Peptutorial/6.6.2003

Tutorial for PEP - Program for Stochastic Forecasts of Population

Juha Alho and Mikko Mustonen

Purpose. This is a purely "how to run the program in practice" tutorial. A more detailed description of *PEP* is available at: http://joyx.joensuu.fi/~ek/pep/. The specification of the various files and parameters needed by *PEP* is a rather more demanding exercise. Here it is assumed that someone has done the work.

1. Preparation

Step 1.1. *PEP* runs under Windows (95 and up - we hope). Create a sub-directory with the name you like, say, **c:\tutorial**. Copy the following files into the sub-directory: **pep.exe, comb1.exe, bds52f.dll, cws3230.dll, owl52f.dll**. The four files take up about 1.48 MB. **pep.exe** makes stochastic forecasts. It needs the dll-files (somewhere in your system) to run. **comb1.exe** is used for aggregating age-specific data. It is optional.

Step 1.2. Copy the file **params.dat** into c:\tutorial. This is not necessary, but we recommend it, as it facilitates the running of the program - provided that you use the file names given below!

2. Setting up the Point Forecast Files

Step 2.1. Copy the file **jpop.dat** into c:\tutorial. This is a *text file* like the other input files to be given below. It contains the *jump-off population* by age and sex. File 1 below provides the example we use in the tutorial. *Highest age* is 7.

Step 2.2. Copy the file **asmr.dat** into c:\tutorial. It contains the *point forecast of age-specific mortality* by age and sex for each forecast year. The *number of forecast years* is 5. See File 2.

Step 2.3. Copy the file **asfr.dat** into c:\tutorial. It contains the *point forecast of age-specific fertility* by child bearing age for each forecast year. Here the *lowest age of child bearing* is 2 and the *highest age of child bearing* is 4. See File 3.

Step 2.4. Copy the file **net.dat** into c:\tutorial. It contains the *point forecast of net-migrants* by age and sex for each forecast year. See File 4.

3. Setting up the Files for Scales and Other Uncertainty Parameters

Step 3.1. Copy the file **sc_asmr.dat** into c:\tutorial. It contains the *scales for age-specific mortality* by age and sex for each forecast year. See File 5.

Step 3.2. Copy the file **kp_asmr.dat** into c:\tutorial. It contains the *kappas for age-specific mortality* by age and sex. See File 6.

Step 3.3. Copy the file **sc_asfr.dat** into c:\tutorial. It contains the *scales for age-specific fertility* by age for each forecast year. See File 7.

Step 3.4. Copy the file **kp_asmr.dat** into c:\tutorial. It contains the *kappas for age-specific fertility* by age. See File 8.

Step 3.5. Copy the file **sc_net.dat** into c:\tutorial. It contains the *scales for net migration* by sex for each forecast year. See File 9.

Step 3.6. Copy the file **kp_net.dat** into c:\tutorial. It contains the *kappas for net migration* by sex. See File 10.

Step 3.7. Copy the file **gross.dat** into c:\tutorial. It contains the *gross migration* by sex for each forecast year. See File 11.

Step 3.8. Copy the file **age_gr.dat** into c:\tutorial. It contains the *age distribution of gross migration* by sex. See File 12.

4. Running PEP

Step 4.1. Start *PEP* by clicking on the **PEP** icon. The program is menu driven. To exit the program, when you are not sure what to do next, click on **Exit**. The current general **Help** is not much of a help. Sorry!

Step 4.2. Click on **Optional parameters**. Choose the option specification file, by first writing the file name **params.dat** into the space provided (see File 13), and then by clicking on the check box on the left of the text **Read the parameter values and file names from specification file**. If you intend to change some parameter values, then they can be stored into a file you must name. Now, the program is, in principle, ready to run. But check first that everything is as you intended.

Step 4.3. Click on **Input files and parameters**. Click on **Parameters for...** We intend to run *30 simulation rounds*, for example. Here, and in the sequel, edit the values given, as necessary. Click on **Point forecasts**. If there are errors in file names, it will be detected as you click **OK**. Click on **Kappas and scales**.

