EKONOMETRIA ECONOMETRICS 32 • 2011
ISSN 1507-3866

Joanna Krupowicz
Wrocław University of Economics

FORECASTING THE NUMBER OF BIRTHS IN POLAND USING LEADING INDICATORS

Summary: The concepts of leading and lagging indicators are of great use in studies of cyclical fluctuations in economic phenomena. The article discusses the possible application of the concepts in analysing and forecasting the birth process in Poland. The article was designed to use leading indicators for projecting the number of births in Poland. For the reference variable (the number of births), leading indicators were sought in the set of potential demographic variables. The set comprised variables describing the structural factors responsible for the birth process. A synthetic leading indicator was constructed and then the projected numbers of births in Poland for 2006-2015 were computed. Subsequently, the projections for the years 2006-2009 were verified in terms of quality.

Key words: cyclical fluctuations in population processes, reference variable, leading indicators, forecasting the number of births.

1. Cyclical fluctuations of demographic variables

The author's earlier research into Poland's demographic situation, covering 1950-1996, enabled her to discover cyclical fluctuations in fundamental demographic processes. Numerous variables were subject to fluctuations, including: population number, number of births, number of deaths of infants, number of marriages contracted, number of divorces, and the average age of mother at childbirth. The fluctuations were pinpointed, leading indicators were identified as symptoms of demographic changes and were subsequently used for developing a forecasting tool for projecting population figures (cf. [Krupowicz 2000]; Krupowicz 2001]).The author's further research focused on analysing and forecasting the number of births by means of leading and lagging indicators. However, the analysis was limited to factors related to the structure of women of childbearing age or to factors related to fertility of women of childbearing age (cf. [Krupowicz 2009a; Krupowicz 2009b]). The research findings presented in this article relate to demographic factors responsible for the birth process, i.e. both the structure of women of childbearing age of 15-49 years, female fertility characterised by fertility rate and the average age of mother at childbirth, and the number of marriages contracted.

The literature on the subject identifies a number of factors determining the birth process. A list of such factors can be found, for instance, in a work by J.Z. Holzer (cf. [1964, pp. 15-16]).He groups birth process determinants into: measurable factors, which can be expressed numerically and whose trends can be found, and non-measurable factors, which are difficult to reflect in figures, but whose directions of impact in the past and in the future can be shown. The first group includes demographic factors, i.e. number of women of childbearing age of 15-49 years, number of population aged 20-34 by sex, number of marriages contracted and newlyweds by sex. The other group of factors comprises: welfare and social factors, economic, psychological, legal-and-administrative factors as well as extraordinary factors (such as war, natural disasters or crises). The above factors exert their impact permanently and over a long period, with the exception of extraordinary factors, i.e. various calamities and disasters, which are incidental in nature and which have an immediate impact on births.

The effect of birth process determinants is subject to constant research and monitoring; this is attested, for instance, by reports of the Governmental Population Council regarding Poland's demographic situation (cf. [*Sytuacja demograficzna*... 2009]) or studies of France's demographic situation (cf. [*La conjuncture*… 1970; Monnier 1989]).A synthetic analysis of Poland's birth levels and intensity is presented below. The changes in the number of births in the years 1950-2006 – shown in Illustration 1 – indicate baby booms and population declines. Changes in the structure of women of childbearing age and a decrease in the number of new marriages, and partly actions of legal nature undertaken under population and socio-economic policies, have contributed to a sizeable drop in the number of births in the years 1984-1996. In the 1990s socio-economic factors related to the transformation period considerably contributed to a reduction in the number of births. In particular, the drop in the number of births in the years 1984-1990 was caused both by decreasing female fertility and by unfavourable changes in the age structure of women of procreative age (especially of the age of the highest reproduction rate, 20-29 years) (cf. [Dzienio, Drzewieniecka 1997; Kędelski 1993]). In turn, the fall in the number of births in the years 1991-2003 was basically caused only by a further sizeable decrease in female fertility.

The number and age structure of women of childbearing age of 15-49 years as well as their total fertility rate (TFR) determine the number of births in the country. In turn, the TFR depends on women's fertility. Over the period in question female fertility in Poland fell dramatically. Although the number of births has been growing since 2003, a birth depression observed for the first time in 1989 has not disappeared. It must be emphasized that women born during the last baby boom are now at the age of the highest or very high fertility, and so the number of births is on the rise.

