Liet. matem. rink., T. 42, spec. nr., 548–552

One parametric fertility model

Remigijus LAPINSKAS (VU),
Ramunė VERIKAITĖ (Lithuanian Department of Statistics)
e-mail: remigijus.lapinskas@maf.vu.lt

In our paper [1], we proposed a parametric model for the Lithuanian age-specific fertility curve. The female fertile period age vector $x = (15, 16, \dots, 49)$ was transformed to $x = (1/36, 2/36, \dots, 35/36)$ and the fertility function (by definition, this is a number of children born by 1000 x-years old females during the tth year) parameterized as

$$f(xx, t) = a(1 - xx)^b \exp\left(-c(1 - xx)^d\right), \tag{1}$$

$t = 1985, \dots, 1999$. The four unknown coefficients a, b, c, and d were found by using nonlinear regression methods. We also established that the evolution of the total fertility rate $TFR = TFR_t = \sum_{i=1}^{35} f(i/36, t)$ is governed by the ARIMA(1,1,0) process

$$TFR_t = -18 + 1.4469 \cdot TFR_{t-1} - 0.4469 \cdot TFR_{t-2} + \varepsilon_t. \tag{2}$$

Now we want to check the validity of our models by applying them to the data of the year 2000. According to formula (2),

$$\begin{aligned} TFR_{2000} &= -18 + 1.4469 \cdot TFR_{1999} - 0.4469 \cdot TFR_{1998} \\ &= -18 + 1.4469 \cdot 1348.9 - 0.4469 \cdot 1363.6 = 1324.3, \end{aligned}$$

while, in fact, we have $TFR_{2000} = 1272.0$, i.e., even less than (quite pessimistic) model (2) predicts. We can also check the predictions given by the linear regression formulas (see [1]) for the coefficients a, b, c, and d:

$$\begin{aligned} &\hat{a}_{2000} = -128.4717 + 0.3013 \cdot TFR_{2000} = 254.7819; && a_{2000} = 253.3965; \\ &\hat{b}_{2000} = 1.0136 + 0.0013 \cdot TFR_{2000} = 2.6672; && b_{2000} = 2.5774; \\ &\hat{c}_{2000} = -1.4296 + 0.0057 \cdot TFR_{2000} = 5.8208; && c_{2000} = 6.1844; \\ &\hat{d}_{2000} = -0.7261 + 0.0096 \cdot TFR_{2000} = 11.4851; && d_{2000} = 11.3366. \end{aligned}$$

In the Lithuanian case, all the calculations were performed on full single-year data. However, if we want to apply our results to other European countries, we are confronted by the fact that the available fertility data are usually given in five-year groups. To extend our model, instead of the model for full data

$$
\begin{aligned}
y_1 &= a(1 - xx_1)^b \exp\left(-c(1 - xx_1)^d\right),\\
y_2 &= a(1 - xx_2)^b \exp\left(-c(1 - xx_2)^d\right),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
y_{35} &= a(1 - xx_{35})^b \exp\left(-c(1 - xx_{35})^d\right),
\end{aligned}
\qquad (3)
$$

we shall use the model for grouped data

$$
\begin{aligned}
Y_1 &= y_1 + \cdots + y_5 = a \sum_{i=1}^{5} (1 - xx_i)^b \exp\left(-c(1 - xx_i)^d\right),\\
Y_2 &= y_6 + \cdots + y_{10} = a \sum_{i=6}^{10} (1 - xx_i)^b \exp\left(-c(1 - xx_i)^d\right),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
Y_7 &= y_{31} + \cdots + y_{35} = a \sum_{i=31}^{35} (1 - xx_i)^b \exp\left(-c(1 - xx_i)^d\right).
\end{aligned}
\qquad (4)
$$

Model (4) proves to be quite accurate (see Table 1 and Fig. 1 below for the Lithuanian-1995 fertility data).

We believe that the coefficients a, b, c, and d contain some intrinsic information about the population. In particular, the increasing and decreasing populations should have different sets of coefficients. We applied a linear discrimination procedure to distinguish between these two classes of populations. As a training set `train` (56 records) we took a few European countries (Lithuania was not included) with a steady trend of development:

Table 1

	Full	Grouped
a	326.362	356.979
b	3.030	3.167
c	6.591	5.902
d	13.548	12.075
TFR	1.4914	1.4914
$\widehat{TFR}$	1.4934	1.4890

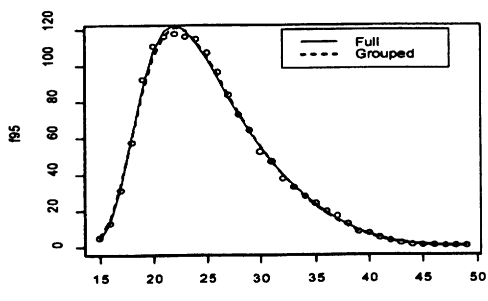

Fig. 1. Lithuanian-1995 fertility data and two approximations based on full and grouped data.

Trend	Country	a	b	c	d
decr	port	389.5070	1.504157	6.652522	10.170671
decr	port	387.0239	1.518283	6.653629	10.425191
....					
decr	roman	324.0535	3.326837	14.599223	23.224870
decr	roman	327.1632	3.473457	12.380310	21.497159
decr	roman	340.0063	3.504578	8.193717	16.814467
....					
decr	bulg	390.3905	4.101414	4.192958	10.705623
decr	bulg	357.9738	3.966205	3.816104	8.962975
decr	bulg	459.1256	4.219507	3.743525	7.336134
decr	bulg	807.1508	3.601820	5.770070	5.907014
....					
incr	austr	817.9103	3.610144	5.801562	5.858395
incr	austr	784.1178	3.542240	6.048635	5.938832
....					
incr	belg	2847.3120	5.277897	9.615142	5.912965
....					

