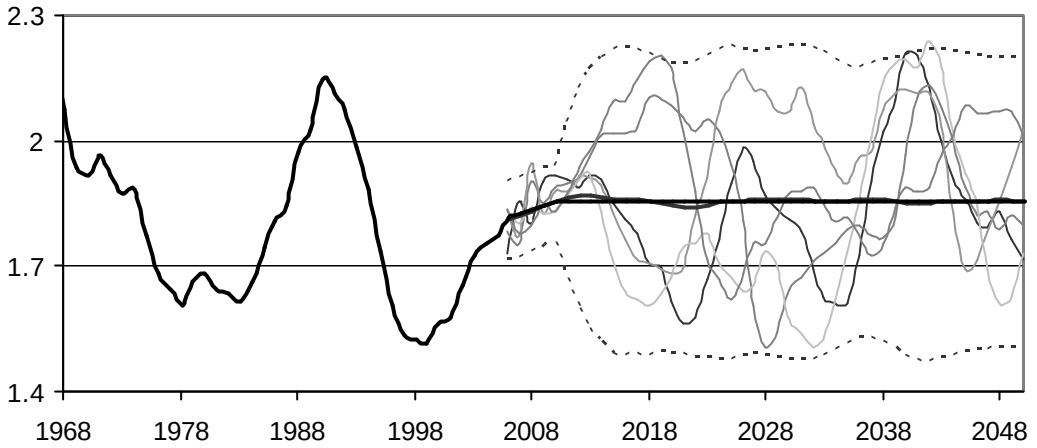
18. The overall level of the fertility rates in our stochastic projection is controlled by one single parameter, the total fertility rate (TFR). The fertility scheme, which is used together with the TFR to calculate age specific fertility rates for each year during the projection period is kept constant during the projection period (we use the same constant fertility scheme as in the official projection).

19. To project the future TFR we use an AR(2) model:

$$TFR_t = a_1 TFR_{t-1} + a_2 TFR_{t-2} + m(1 - a_1 - a_2) + e_t .$$

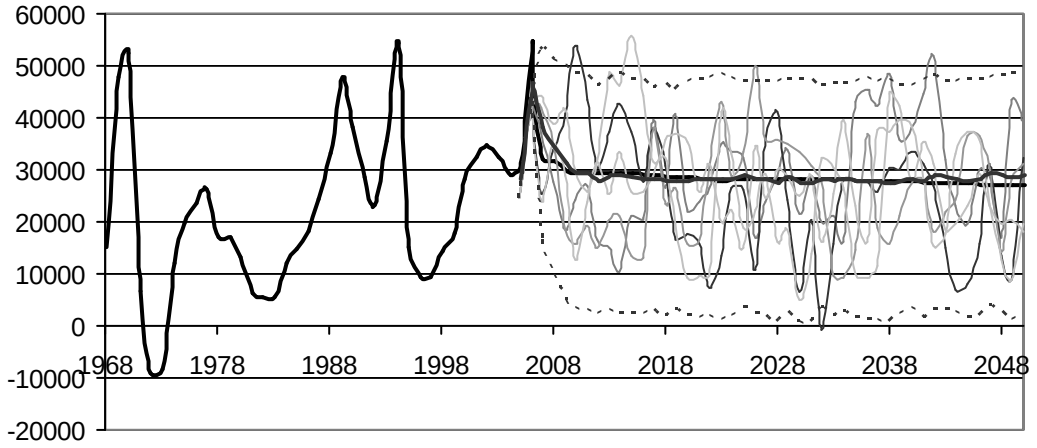
20. We first fit the model to the observed TFR for the years 1976-2005. We then make a small adjustment of the m parameter in order for the projection of the TFR using the AR(2) model to give results consistent with those in the official population projection. The fitted and adjusted model is then used to produce 500 projections of the TFR, where we simulate the error term e_t with independent random (Gaussian) noise with the residual variance from the fit of the model. The result is shown in Diagram 3.1. The two almost coinciding lines in the middle of the diagram shows the projected TFR in the official projection and the mean of the 500 projections made in our stochastic projection. The area between the broken lines constitute 95 % projection interval for the TFR.

