

The following presentation describes the outline for a simple R (www.r-project.org) script that I wrote to forecast population

The script was shared with a statistical computing class, just as an application of R

Though I am interested in this type of population forecast, I am not an expert on the subject

To understand probabilistic population forecasts, the following text is a better place to start:

Lutz, Vaupel and Ahlburg [editors] (1998), *Frontiers of Population Forecasting*, Population Council

Please send me a message if you have any knowledge to share on the subject, or if you find this presentation useful

A Probabilistic Population Forecasting Technique Applied to Alaska: 2005-2030

Eddie Hunsinger
Demographer, Alaska Dept. of Labor

eddieh@demog.berkeley.edu
www.demog.berkeley.edu/~eddieh

Presented to Demography 212 (Computing for Demographers)
UC Berkeley, Spring 2007

The technical approach described in these slides has not yet been adopted by the State
of Alaska for its official population forecast

Please treat the work described here as part of a working paper

How did this forecast happen?

Alaska State Demographer requested a probabilistic technique to forecast population

To develop what I think is a simple but reasonable technique, I tried to follow:

1. 110, 210 and 213

(Berkeley Demography Courses led by Wachter (110,210) and Mason (213))

2. The work of others on the subject

Lutz, Vaupel and Ahlburg [editors] (1998), *Frontiers of Population Forecasting*, Population Council

Miller (2002), "California's Uncertain Population Future" Online.

<http://www.demog.berkeley.edu/~tmiller/papers/p2002.california.forecast.pdf>

Smith, Tayman and Swanson (2001), *State and Local Population Projections*, Springer

Pflaumer (1988), "Confidence intervals for population projections based on Monte Carlo methods" *International Journal of Forecasting*, 4: 135-142

3. Independent reasoning

What I think probabilistic population forecasts offer

Probability of different future population levels,
given:...

specified distributions and autocorrelations for
future fertility, survival and migration index
levels,...

and assumptions regarding the age profiles of
fertility, survival and migration.

How simple is this one?

Requirements:

1. A distribution of index rates (such as the TFR)
2. Fixed age profiles for index multipliers
3. A simple simulation technique

Optional: autocorrelation level for the index rates
(for five year steps I allow zero autocorrelation)

How simple is this one?, cont.

Project the population with a set of random index values (random based on selected variance)

Repeat many (maybe 1000) times

This yields an inferred probability distribution for future populations

This is often called a “Monte Carlo” approach

Variance selection

This specific presentation is focused on just the organization of the population forecast

I selected the variance of the index distributions (all Gaussian) presented here subjectively, and they are for presentation purposes only

I hope to develop and describe my variance selection method as the work develops further

References on the subject include:

Lutz, Vaupel and Ahlburg [editors] (1998), *Frontiers of Population Forecasting*

Miller (2003), "California's Uncertain Population Future"

Pflaumer (1988), "Confidence intervals for population projections based on Monte Carlo methods"

Index rates that I used

Fertility: TFR with a fixed age profile

Migration: total net “migration” ratio (similar to a TFR) with a fixed age profile

Survival: did (almost) nothing-- used Brass to deterministically increase life expectancy in line with Social Security projections for the US

How I estimate and vary the F_x 's

Lee (1974), "Forecasting births in post-transitional population: Stochastic renewal with serially correlated fertility,"
Journal of the American Statistical Association, 69: 607-617