The demographic changes of the 1990s, among other things, shifted the highest female fertility from the 20-24 age group to the 25-29 age group, and significantly boosted fertility in the 30-34 age group, which is largely a result of women from the

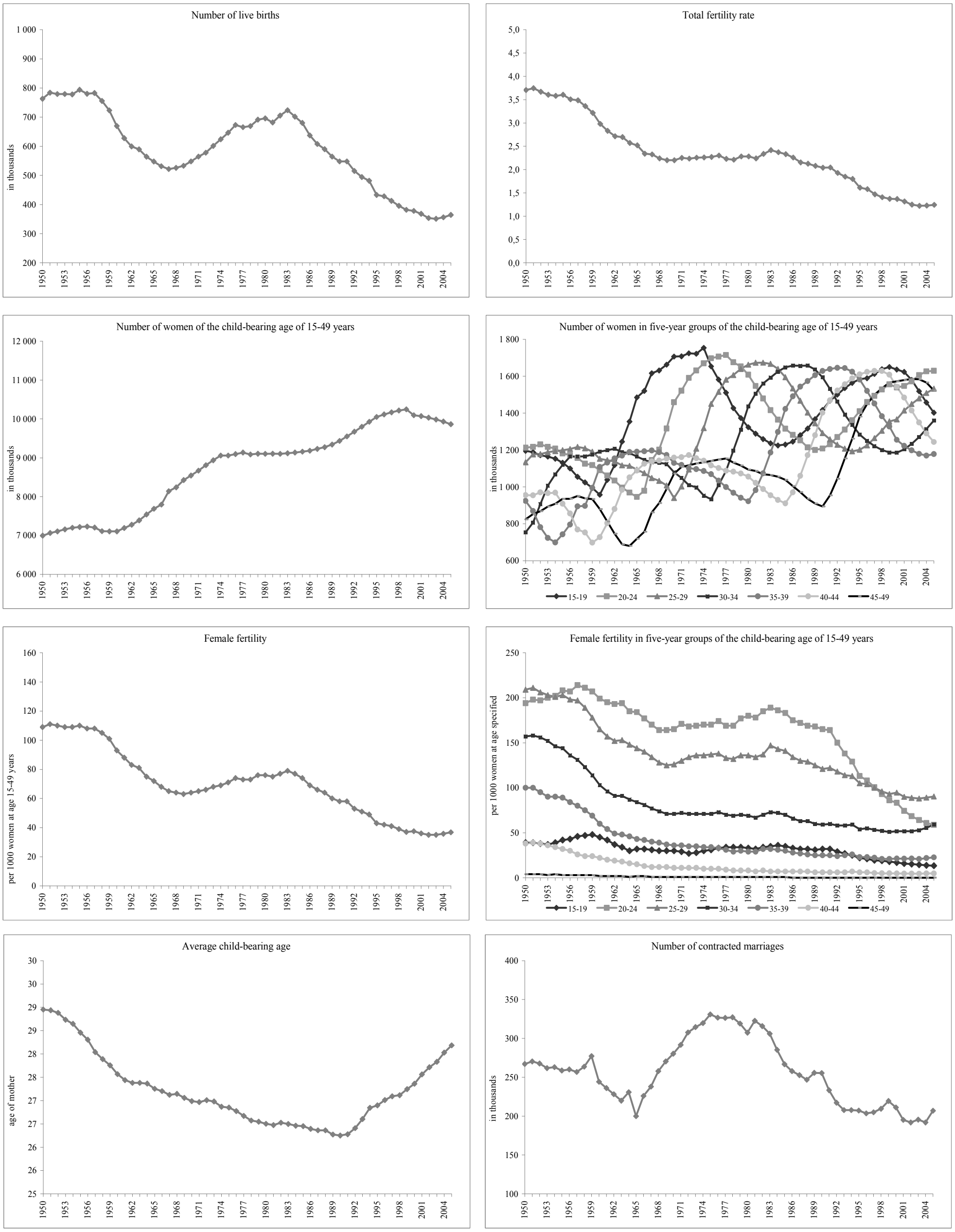

Figure 1. Number of live births and factors affecting the number of births in Poland in the years 1950-2006

Source: Yearly Statistics for 1951-2007, data collected by the Central Statistical Office.

latter group giving 'previously postponed' births. The modification of the pro-family behaviour has increased both the average age of mother and the average age of first-time mother.