Diagram 3.1. Stochastic and deterministic projections of TFR in Sweden



21. In our stochastic projection we do not treat in migration and out migration separately (in the official projection they are however treated separately). Instead, we work directly with net migration. The net migration is treated as an exogenous variable in the sense that the projected size of the future population (or any other "outer" variable) does not influence the projection of the net migration. The future net migration in the stochastic projection is projected in almost complete analogy with how we project the TFR. To project the net migration we use an AR(2) model which we fit to observed data for the years 1976-2005. The m parameter of the AR(2) model is then fine tuned in order for the model to produce projections which in the mean coincide with the corresponding net migration in the official projection. The result is shown in Diagram 3.2. The net migration is finally divided among ages and sexes according to a migration scheme which is kept constant during the projection period and which coincides with the corresponding migration scheme in the official projection.

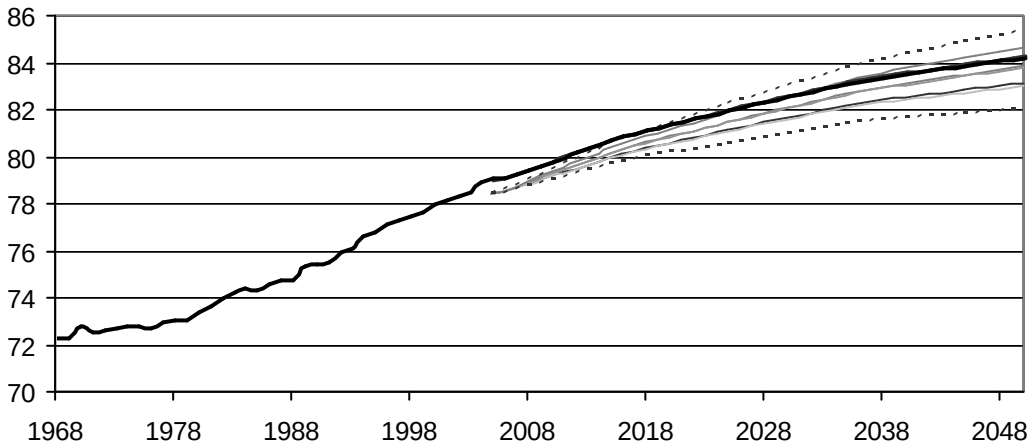
Diagram 3.2. Stochastic and deterministic projections of net migration in Sweden



22. In our stochastic projection we use a Brass model to model the mortality (see e.g. Brass, 1971 and Brass, 1974). The future age and sex specific death rates are controlled in the Brass model by two parameters, alpha and beta (actually one alpha and one beta per sex). When the Brass model is used in a population projection the usual course of action consists of several steps. First, values of alpha and beta for past years are calculated from observed mortality data. The resulting time series for alpha and beta are then projected into the future using some time series models which has been fitted to the calculated alphas and betas. Finally, the future values of alpha and beta are transformed back to future mortality data.

23. In our stochastic projection we have chosen a different path along which we project future mortality data. Although in technical terms we do use the Brass model, the main idea behind our approach is illustrated here by giving reference directly to the life expectancy at birth. Both the TFR and the net migration have been treated above as stochastic processes which follow a stationary horizontal trend where the difference between the years seems to be mostly due to random jumps up or downs. The life expectancy at birth is, as can be seen from Diagram 3.3, a process of a different breed. The difficulty in projecting future values of this series is not primarily about the placement and shape of the future tops and valleys. Instead, the problem seems to be to find the future direction of the curve describing the life expectancy. The life expectancy in Sweden will with high probability continue upwards but the question is at what rate.