F_x = Fertility rate at age x

$TFR = \text{sum}(F_x)$

I create a proportional age profile

$\text{Prop}F_x = F_x / TFR$

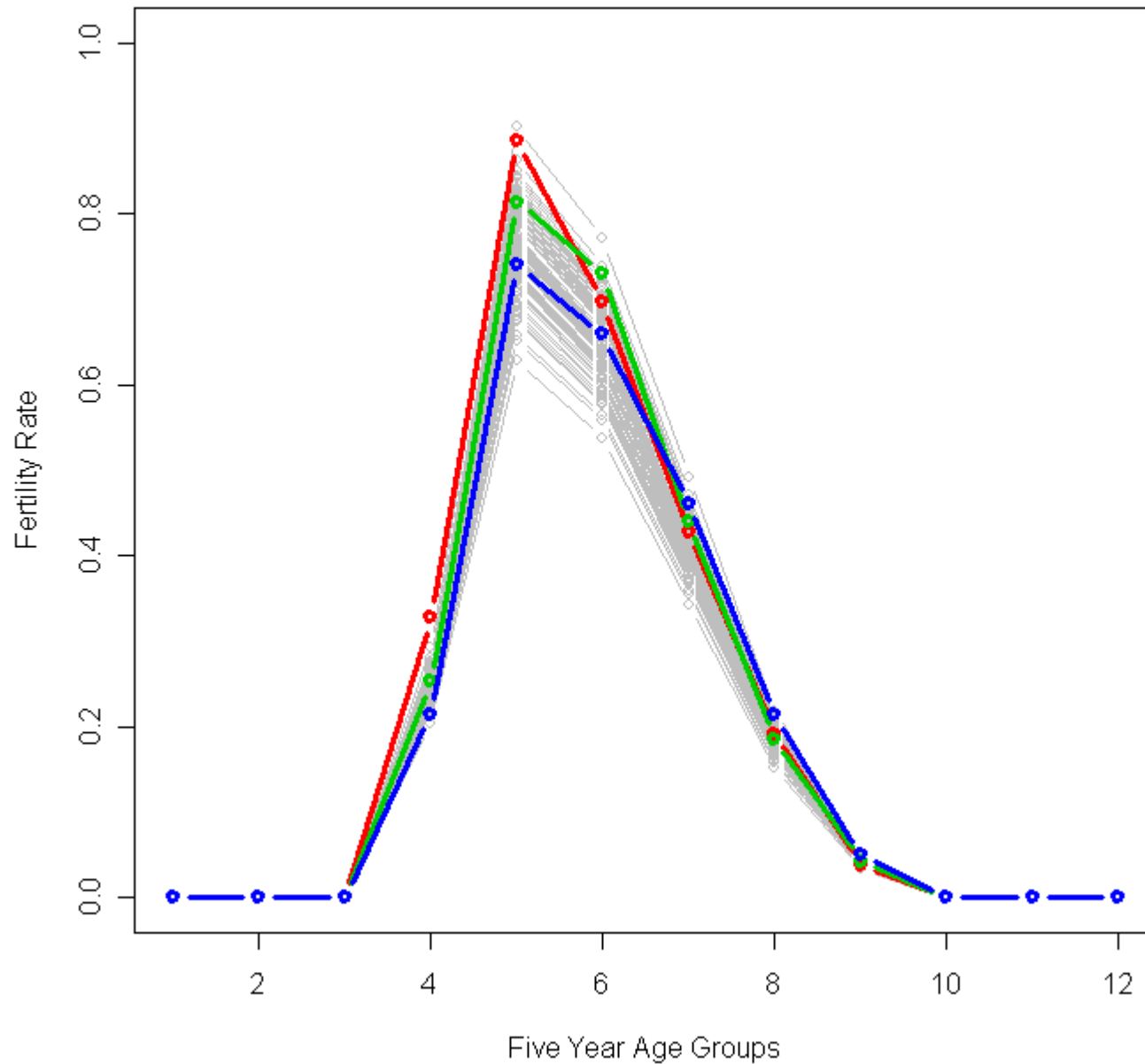
Use average of standardized Alaska ASFRS 90-05, from AK DHHS

I now create simply modeled F_x 's

$F_x = \text{Prop}F_x * TFR$

Fertility Rates: 1990-94, 1995-99, 2000-04, Model

Model variance selected for presentation purposes only



Too simple for the instability of fertility?

Clearly this is a very simple way to manage age specific fertility

There are more complex approaches that account for changes in the age profile of fertility

Lee (1993) "Modeling and Forecasting the Time Series of US Fertility: Age Patterns, Range, and Ultimate Level," *International Journal of Forecasting* 9:187-202

Net migration?

I've created several versions of the script that each manage migration differently

The version I'm presenting is the only one that attempts to manage net migration rather than gross migration

I think this version is the most imaginative, that it performs as well as any of the others, and that it is the easiest to replicate...and it's my favorite

Net migration?