Below is the output of the `lda` function from the R statistical package.

```
Call:
lda.formula(trend ~ a + b + c + d, data = train)

Prior probabilities of groups:
  decr   incr
 0.375  0.625

Group means:
              a          b          c          d
decr    391.3525   3.074562   7.040213   12.88016
incr   1446.6652   3.627961   7.603910    6.88000

Coefficients of linear discriminants:
            LD1
a    -0.001164792
b     0.306927520
c     0.683894824
d    -0.520167306
```

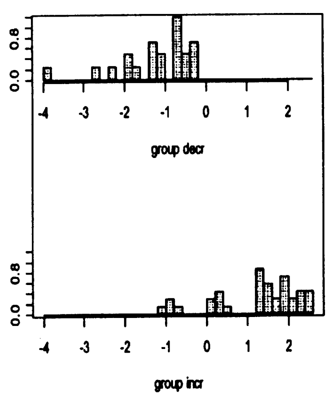

Fig. 2. Values of the `lda` for the decreasing and increasing populations.

Table 2

Lithuania			Romania			Bulgaria		
1	incr	1985	1	decr	1980	1	decr	1975
2	incr	1986	2	decr	1981	2	decr	1976
3	incr	1987	3	incr	1982	3	decr	1977
4	incr	1988	4	incr	1983	4	decr	1978
5	incr	1989	5	decr	1984	**5**	**decr**	**1979**
6	incr	1990	6	decr	1985	6	decr	1980
7	incr	1991	7	decr	1986	7	incr	1981
8	incr	1992	8	decr	1987	8	decr	1982
9	**decr**	**1993**	9	decr	1988	9	decr	1983
10	**decr**	**1994**	10	decr	1989	10	decr	1984
11	**decr**	**1995**	11	decr	1990	11	decr	1985
12	**decr**	**1996**	**12**	**decr**	**1991**	**12**	**decr**	**1986**
13	**decr**	**1997**	**13**	**decr**	**1992**	13	decr	1987
14	**decr**	**1998**	**14**	**decr**	**1993**	14	decr	1988
15	**decr**	**1999**	**15**	**decr**	**1994**	15	decr	1989
16	**decr**	**2000**	**16**	**decr**	**1995**	**16**	**decr**	**1990**
			17	**decr**	**1996**	**17**	**incr**	**1991**
			18	**decr**	**1997**	**18**	**decr**	**1992**
			19	**decr**	**1998**	**19**	**decr**	**1993**
			20	**incr**	**1999**	**20**	**decr**	**1994**
			21	**incr**	**2000**	**21**	**decr**	**1995**

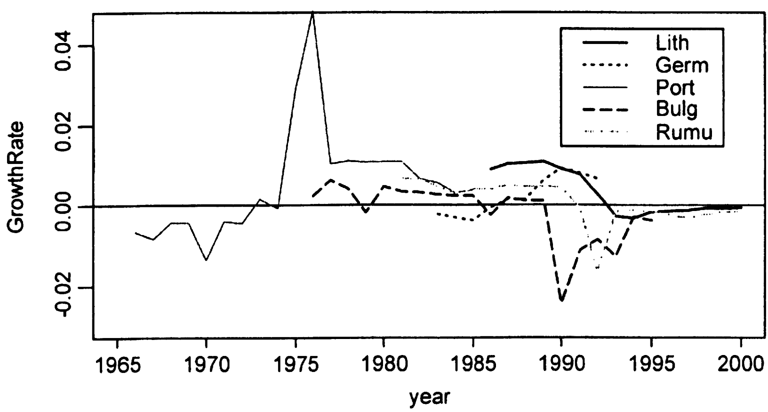

Fig. 3. Population growth rate of some European countries.

This procedure just ideally discriminates the periods of the population growth and decline for Lithuania, but not so succesfully for, say, Romania or Bulgaria (see Table 2; the years in bold letters mark the years with actual decline in population).

It was difficult to expect a perfect discrimination whatever is the selection for a train set and, therefore, the Lithuania's case comes as a surprise. One of the possible explanations is the "regular" behavior of its development compared with somewhat erratic curves for other countries (see Fig. 3).

References

[1] R. Lapinskas, R. Verikaitė, Population projection: a parametric approach, *Liet. Matem. Rink.*, **41** (spec. nr.), 538–541 (2001).

[2] P.A. Thompson, W.R. Bell, J.F. Long, R.B. Miller, Multivariate time series projections of parameterized age-specific fertility rates, *Journal of the American Statistical Association*, **84**(407), 689–699 (1989).

Vienas parametrinis fertilumo modelis

R. Lapinskas, R. Verikaitė

Šiame darbe mes praplėtėme savo ankstesnį netiesinį fertilumo modelį ir pritaikėme jį grupuotiems duomenims. Praplėstą modelį naudojome, bandydami atskirti Europos augančias ir mažėjančias populiacijas. Taip pat nustatėme, kad 2000 metais Lietuvos moterų suminis fertilumas buvo mažesnis nei ankstesnio modelio prognozuojamas.