Diagram 3.3. Stochastic and deterministic projections of life expectancy at birth for males in Sweden



24. In our stochastic projection we have based our projection of the life expectancy on the projection of the future life expectancy in the official projection. Let $Life(off, y),\ y = 2006,...,2050$ denote the projection of the life expectancy in the official population projection. Our stochastic projection of the life expectancy consists then of 500 projections of the form

$$Life(k, y) = Life(off, y) + d(k)(y - 2005) + e(k, y), \qquad k = 1,...,500 \quad y = 2006,...,2050,$$

where the 500 numbers *d(k)* and the 22500 numbers *e(k,y)* are chosen randomly and independently. In other words, we produce 500 new life expectancies by adding linear functions with random coefficients of direction, *d(k)*, and by adding a small noise term, *e(k,y)*, to the projected life expectancy in the official population projection. An evaluation study of deterministic predictions of the life expectancy at birth undertaken at Statistics Sweden has shown that a deterministic projection ten years into the future has a mean error of one year. The spread of our random coefficients of direction, *d(k)*, has been decided according to this.

25. Using our 500 projections of the TFR, our 500 projections of the net migration and our 500 projections of the life expectancy at birth (actually of alpha and beta) and the population simulation machinery mentioned in Section 2 we now iterate 500 population projections, i.e. 500 projections of the number of children born, the number of people deceased and the population divided after age and sex. Diagrams 3.4-3.8 display the outcome. The thick lines in the middle of the diagrams always consist of two lines, one showing the mean value of the 500 iterations in our stochastic projection and one showing the corresponding prediction from the Swedish official 2006 population projection. We claim that these lines coincide closely enough for us to say that we have replicated the official projection with a stochastic projection. The thinner non broken lines in the diagrams shows some individual projections from our stochastic projection and the area between the broken lines constitute 95 % projection intervals.

Diagram 3.4. Stochastic and deterministic projections of the number of children born in Sweden

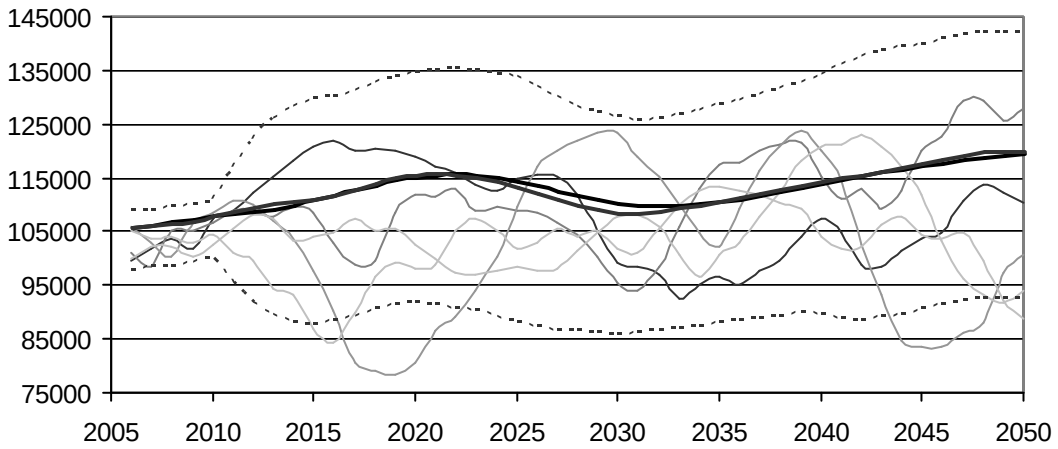


Diagram 3.5. Stochastic and deterministic projections of the number of people deceased in Sweden