My net “migration” ratios don't track real rates of migration

Real rates of migration can only be calculated and implemented through multi-regional data and models

The ratios I use take us outside of real demographic processes

They *may* be better understood as net “correction” ratio estimates

Other migration options?

Age specific in- and out-migration rates with US population as the denominator for in- and Alaska for out-

The same, but with Alaska as the denominator for in-

Numbers or crude rates of in- and out- migrants, shared out to the age groups with a fixed profile

How I estimate the age specific net migration ratios (N_x)

$K_{x+n}[1]$ = Population age x plus n at time 1

$$K_{x+n}[1temp] = K_x[0] + Births_x - Deaths_x$$

$$Residual_{x+n} = K_{x+n}[1actual] - K_{x+n}[1temp]$$

$$N_{x+n} = Residual_{x+n} / K_{x+n}[1temp]$$

$$TNR = \text{sum}(N_x)$$

How I vary the age specific net migration ratios

I've not seen migration managed this way before. If this approach has been used before (for me to cite), or if it is just ridiculous, please send me a message indicating so (eddieh@demog.berkeley.edu)

$$G_x = \text{Migrants}_x[1] / K_{x-n}[0]$$

Using real age specific migration flow data from Census "Place of Residence" 95-00 for Alaska

I create a fixed age gross migration profile (G_x)

$$\text{Prop}G_x = G_x / \text{sum}(G_x)$$

Could also use a model age profile for migration here

I then create a "proportional" net migration profile

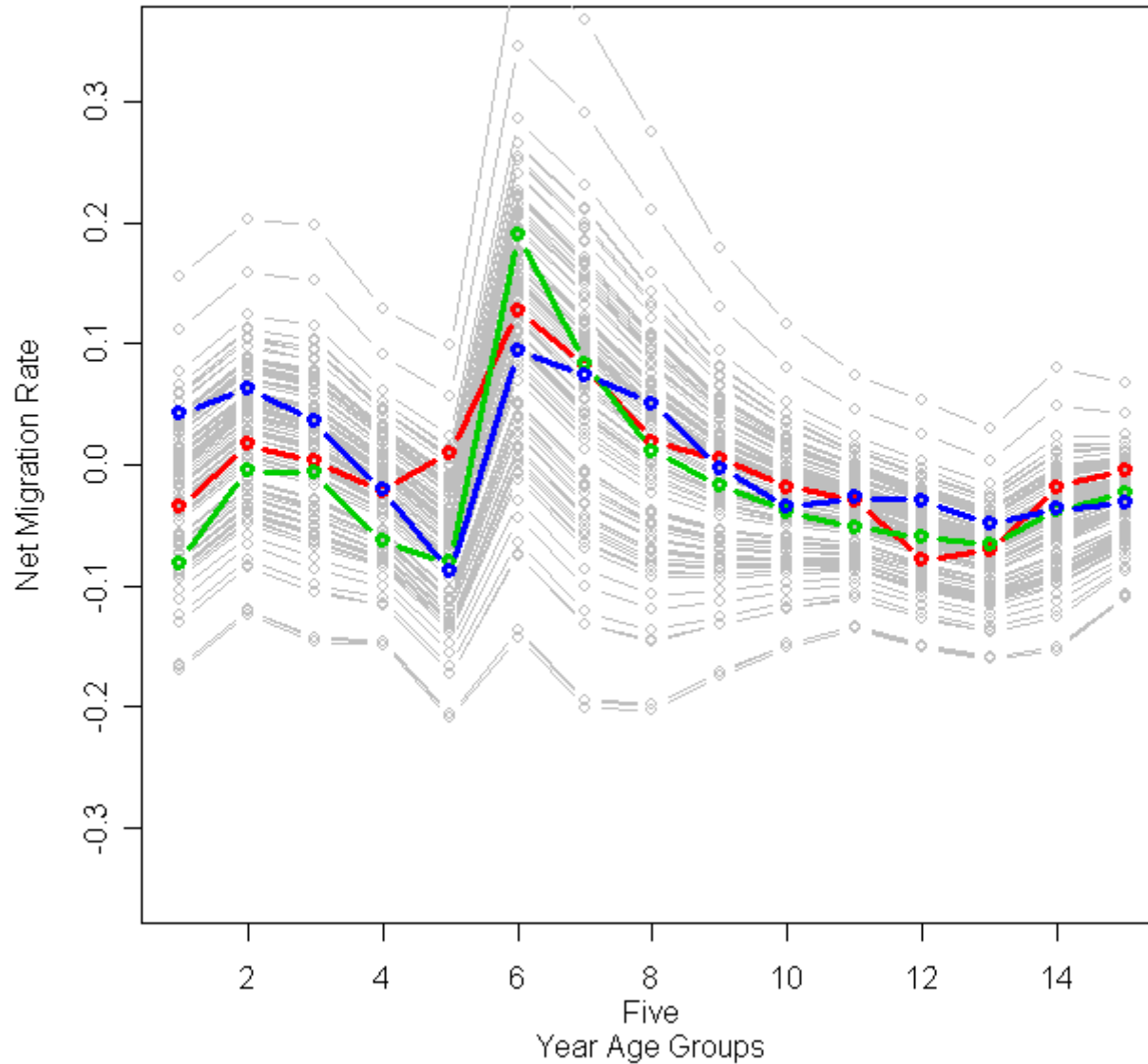
$$\text{Zero}N_x = N_x - (\text{TNR} * \text{Prop}G_x)$$