The TFR is markedly affected by the number of marriages contracted. A predominant majority of children are born into marriages, with nearly half of all children being born during the first three years of marriage. The growth in the number of marriages contracted observed since 2002 indicates a possible further increase in the number of births, but at the same time the percentage of illegitimate births has been rising for the last fifteen years or so. In the early 1990s about 6-7% of all children were born to such couples, and the rate has recently increased up to 17-20%. As far as extraordinary factors are concerned, no such factors have been identified in the post-war period. However, the results of WWII are still visible in fundamental population processes and the country's population structure.

Figure 1 shows the values of the indicators in question: number of live births[1] and factors influencing the number of births: total fertility rate, number of women of childbearing age 15-49 years, number of women in five-year groups of childbearing age, fertility rate, partial fertility rates in five-year groups of childbearing age, average childbearing age, number of marriages contracted. An analysis of the charts shows a variable speed of such changes over time, which suggests cyclical fluctuations in the factors in question in the period under consideration. The number of live births oscillates visibly around a falling trend, while the number of women of childbearing age fluctuates around a growing trend (Figure 1).

Changes in the numbers of women of childbearing age are more noticeable in five-year age groups. Cyclical rises and falls in the number of women in appropriate age groups are observed around growing trends. In the case of fertility rate, which reflects the ratio of births to the number of women of childbearing age (both general and partial), cyclical variations occur around a falling trend. The last factor – the number of marriages contracted – behaves similarly.

2. The concepts of leading and lagging indicators

Earlier research into cyclical changes in Poland's demographic situation has lead to the identification of leading indicators that can be used to draw conclusions regarding the future demographic situation. The starting point for the analysis was the choice of a reference variable, i.e. a base series reflecting main changes in the demographic situation. Continuing research into the cyclical nature of demographic changes, an attempt was made to find leading indicators for the number of births. As in the case of previous analyses, the reasons for changes in the birth process were exclusively

[1] During the period various definitions of a 'live birth' were in use, which had an effect on the numbers of live births and deaths of infants. This, however, does not have any bearing on the long-term trend, and so for the purposes of this article it has been accepted that the data is comparable.

sought in demographic factors. Thus, the number of births was regarded as a reference variable and the set of potential variables to be divided into leading and lagging indicators comprised a total of 18 variables: number of women of childbearing age (15-49 years old), number of women in five-year childbearing age groups, fertility rate, partial fertility rates in five-year childbearing age groups, total fertility rate, average age of woman at childbirth, and number of marriages contracted. The group of variables under consideration did not include female fertility for the 45-49 age group, as the variable oscillated almost imperceptibly, which excluded it from the set of potential indicators. The available data enabled the study to cover the years 1950-2005.

2.1. The procedure followed to create the set of leading and lagging indicators

The procedure leading to the formation of a set of leading and lagging indicators, adjusted to suit the purpose of the study, constituted in implementing the following stages (cf. [*OECD* 1987; Krupowicz 2000; Krupowicz 2009a]):

1) Elimination of development trends from time series of indicators; absolute deviations for reference variable Y and potential variables X_i were determined:

$$y_t^{'} = y_t - f(t), \tag{1}$$

$$x_{it}^{'} = x_{it} - f_i(t), \tag{2}$$

where: y_t − value of reference variable Y in period t,
$f(t)$ − value of function of trend of reference variable Y in period t,
x_{it} − value of variable X_i in period t,
$f_i(t)$ – value of function of trend of variable X_i in period t.

2) Determination of lags or leads in relation to the reference variable; the coefficients of correlation between deviations from the trend of the reference variable and deviations from the trend of variables from the set of potential variables, accounting for shifts in time, were calculated.

3) Determination of variable classes: leading indicators, converging and lagging indicators; for this the maximum value of correlation coefficient criterion was used.