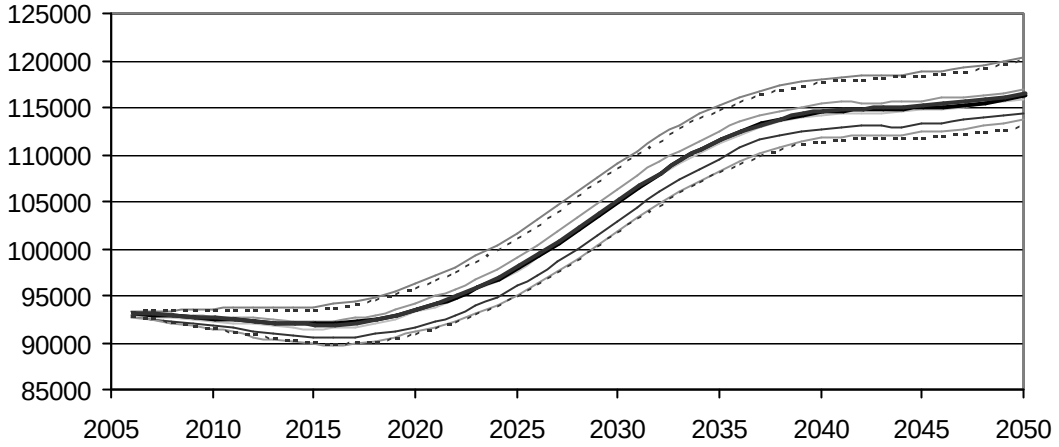


Diagram 3.6. Stochastic and deterministic projections of the total population in Sweden

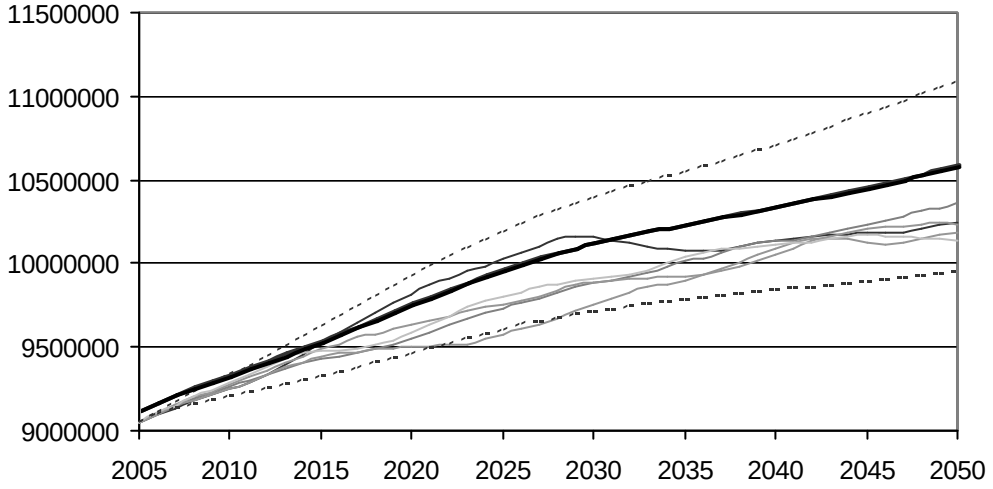


Diagram 3.7. Stochastic and deterministic projections of population males by age in 2050.