I vary the profile by

$$N_x = \text{Zero}N_x + (\text{TNR} * \text{Prop}G_x)$$

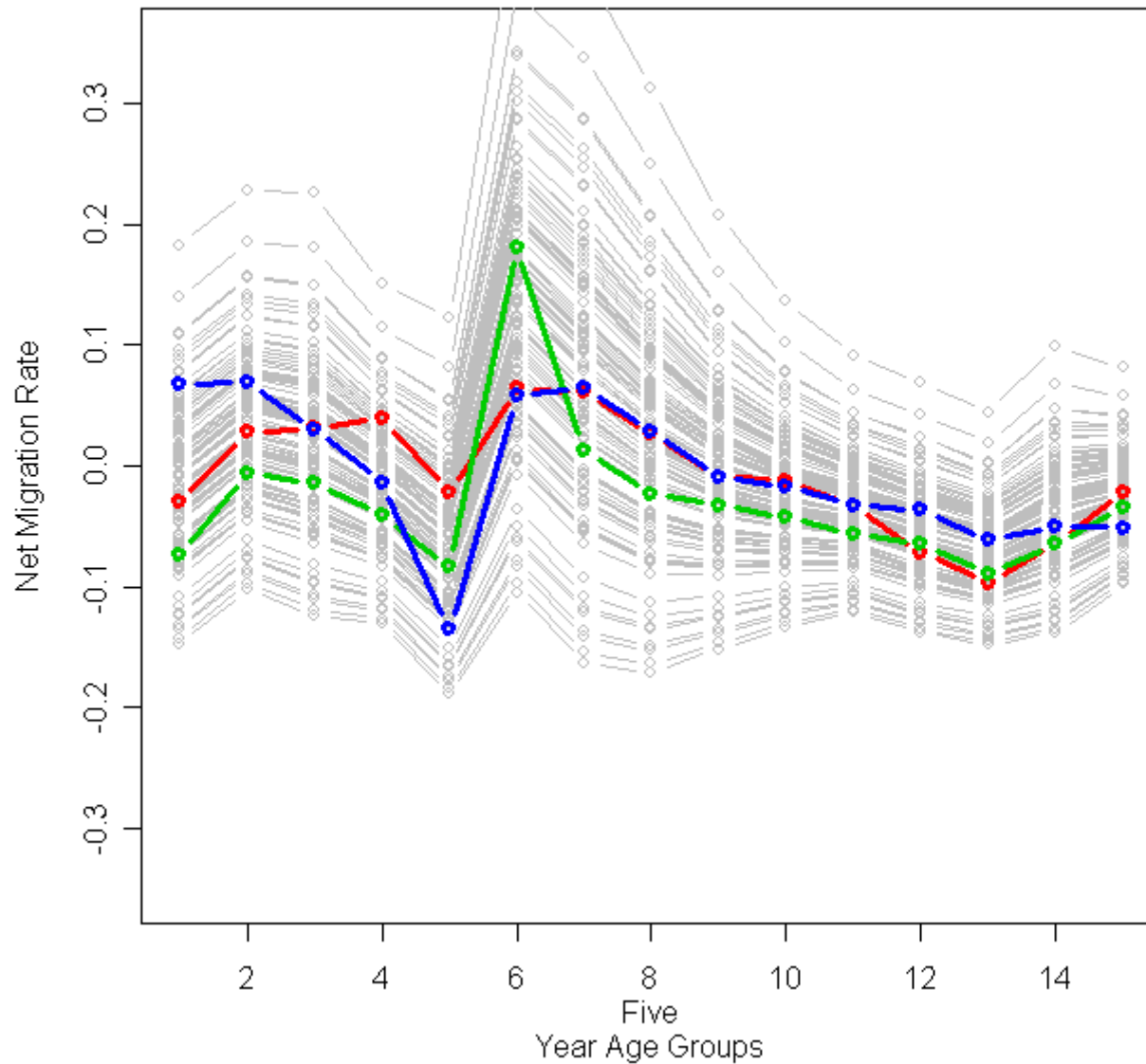
Female Net Migration Ratios: 1995, 2000, 2005, Model