4) Aggregation of leading indicators into a synthetic leading indicator; the synthetic leading indicator was constructed as the arithmetic mean of standardized leading indicators, taking account a shift in the time series of leading indicators by the established lead period. Variables from the set of leading indicators were normalized using the following formulas:

$$z_{it} = \frac{x_{it} - \bar{x}_i}{\bar{x}_i} + 100, \qquad \text{for } 0 < r_i \leq 1 \tag{3}$$

$$z_{it} = \frac{\overline{x}_i - x_{it}}{\overline{x}_i} + 100, \qquad \text{for } -1 \leq r_i < 0 \tag{4}$$

where: z_{it} − normalized value of indicator X_i in period t,
x_{it} − value of indicator X_i in period t,
$\overline{x}_i$ − mean value of indicator X_i,
r_i − coefficient of correlation between indicator Y and indicator X_i.

5) Making the aggregated leading indicator comparable with the time series of the reference variable; the procedure was similar to that applied at stage one.

2.2. Outcome

In the period covered by the study, i.e. from 1950 to 2005, the trend functions were determined for the reference variable, i.e. number of births, and for the considered potential variables. In an overwhelming majority of cases they were linear trend functions; only for four variables functions in the form of a second-degree polynomial were used. The selected trend function forms reflect long-term single-directional changes in individual variables over time, but they do not account for the oscillation visible in the graphs. The analysis of the values of the deviations from the trend shows a distinctive cyclical nature of each variable under consideration.

In order to form a set of leading indicators correlation analysis was used, in line with the second stage of the procedure. Table 1 presents maximum and minimum values of coefficients of correlation between absolute deviations from the trend for the reference variable and the variables from the set of potential variables. A negative shift value (-p) means that variable X leads reference variable Y by p periods. A positive shift value (p) means that variable X lags in relation to reference variable Y by p periods.

Using the calculated coefficients of correlation, sets of indicators leading and lagging in relation to the reference variable were determined and the appropriate lead and lag periods for each individual indicator were established. The sets included indicators characterised by the highest (in relation to a given modulus) coefficient of correlation (in Table 1 the figures were highlighted in bold). The four-element set of leading indicators comprised:

- number of women of childbearing age 15-59 years with a lead of 24 years,
- fertility rate of women aged 25-29 years with a lead of 14 years,
- fertility rate of women aged 30-34 years with a lead of 16 years,
- total fertility rate with a lead of 14 years.

Because of an extremely short lead or lag period, two variables can be regarded as converging: fertility rate (p = 1) and the number of marriages contracted (p = −4). The remaining variables made up a 12-element set of lagging indicators.

Table 1. Maximum and minimum values of coefficients of correlation between absolute deviations from the trend for the reference variable and potential variables

Variable	a – Shift in time b – Coefficient of correlation						
Number of women of childbearing age	a	**–24**	–7	8	37		
	b	**–0.897**	0.827	–0.456	0.617		
Number of women aged 15−19 years	a	–27	–11	4	18	**30**	
	b	–0.844	0.821	–0.791	0.489	**–0.966**	
Number of women aged 20−24 years	a	–37	–23	–6	9	24	**35**
	b	0.919	–0.821	0.816	–0.879	0.694	**–0.957**
Number of women aged 25−29 years	a	–32	–19	–1	**14**	29	
	b	0.857	–0.718	0.829	**–0.865**	0.838	
Number of women aged 30−34 years	a	–26	–10	4	19	**33**	
	b	0.616	–0.449	0.831	–0.915	**0.950**	
Number of women aged 35−39 years	a	–36	–21	–5	9	**24**	37
	b	–0.756	0.653	–0.530	0.733	**–0.981**	0.976
Number of women aged 40−44 years	a	–26	–16	1	15	**29**	
	b	–0.266	0.296	–0.551	0.778	**–0.988**	
Number of women aged 45−49 years	a	–36	–20	–7	7	21	**34**
	b	0.797	–0.362	0.209	–0.671	0.898	**–0.987**
Fertility rate	a	–31	–15	**1**	18	31	
	b	0.882	–0.892	**0.912**	–0.660	0.850	
Fertility rate of women aged 15−19 years	a	–27	–12	3	20	**35**	
	b	0.412	–0.551	0.885	–0.856	**0.936**	
Fertility rate of women aged 20−24 years	a	–19	2	**20**	35		
	b	–0.375	0.804	**–0.946**	0.932		
Fertility rate of women aged 25−29 years	a	–32	**–14**	2	22	34	
	b	0.926	**–0.971**	0.761	–0.630	0.841	
Fertility rate of women aged 30−34 years	a	–36	**–16**	0	20	36	
	b	0.887	**–0.923**	0.904	–0.806	0.905	
Fertility rate of women aged 35−39 years	a	–36	–16	0	21	**37**	
	b	0.855	–0.887	0.850	–0.879	**0.949**	
Fertility rate of women aged 40−44 years	a	–37	–17	0	22	**37**	
	b	0.895	–0.926	0.728	–0.878	**0.947**	
Total fertility rate	a	–31	**–14**	3	22	34	
	b	0.932	**–0.971**	0.816	–0.718	0.934	
Average age of woman at childbirth	a	–26	–10	**7**	24	35	
	b	–0.724	0.695	**–0.891**	0.846	–0.861	
Number of marriages contracted	a	–22	**–4**	12	32		
	b	–0.849	**0.881**	–0.796	0.696		