Step 4.4. Click on **Error terms**. Click on **Mortality**. We use *AR(1) correlation structures* with a number of autocorrelation and crosscorrelation parameters. If you don't know what **Limit year for mortality** means, put some number in that is *larger* than the number of forecast years. Click on **Fertility**. Click on **Migration**.

Step 4.5. Click on **Specification of mortality**. Verify that there is a check mark next to **Projective mortality rates**. Alternative choices are discussed in Appendices.

Step 4.6. Click on **Aggregated files**. This produces simulated results for *age-sex groups you choose*, in addition to the single age by sex data that *PEP* always produces. In our example, aggregation is done for annual files (not sample paths), for both sexes (as opposed to just one), together (as opposed to separately), for three age-groups.

Step 4.7. Click on **Parameters and Options**. Click on **Sex ratio**. Choose 1.05, or other value you like.

Step 4.8. Click on **Optional parameters** (again!). Click on **Testing files**. This will check, before the program starts simulating, that the data files are compatible with the ages and number of forecast years you have given. It is not uncommon that the check routine complains about the presence of characters that cannot be seen on screen. One can avoid this by making the input files with a program that is capable of producing formatted text files.

Step 4.9. Click on **Run**. The files are first checked, if you like. Then, the simulation is carried out. When the program has run, a file **pep.log** is created. See File 14. It describes what parameter values were used, and what the names of the output files were. In this case we get sample paths as files **Px_d1.S1**. For x = 0 we get a point forecast with no stochastics. For x = 1,..., 30 we get 30 simulated sample paths. Files **Yy_d1.S1** for y = 1,..., 5 give the simulated counts for the 5 forecast years. Files **Yx_0.C1**, x = 1,...,5 have the aggregated data (ages 0-1, 2-4, 5-7) for 5 forecast years. See File 15 for an example. Now we are finished!

5. Use of Simulated Data

5.1. The basic output of *PEP* consists of sample paths for the population by single years of age and sex; and of annual files by single years of age and sex. A separate program **comb1** can be run to produce additional aggregated files. The logic is the same as in Step 4.6 above. For example, one can choose all ages and both sexes together to form the total population. Or, we can, e.g., request the sample paths for the three age-groups that were not (chosen in **Step 4.6** to be) output by *PEP* already. The latter results are in files **Px_1.C1**, for x = 0,..., 30. For an example (x = 1), see File 16. A file **combine.log** contains the parameter values used. In the next run of **comb1**, the file names would be **Px_2.C1**, etc.

5.2. The data of File 16 can be used to calculate for example the age dependency ratio, here defined as (the population in ages 0-1 and 5-7) ÷ (population in ages 2-4). This is done by importing the data of the text file into some statistical program. A graph is shown in **Figure 1** was produced with *Minitab*. This allows the study of time paths of population functionals of interest. Or, such paths can be used as inputs in other work, e.g., in macroeconomic models.

5.3. The data of File 15 can be used to assess what the age-dependency ratio will look like in year t = 5. Figure 2 has a histogram of the predictive distribution. Numerical summaries can, of course be produced, once the data have been imported into a statistical program. The mean of the distribution is 1.59, the median is 1.61, the standard deviation is 0.10 etc.

5.4. The files **LE_Yx.dat** contain simulated values of *life expectancies* at birth for males and females.

6. A Note on the Preparation of Input Files

In actual applications we might have, say, 100 ages and 50 forecast years. There is no way the input files could be produced by entering the text by hand, character by character. A statistical program, such as *Minitab*, can be used to write the files in the required format.

Appendices

I. Mortality Rates

Suppose the mortality rates come from the rectangles of the Lexis-diagram (as opposed to parallelograms that correspond to survival to a given age at the end of the year). Formally the input files look like Files 5 and 6, but the numbers have a different interpretation.

Step 4.5N. Click on **Specification of mortality**. Verify that there is a check mark next to **Mortality rates**. (This opens two options, to be discussed in Step 4.7N, in **Parameters and Options**. Depending on how you started, you may have to repeat the choice, until this happens.)