Model here uses average of 95-05 ZeroNx profiles
Model variance selected for presentation purposes only



Male Net Migration Ratios: 1995, 2000, 2005, Model

Model here uses average of 95-05 ZeroNx profiles
Model variance selected for presentation purposes only



How well should “ N_x ” work with regard to in-migrants?

I think it should be that the closer the shape of the population age profiles for the sending and receiving pools of potential migrants, the better the estimate of in-migrants by “ N_x ”

I think one problem is that the variances for in- and out-migrants are surely distinct, so that, when certain ages are dominated by in- or out-migration, a single “net” migration variance estimate is flawed

How I managed survival

Solved for the Brass alpha and beta's on SSA's 1980 to 2025 (projected) l_x 's, with interpolated 2005 l_x 's as the Brass standard

Applied the same coefficients to Alaska 2005 l_x 's to estimate a time series of Alaska l_x 's

How I managed the simulation technique

Create an array of random numbers (0,sd): length [i,] equal to the number of iterations (1000), width [,i] equal to the number of projection steps plus one (6)

These will be the Error term for the iterations

How I managed the simulation technique, cont.

BaseTFR = TFR at time 0

Fert = array of length [i,] 1000 and width [,i] 6

Fert[:,1] = BaseTFR

AC = Autocorrelation

Flat Constant = BaseTFR * (1-AC)

Fert[:,i] = Fert[:,i-1] * AC + Constant + Error

How I managed the simulation technique, cont.

Repeat for Net Migration

How I managed the simulation technique, cont.

Create an array of the same length and width, with a third dimension for number of age groups (21), and fill it with the age specific fertility rates

Create another array of the same three dimension sizes, and fill it with the N_x vectors

How I managed the simulation technique, cont.

Create an array of Leslie transition matrices:
length [i,] equal to the number of iterations, width
[,i] equal to the number of projection steps

Fill in the sub-diagonals using survival data for the
respective projection step

Fill in the top row using survival and fertility data
for the respective projection step

How I managed the simulation technique, cont.

Do the matrix multiplication to project the population. Add the $K_x[1temp]$ times N_x to the projected population:

$K_x[0]$ = Population age x at time 0

A = Leslie matrix

$K_x[1temp] = K_x[0] * A$

$K_x[1final] = K_x[1temp] + K_x[1temp] * N_x$

Extra: A way to manage the male top row on the Leslie matrix

Keyfitz (1977), *Applied Mathematical Demography*, John Wiley

$$A_{[1,j]} = (L0[\text{male}] / (2 * l_0)) * (F_x + F_{x+n} * (S_x[\text{female}])) * \text{fmab}$$

$$\text{fmab} = 1 - .4886$$

$$K_0[1\text{males}] = K_j[0\text{females}] * A$$