Distinguished coefficients of correlation significant at $\alpha = 0.05$.

Source: own calculations.

Subsequently, a synthetic leading indicator was designed as the arithmetic mean of standardized variables from the set of leading indicators. Standardization was conducted according to formula (4). Aggregation required a shift in the time series of standardized variable values in line with the established lead periods. When designing the synthetic variable, alternative sets of leading indicators were considered; ultimately, two synthetic variables were created:

– synthetic variable I, which uses all variables determined as leading,
– synthetic variable II, which includes: fertility rate of women aged 25-29 years with a lead of 14 years, fertility rate of women aged 30-34 years with a lead of 16 years and total fertility rate with a lead of 14 years.

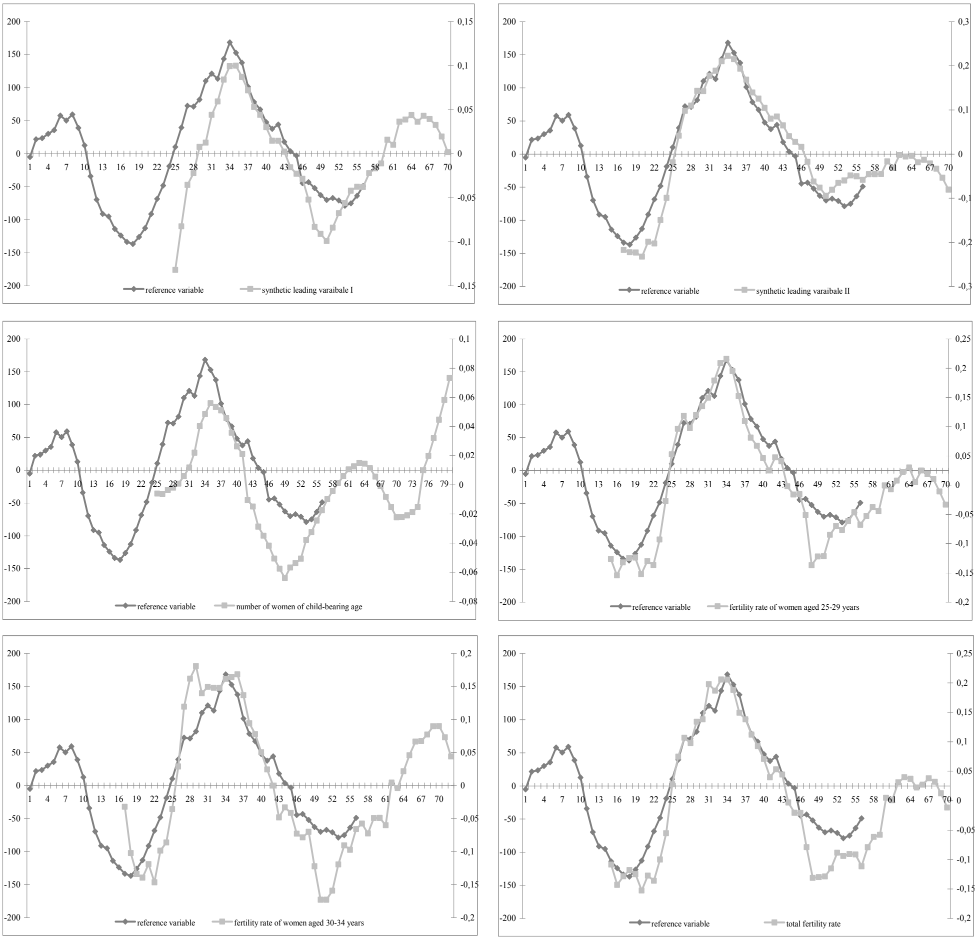

Figure 2. Absolute deviations from the trend for reference variable and synthetic leading indicators (top), and for reference variable and individual leading indicators (bottom)

Source: own calculations.

