

Cite: Shevchuk, P. E., Yehorchenkov, O. V. (2020). Probabilistic Demographic Forecasts for Metropolises of Ukraine. *Demography and Social Economy*, 3 (41), 76-90.

УДК 314.303.(477)

JEL Classification: R10

**P.E. SHEVCHUK**

PhD (Economics), Leading scientist

Ptoukha Institute for Demography and Social

Studies of the National Academy of Sciences of Ukraine

01032, Ukraine, Kyiv, Blvd Taras Shevchenko, 60

E-mail: pavlo-shevchuk@ukr.net

ORCID 0000-0003-1158-4438

**O.V. YEHORCHENKOV**

PhD (Engineering), Associate professor,

Taras Schevchenko National University of Kyiv,

04116, Ukraine, Kyiv, Bohdana Havrylyshyna Str., 24

E-mail: alexee@ukr.net

ORCID 0000-0003-1390-5311

## PROBABILISTIC DEMOGRAPHIC FORECASTS FOR METROPOLISES OF UKRAINE

*Uncertainty is an intrinsic characteristic of demographic processes. This applies even more to the future. Accurate deterministic forecasts are fundamentally impossible. This determines the necessity to quantify the future uncertainty. The purpose of this research is to develop probabilistic demographic forecasts for the metropolises of Ukraine and analyze the outcome results. For the first time, probabilistic demographic forecasts have been developed for individual cities of Ukraine. The study was carried out using the functional data approach which incorporates wide set of demographical methods and models implemented in several packages of R-programming language. Chosen methodology is based entirely on statistics and does not require introducing any additional arbitrary hypotheses.*

*At three cases (namely for fertility in Kyiv, Lviv and Kharkiv) the default method (ARIMA) showed implausible results which could be induced by unreliable current data. In these cases we used random walk model. For Odesa the both models give similar results. It is possible that in this city the underestimation of the departed population is compensated by the underestimation of the arrived, which leads to the relevance of the current fertility rates (namely their denominators) and, consequently, the consistency of the forecast results regardless of the method. Mortality forecasts are consistent with the dynamics of mortality rates being observed and the quality of current data. The model captured upward life expectancy trends for Dnipro and Odesa and stagnation for other cities. This is also could be caused by denominator inconsistency*

for the latter ones. Computation showed that the population size of Dnipro, Lviv and Kharkiv in 2040 is most likely to be below the population number reached in early 2019. Some chances for population growth remain in Odesa and Kyiv is likely to have a larger population. The age distribution of the population in all cities in future looks similar. The number of people over 40 years of age has least uncertainty. At the age of 20 to 40 years, the uncertainty is much higher. This is a consequence of the uncertainty of youth migration during the forecast horizon of 2019–2039, because all these cities are powerful educational centres and attract students. In 2040 those who were students in 2020 will reach the age of 40 and can stay in the big city or leave. Uncertainty of the number of persons under 20 is formed from two sources: uncertainty of fertility forecasts and uncertainty of the number of reproductive cohort, i.e. those 20-40-year-olds. It is needed to review these forecasts after receiving the results of the closest census.

**Keywords:** probabilistic forecast, functional time series, uncertainty, big city.

‘... to predict future events in a statistical fashion — that is, by stating the probability for alternate sets of events, rather than flatly predicting that one set will take place.’

*I. Asimov, Prelude to Foundation*

**Formulation of the problem.** Demographic forecasts have always been inaccurate. They were relatively more accurate when their development fell on a period of relatively stable demographic indicators. Times of abrupt changes in fertility or mortality, not to mention migration flows, have always been unexpected. As N. Keilman convincingly shows, the development of the methodology of demographic forecasting and the availability of computing power did not increase the accuracy of population forecasts [1]. Deterministic forecasts cannot be accurate in principle. The probability that the future population will be exactly as predicted is close to zero [2, p. 9]. The reason for this is three main sources of error. The most obvious type of error is an incorrect model specification, namely, inadequate choice of model or its parameters. This includes incorrect statement of hypotheses about future fertility, mortality and migration. Inherent randomness of demographic processes makes it impossible to accurately predict in principle. The second source of error is the current population count data, on the basis of which the model is set up. Unfortunately, a serious challenge for Ukraine in the 21st century is the quality of current population statistics due to the long absence of census and, consequently, the impossibility to specify the denominator of the most important indicators of demographic development. Thus, not only finding the parameters of the model, but also the actual starting population itself and its distribution by sex and age may turn out to be such that do not correspond to reality. The third factor, that may affect the divergence of the forecast from actually observed, once the numbers become available, is the forecast itself. A clear example of the impact on itself is the HIV / AIDS mortality forecast published in 2006. According to this forecast, the number of AIDS deaths in Ukraine in 2014 was supposed to be 59,000 people in the average scenario [3, p. 12]. Public concern about this possible development led to the introduction of a number of

public and private programs for HIV prevention, treatment, care and support for HIV-infected and AIDS patients. As it is already known according to the State Statistics Service of Ukraine, 4,399 people died of HIV-related diseases in 2014. Thus, the forecast-warning contributed to the abolition of itself. Fertility or migration forecasts can work in a similar way, urging the government to take action to regulate certain processes and change their course accordingly.