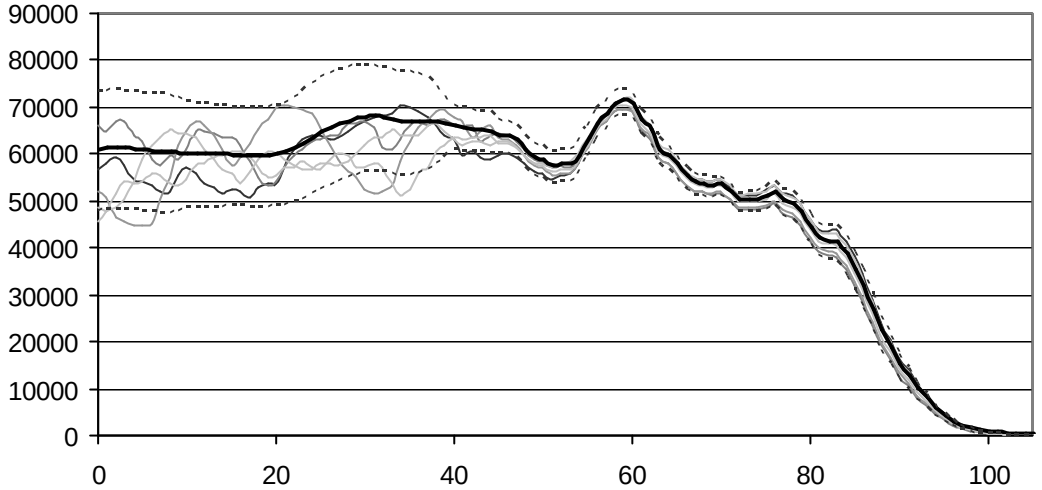
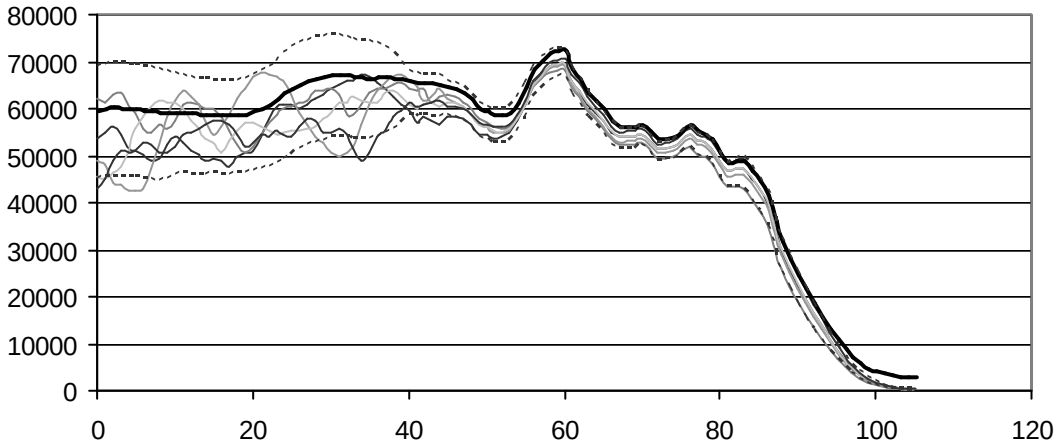


Diagram 3.8. Stochastic and deterministic projections of the population females by age in 2050



26. From Diagram 3.7 and Diagram 3.8 we conclude that it is much more difficult to say anything certain about the number of young people in Sweden in the year 2050 than it is to guess the number of old people that year. This means that care for older people during the next 45 years or more can be planned today (as long as the only variable under concern is the number of people) whereas care and education for younger people can only be planned a few years into the future.

IV. REFERENCES

Alho, J (1990). Stochastic methods in population forecasting. *International Journal of Forecasting,* 6,4, 521-530.

Brass, W (1971). *On the Scale of Mortality* in Biological Aspects of Demography, ed. By William Brass, pp. 69-110.

Brass, W (1974). Mortality Models and their Uses in Demography. *Transactions of the Faculty of Actuaries,* Vol. 33, No. 2, pp. 132-142.

Goldstein, J (2004). Simpler Probabilistic Population Forecasts: Making Scenarios Work. *International Statistical Review,* 72, 93-106.

Hartmann, M & Strandell, G (2006). *Stochastic Population Projections for Sweden,* Statistics Sweden, Research and Development report.

Keilman, N., Pham, D.Q. & Hetland, A (2002). Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research*, 6, Article 15. Max Planck Institute for Demographic Research.

Statistics Sweden (2006), *Sveriges framtida befolkning 2006-2050,* Statistics Sweden.

*****