Individual leading indicators were also considered. For constructed synthetic leading indicators and individual leading indicators, linear trend functions were determined. Figure 2 shows comparatively, i.e. after eliminating the trend, respectively: the reference variable and the synthetic leading indicator, the reference variable and an individual leading indicator.

Analysis of the graphs discloses a similarity between absolute deviations of the reference variable and the synthetic variables, as well as the reference variable and individual leading indicators. Turning points for the synthetic variables and individual variables are observed, growth and fall phases of the variables in question coincide.

Considering the high values of correlation coefficients attesting to similar absolute deviations for the number of births and the variables in question, it can be assumed that both the synthetic leading indicators and individual leading indicators may be regarded as symptoms of changes in the number of births in Poland. The observed deviations of synthetic and individual leading indicators show that the indicators go through the successive change phases earlier than the number of births, which was analysed as the reference variable. Changes in the number of births thus result from earlier changes in the structure of women of childbearing age, fertility rate of women aged 25-29 years and 30-34 years, and total fertility rate. Considering the above, the established leading indicators can be used to predict the number of births in the future.

3. Forecasting number of births by means of leading indicators

3.1. Procedure resulting in developing a projection and its verification

Using the constructed synthetic leading indicators and individual leading indicators, the numbers of births in Poland for the years 2006-2015 were projected. In order to perform the task symptomatic models had to be developed, in which synthetic or individual leading indicators acted as predictors. The constructed models had the following forms:

$$y_t' = a_0 + a_1 \cdot q_{t-p}' \tag{5}$$

$$y_t' = a_0 + a_1 \cdot x_{t-p}' \tag{6}$$

where: y_t' – deviations of reference variable Y from the determined trend function,

q_{t-p}' – deviations of the synthetic leading indicator from determined trend function with lead p in relation to reference variable Y,

x_{t-p}' – deviations of individual leading indicator X from determined trend function with lead p in relation to reference variable Y,

a_0, a_1 – model parameters.

The forecast number of births – reference variable – was established as follows:

$$y_t^* = f(T) + y_t^{*'}, \qquad T > n \tag{7}$$

where: y_t^* – ultimate forecast reference variable Y,
$f(T)$ – forecast reference variable Y from extrapolation of trend function,
$y_t^{*'}$ – forecast deviations of reference variable Y – from a model with a synthetic variable or an individual leading indicator.

In order to verify the established projected numbers of births *ex post* relative errors of the projections were calculated using the following formula (cf. [*Prognozowanie*… 2005, p. 50]):

$$\Psi_t = \frac{y_t - y_t^*}{y_t} \times 100 \qquad t = n + 1, ..., T, \tag{8}$$

where: y_t – actual value of variable Y in period t,
y_t^* – forecast value of variable Y in period t.

3.2. Results obtained

Table 2 presents the parameters of the developed models and the matching of the models with empirical data. Three out of the six developed models were a very good match with empirical data (coefficient of determination in excess of 0.9) and two – a good match ($R^2 > 0.8$). For all the models the parameters were significant at $\alpha = 0.05$.

Figure 3 presents projected ultimate numbers of births in Poland in the years 2006-2015. The projected number of births resulting from the extrapolation of the trend function was obtained on the basis of a linear trend function expressed as follows:

$$f(T) = 774.695 – 6.454 \times T.$$

Table 2. Parameters of models with a synthetic leading indicator and individual leading indicators

Item	a_0	a_1	R^2	s
With a synthetic leading indicator				
I	38.036	1040.866	0.699	43.928
II	–1.204	603.207	0.916	25.706
With individual leading indicators				
Number of women of child-bearing age	39.802	1896.543	0.804	35.454
Fertility rate of women aged 25-29 years	–2.039	758.138	0.942	21.778
Fertility rate of women aged 30-34 years	10.027	689.284	0.851	34.282
Total fertility rate	–0.217	720.533	0.943	21.703

Source: own calculations.