Step 4.7N. Click on **Parameters and Options**. Click on **Sex ratio**. Choose 1.05, or other value you like. Click on **Correction term for survival probability**. Choose either **Keyfitz' method**, or **Reed-Merrell formula**. Click on **Separation factors**. This is the *fraction of deaths in age zero that are due to those born during the previous year*. Choose 0.15, or other value you like, for both males and females.

II. Hazard Rates

Suppose exact hazard increments have already been estimated (as opposed to having to use age-specific rates). Formally there are two input files that look like Files 5 and 6. Only now we input annual hazard increments: from birth to age 0, from exact age 0 to exact age 1,..., from exact age 6 to exact age 7. An additional input file for the average probability of survival in the highest age is needed.

Step 2.5O. Copy the file **sur.dat** into c:\tutorial. It contains the *point forecast of survival in the highest age* by sex for each forecast year. See File 18.

Step 4.5O. Click on **Specification of mortality**. Verify that there is a check mark next to **Hazard rates**. (This opens two options, to be discussed in Step 4.7N, in **Parameters and Options**. Depending on how you started, you may have to repeat the choice, until this happens.)

Step 4.3O. Click on **Input files and parameters**. Click on **Parameters for...** We intend to run *30 simulation rounds*, for example. Here, and in the sequel, edit the values given, as necessary. Click on **Point forecasts**. The file **sur.dat** must be added as the last required file name.

The order of the last two steps depends on the state *PEP* is in, when you start.

File 1. jpop.dat

```
(0) M   100
(1) M   105
(2) M   110
(3) M   100
(4) M    90
(5) M    80
(6) M    50
(7) M    50
(0) F   100
(1) F   105
(2) F   112
(3) F   105
(4) F    95
(5) F    90
(6) F    65
(7) F    60
```

Notes:
* The file must begin with “(“; then comes age 0, 1,..., up to the highest age, here 7; then comes “)”; then “M” for males “F” for females; then the population size. The format is intended to help the user to check his or her input files, in case there are errors.
* First males then females, M/F written in upper case.

File 2. asmr.dat

```
(0) M   0.01    0.01    0.009   0.008   0.007
(1) M   0.05    0.05    0.047   0.046   0.045
(2) M   0.05    0.05    0.045   0.044   0.04
(3) M   0.07 0.07   0.065   0.064   0.06
(4) M   0.09 0.09   0.085   0.084   0.08
(5) M   0.15 0.15   0.145   0.14        0.13
(6) M   0.30 0.30   0.28        0.27        0.26
(7) M   0.50 0.50   0.49        0.47        0.45
(0) F   0.01 0.01   0.009   0.007   0.005
(1) F   0.04 0.04   0.035   0.034   0.03
(2) F   0.04    0.04    0.037   0.035   0.033
(3) F   0.06    0.06    0.055   0.05        0.04
(4) F   0.08    0.08    0.075   0.07        0.06
(5) F   0.12    0.12    0.115   0.11        0.1
(6) F   0.25    0.25    0.24        0.2     0.18
(7) F   0.40    0.40    0.37        0.35        0.33
```

Notes:
* In this illustration we assume that these are “projective mortality rates”, i.e., rates from the parallelograms of the Lexis diagram that correspond to birth cohorts. The first is the probability of survival from birth to age 0, the second from age 0 to age 1. PEP will extrapolate the last survival probability from 7 to 7 automatically.

File 3. asfr.dat

```
(2  )   FER 500.0   500 450 450 450
(  3) FER   1000.0  1000    900 900.0   900.0
(4) FER 200.0   200.0   200 200     200
```

Notes:
* the rates are given as multiplied by 1000
* These are rates from the rectangles of the Lexis diagram.
* PEP reads in the files using free format. It assumes that the next age begins with “(“.
* Space is a delimiter here as elsewhere.