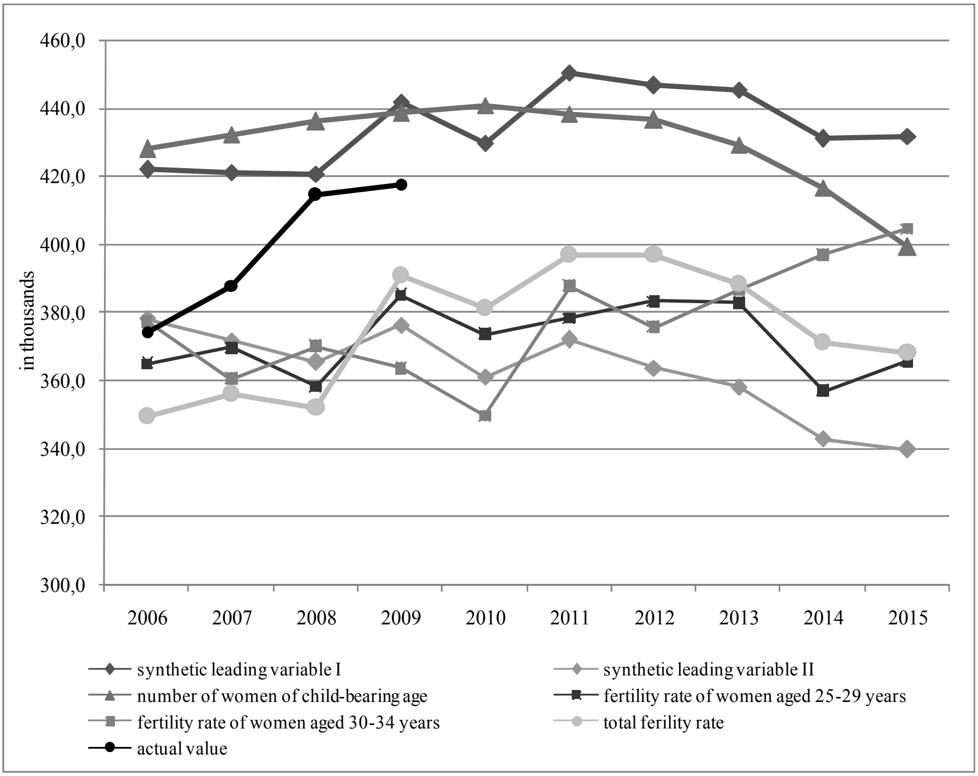

Figure 3. Actual and forecast numbers of births in Poland 2006-2015

Source: own calculations.

The projected numbers of births in the years 2006-2009 were verified. Projections for which the *ex post* relative error did not exceed 10% were recognized as accurate. The values of the ex post relative errors and the average errors being the mean of moduli of individual ex post errors of the projections are gathered in Table 3.

An evaluation of the average accuracy of the projections indicates that the projections proved accurate (the average *ex post* errors did not exceed the adopted critical value of the relative error). A comparison of the quality of the obtained projections shows that the best results, i.e. projections with the smallest errors were obtained when using synthetic variable II, with the following variables as leading indicators, fertility rate of women aged 25-29 years, fertility rate of women aged 30--34 years and total fertility rate, although the 2008 projection turned out inaccurate. Among projections obtained on the basis of individual leading indicators the best qualitatively were those determined using the fertility rate of women aged 25-29 years. Also in this case the 2008 projection proved inaccurate. The observed discrepancies in the *ex post* errors are definitely a result of casual fluctuations

Table 3. *Ex post* errors (in %) of forecast numbers of births in Poland 2006-2009

Item	Relative error				Average relative error
	2006	2007	2008	2009	
On the basis of a synthetic leading indicator					
I	–12.7	–8.5	–1.4	–5.8	7.1
II	–1.0	4.2	11.8	9.9	6.7
On the basis of an individual leading indicator					
Number of women of child-bearing age	–14.4	–11.4	–5.2	–5.0	9.0
Fertility rate of women aged 25-29 years	2.5	4.7	13.6	7.8	7.1
Fertility rate of women aged 30-34 years	–0.8	7.1	10.7	12.9	7.9
Total fertility rate	6.6	8.3	15.1	6.4	9.1

Source: own calculations.

occurring in the obtained absolute deviations of synthetic variables and individual leading indicators (cf. Figure 2). An attempt can be made to improve the quality of the projections by calculating them on the basis of models with smoothed-out values of absolute deviations. Such an attempt has been made, but models of similar quality and projections with similar *ex post* error values were obtained. The average errors did not change and individual errors changed only slightly (not always towards more accuracy).

4. Conclusion

As part of the study an analysis was conducted of changes to the number of births in time and factors affecting the number of births in the years 1950-2005. The concepts of leading and lagging indicators were used for the study. As part of the continuing research into the cyclical nature of demographic changes, an attempt was made to find leading indicators for number of births. As in the case of previous analyses, the reasons for changes in the birth process were exclusively sought among demographic factors. The number of births was regarded as the reference variable. A set of leading indicators for the reference variable was established. The set included: the number of women of childbearing age of 15-49 years with a lead of 24 years, fertility rate of women aged 25-29 years with a lead of 14 years, fertility rate of women aged 30-34 years with a lead of 16 years and total fertility rate with a lead of 14 years. The above leading indicators were used to forecast the number of births in Poland. The obtained projections proved medium accurate in the 2006-2009 verification bracket, which attests to the usefulness of the created set of leading indicators for forecasting the number of births.

Literature

Dzienio K, Drzewieniecka K., *Sytuacja demograficzna Polski. Raport 1996 Rządowej Komisji Ludnościowej*, „Studia Demograficzne" 1997, nr 1.

Holzer J.Z., *Urodzenia i zgony a struktura ludności Polski,* PWE, Warszawa 1964.

Kędelski M., *Demographic decline in Poland (1991−1992),* „Studia Demograficzne" 1993, nr 4.

Krupowicz J., *Koncepcja zmiennych wyprzedzających i naśladujących w badaniach koniunktury demograficznej w Polsce*, „Studia Demograficzne" 2000, nr 1.

Krupowicz J., *Sygnalizatory przemian demograficznych w Polsce,* „Zeszyty Naukowe Sekcji Analiz Demograficznych KND PAN" 2001, nr 3/2001.

Krupowicz J., *Wykorzystanie zmiennych wyprzedzających do prognozowania procesu urodzeń,* Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 38, UE, Wrocław 2009.

Krupowicz J., *Zmiany struktury populacji kobiet w okresie zdolności rozrodczej a kształtowanie się procesu urodzeń w Polsce – propozycje w zakresie prognozowania*, „Zeszyty Naukowe Sekcji Analiz Demograficznych KND PAN" 2009, nr 20.

La conjoncture démographique: La France, "Population" 1970, nr 2.

Monnier A., *La conjoncture démographique: l'Europe et les pays dévelopées d'Outre-Mer,* „Population" 1989, nr 4-5.

OECD Leading Indicators and Business Cycles in Member Countries. Sources and Methods 1960--1985, OECD, Paris 1987.

Prognozowanie gospodarcze. Metody i zastosowania, red. M. Cieślak, PWN, Warszawa 2005.

Roczniki Demograficzne GUS z lat 1951-2007, GUS, Warszawa.

Sytuacja demograficzna Polski – raport 2007-2008, GUS, Warszawa 2009.

PROGNOZOWANIE LICZBY URODZEŃ W POLSCE Z WYKORZYSTANIEM ZMIENNYCH WYPRZEDZAJĄCYCH

Streszczenie: Koncepcja zmiennych wyprzedzających i naśladujących ma szerokie zastosowania w badaniach wahań cyklicznych w zjawiskach ekonomicznych. W artykule została wskazana możliwość zaaplikowania tej koncepcji na grunt analizy i prognozowania procesu urodzeń w Polsce. Celem artykułu było wykorzystanie zmiennych wyprzedzających do prognozowania liczby urodzeń w Polsce. Dla zmiennej referencyjnej (liczby urodzeń) poszukiwano zmiennych wyprzedających w zbiorze potencjalnych zmiennych demograficznych. Zbiór ten tworzyły zmienne charakteryzujące strukturalne czynniki odpowiedzialne za proces urodzeń. Zbudowano syntetyczną zmienną wyprzedzającą, a następnie skonstruowano prognozy liczby urodzeń w Polsce na lata 2006-2015. Przeprowadzono weryfikację jakości prognoz utworzonych na lata 2006-2009.