File 4. net.dat

```
(0) M   1   1   0.0 0.0 0.0
(1) M   5   5   0   -4  -3
(2) M   10  5   0   -4  -3
(3) M   5 3 0   0   0
(4) M   5 3 0   0   0
(5) M   3 0 0   0       0
(6) M   0 0 0       0       0
(7) M   0 0 0       0       0
(0) F   0 0 0   0   0
(1) F   4 4 0   -3  -3
(2) F   4   4   0   -3  -2
(3) F   0   0   0   0       0
(4) F   0   0   0   0       0
(5) F   0   0   0   0       0
(6) F   0   0   0       0       0
(7) F   0   0   0       0       0
```

Notes:
* Absolute numbers given.
* Net migration can be negative.

File 5. sc_asmr.dat

```
(0) M   0.04    0.04    0.04    0.04    0.04
(1) M   0.04    0.04    0.04    0.04    0.04
(2) M   0.07    0.07    0.04    0.04    0.04
(3) M   0.07 0.07   0.04    0.04    0.04
(4) M   0.04 0.04   0.04    0.04    0.04
(5) M   0.04 0.04   0.04    0.04    0.04
(6) M   0.04 0.04   0.04    0.04    0.04
(7) M   0.04 0.04   0.04    0.04    0.04
(8) M   0.04 0.04   0.04    0.04    0.04
(0) F   0.04 0.04   0.04    0.04    0.04
(1) F   0.04 0.04   0.04    0.04    0.04
(2) F   0.06    0.06    0.06    0.06    0.06
(3) F   0.06    0.06    0.06    0.06    0.06
(4) F   0.04    0.04    0.04    0.04    0.04
(5) F   0.04    0.04    0.04    0.04    0.04
(6) F   0.04    0.04    0.04    0.04    0.04
(7) F   0.04    0.04    0.04    0.04    0.04
(8) F   0.04 0.04   0.04    0.04    0.04
```

Notes:
* Scales give the standard deviation of unit increment of error for log-mortality.
* “0" refers to survival from birth to age 0, “8" refers to survival from age 7 to age 7.

File 6. kp_asmr.dat

```
(0) M   0.15
(1) M   0.15
(2) M   0.15
(3) M   0.15
(4) M   0.15
(5) M   0.15
(6) M   0.15
(7) M   0.15
(8) M   0.15
(0) F   0.15
(1) F   0.15
(2) F   0.15
(3) F   0.15
(4) F   0.15
(5) F   0.15
(6) F   0.15
(7) F   0.15
(8) F   0.15
```

Notes:
* Kappas don’t have to be the same for each age.
* Kappas determine the constant correlation of unit increments.

File 7. sc_asfr.dat

```
(2) FER 0.06    0.06    0.06    0.06    0.06
(3) FER 0.06    0.06    0.06    0.06    0.06
(4) FER 0.06    0.06    0.06    0.06    0.06
```

Notes:
* Scales for fertility need not to be the same as for mortality, but empirical work shows they are often similar.

File 8. kp_asfr.dat

```
(2) FER 0.01
(3) FER 0.01
(4) FER 0.01
```

File 9. sc_net.dat

```
() M    0.1 0.2 0.3 0.5 0.6
() F    0.1 0.2 0.3 0.5 0.5
```

Notes:
* Scales give the standard deviation for absolute numbers (not logs).
* Parentheses are empty. They are used to conform to the other read-in procedures.

File 10. Kp_net.dat

```
() M    0.5
() F    0.5
```

File 11. gross.dat

```
(0) YEAR    20 20
(1) YEAR    25 25
(2) YEAR    30 30
(3) YEAR    40 40
(4) YEAR    40 40
(5) YEAR    40 40
```

Notes:
* A guess for the number of gross migrants is used to specify uncertainty.

File 12. age_gr.dat

```
(0) M   0.1
(1) M   0.1
(2) M   0.3
(3) M   0.2
(4) M   0.1
(5) M   0.1
(6) M   0.1
(7) M   0.0
(0) F   0.1
(1) F   0.1
(2) F   0.3
(3) F   0.2
(4) F   0.1
(5) F   0.1
(6) F   0.1
(7) F   0.0
```

Notes:
* This is the age-distribution of gross migration separately for each sex. Together with File 11 it determines the absolute size of gross migration by year. Multiplying these numbers with the scales of File 9 gives the standard deviations for error in the annual forecast of net-migration by age and sex for each future year.
* As the numbers from the three files are multiplied together, the representation is not unique!

File 13. params.dat

```
3
30 5 2 4 7
1.05
1
0
0.000  0.000
  0.89  0.85  0.8  0.95  0.905
  0.877  0.879  0.795  0.963  0.906
2 2
2 2
10 10
JPOP.DAT
ASMR.DAT
KP_ASMR.DAT
SC_ASMR.DAT
ASFR.DAT
KP_ASFR.DAT
SC_ASFR.DAT
NET.DAT
KP_NET.DAT
SC_NET.DAT
AGE_GR.DAT
GROSS.DAT
1
2 3 1 3
0 1
```

Notes:

* The significance of the various, admittedly cryptic, codes can be discerned by going through the pull down menus, before the program is run, or from the log file that is created after PEP has run.
* It is not necessary to use the parameter file at all. All parameters can be given interactively by going through the pull down menus. However, we recommend that you use the parameter file, in order to minimize the possibility of typos, and that you go through all menus to check that everything is as intended.

File 14. pep.log

```
*************************************************************

                PEP.LOG

*************************************************************


Date 8.5.2003
Time 10:48:32.



The following parameter values were used:

- number of simulation rounds: 30
- number of forecast years: 5
- lowest age of child-bearing: 2
- highest age of child-bearing: 4
- highest age: 7


The following data files were used:

 (1) jump-off population: JPOP.DAT
 (2) projective mortality rates: ASMR.DAT
 (3) kappas for mortality rates: KP_ASMR.DAT
 (4) scales for mortality rates: SC_ASMR.DAT
 (5) future fertility rates: ASFR.DAT
 (6) kappas for fertility rates: KP_ASFR.DAT
 (7) scales for fertility rates: SC_ASFR.DAT
 (8) future net-migration rates: NET.DAT
 (9) kappas for net-migration numbers: KP_NET.DAT
(10) scales for net-migration numbers: SC_NET.DAT
(11) age-distribution of gross-migration numbers: AGE_GR.DAT
(12) future gross-migration numbers: GROSS.DAT


- sex ratio at birth between boys and girls: 1.050
- seed to initialize the random number generator: 1
- correlation structure of forecast error for the eta term
  of mortality: AR(1) model
- rho (eta) for male mortality: 0.890
- rho (eta) for female mortality: 0.850
- rho (eta) for the correlation of mortality between the
sexes: 0.800

- correlation structure of forecast error for the delta term
  of mortality: AR(1) model
- rho (delta) for male mortality: 0.877
- rho (delta) for female mortality: 0.879
- rho (delta) for the correlation of mortality between the
sexes: 0.795
```

- correlation structure of forecast error for the eta term
  of fertility: AR(1) model
- rho (eta) for fertility: 0.950

- correlation structure of forecast error for the delta term
  of fertility: AR(1) model
- rho (delta) for fertility: 0.963

- rho (eta) for the correlation of net-migration between the sexes: 0.905
- rho (delta) for the correlation of net-migration between the sexes: 0.906

- limit year for mortality: 10
- limit year for fertility: 10

Files Px_d1.S1, x = 0 to 30, and files Yy_d1.S1, y = 1 to 5, are written into C:\tutorial.

The aggregated output:

- annual files were chosen to be output
- both males and females were chosen to be output combined together
- user-specified age-groups were used
- number of age-groups was 3
- the low and high limit of each group were:
    group 1: 0 and 1
    group 2: 2 and 4
    group 3: 5 and 7

Files Yx_0.C1, y = 1 to 5, are in the directory C:\tutorial.

File 15. Y5_0.C1

0-1	2-4	5-7
214	436	495
282	524	479
260	440	449
219	431	486
350	583	463
214	410	413
303	433	466
289	471	422
169	329	395
288	504	473
254	425	426
174	313	369
258	454	425
231	425	463
259	455	452
316	502	360
207	346	391
221	387	441
247	448	427
251	433	433
232	462	440
269	422	406
208	412	419
228	407	464
191	381	436
310	483	452
216	421	474
208	407	463
277	434	436
201	388	428

File 16. P1_1.C1

0-1	2-4	5-7
366	626	412
324	623	427
279	574	472
237	516	474
214	436	495

File 17. Combine.log

```
***************************************************************

                    COMBINE.LOG

***************************************************************


Date 8.5.2003
Time 11:20:10.



The aggregated output:

- sample paths were chosen to be output
- both males and females were chosen to be output combined
together
- user-specified age-groups were used
- number of age-groups was 3
- the low and high limit of each group were:
     group 1: 0 and 1
     group 2: 2 and 4
     group 3: 5 and 7


Files Px_1.C1, x = 0 to 30, are in the directory C:\tutorial.
```

File 18. Probabilities of surviving in highest age.

```
(1) YEAR   0.10   0.10
(2) YEAR   0.10   0.09
(3) YEAR   0.10   0.08
(4) YEAR   0.09   0.08
(5) YEAR   0.09   0.08
```

The first column contains the probability for males, the second column the probability for females of the year in question.

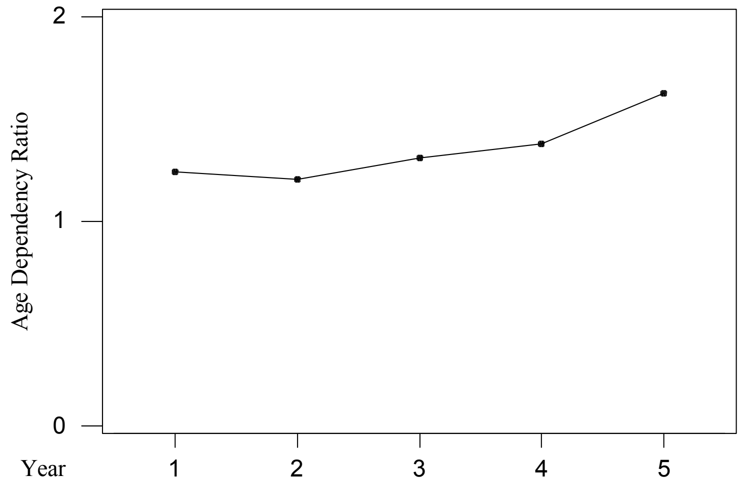

Figure 1. A Sample Path of the Age Dependency Ratio.

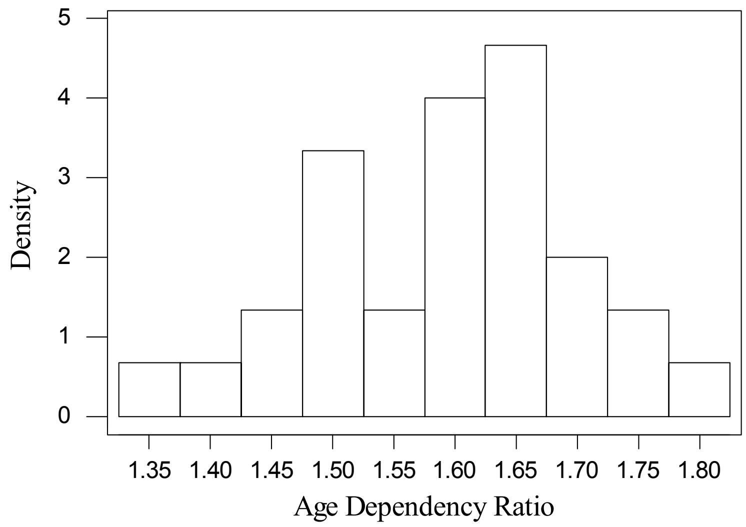

Figure 2. The Predictive Distribution of the Age Dependency Ratio in Year 5.