**Relevance of the research.** The fundamental impossibility of an accurate demographic forecast necessitates the description of the uncertainty of the future. The traditional approach is to develop a multivariate forecast. However, as with one variant, the probability of implementing each of the forecast variants is the same (close to zero), and their number is still too small to describe the diversity of the uncertain future. Keilman identifies two shortcomings of the multivariate approach [4, p. 410]. First, there is no probability attached to the projected range of variants, because the user does not have information on how likely future indicators will fall in the range between “low” and “high” scenarios. Second, the “low” and “high” variants are not realistic because, for example, the “low” variant assumes that the fertility will be low each year of the forecast horizon. Also, the combinations of “low fertility — low life expectancy” and “high fertility — high life expectancy” form quite similar age structures of the population, which artificially narrows the uncertainty of the dynamics of the share of the elderly in the future. The latter problem can be solved by computing the forecast variants in the combinations “low fertility — high life expectancy” and “high fertility — low life expectancy”. However, the user still cannot determine the degree of probability of such a development of events.

Instead, the probabilistic approach offers an alternative and allows, if not to get rid of uncertainty, then at least to quantify it. Probabilistic forecast is not necessarily more accurate than deterministic one. However, it contains more information for user [5, p. 25], on the basis of which the user can make decisions taking into account the degree of uncertainty. Therefore, developers of demographic forecasts, such as the United Nations Department of Economic and Social Affairs (DESA), the national statistical agencies of the Netherlands, New Zealand, and South Korea, publish probabilistic demographic forecasts on a regular basis. The United Nations Economic Commission for Europe (UNECE) advises to represent uncertainty directly [6, p. 2]. In addition, a study in Australia showed that most users of subnational demographic forecasts need information about the degree of uncertainty of the forecast and understand it correctly [7, p. 374]. Thus, the development of probabilistic demographic forecasts is becoming widespread and relevant.

**Analysis of recent research and publications.** The first approaches to the probabilistic interpretation of the demographic forecast were made in the middle of the 20<sup>th</sup> century. Later, there were attempts of an analytical solution. Finally, the development of computer technology has made it possible to make appro-

ximate calculations using simulations. At the end of the 20th century it was proposed to combine the advantages of the probabilistic approach with an expert- and argument-based judgment [8]. To reduce the impact of the subjectivity of an expert opinion, the use of the Bayesian approach was proposed [9]. Namely, the UN Population Division uses the Bayesian hierarchical model to develop probabilistic forecasts [10]. This model is implemented in the form of a system of R language packages [11] and can be applied both to the country as a whole and at the level of administrative districts [12, p. 17]. The results of the computation are published in a form adapted to the format of UN publications: by 5-year age groups and 5-year time intervals [13, p. 818]. Models have been developed that allow for a coherent forecasting [14] of indicators of close (e.g., neighbouring regions) or related (e.g., male-female) populations.

**The purpose of this research** is to develop probabilistic demographic forecasts for the metropolises of Ukraine and analyze the outcome results.

**Scientific novelty.** For the first time, probabilistic demographic forecasts have been developed for individual cities of Ukraine.

**Materials and methods.** This study used the data of the State Statistics Service of Ukraine for Dnipro, Kyiv, Lviv, Odesa, and Kharkiv. It was decided to cancel the development of the forecast for Donetsk due to incomplete population statistics and registration of demographic events in recent years in this city. The forecast for Donetsk from the launch year of 2013 is inconclusive due to the impossibility of taking into consideration the unforeseen events in this region in the coming years. To develop probabilistic demographic forecasts, data were used for the period after the last population census on the number of: permanent population by sex and age; born by mother's age; born by sex; died by year of birth; died by sex and year of birth under the age of 1 year. Data on the number of deaths by age and year of birth in urban settlements of the respective regions were used as additional information.

Newborns whose mother's age is unknown were distributed according to the age structure of the mothers of the respective marital status. In addition, for the correct computation of mortality rates, the number of deaths was distributed by age. As already mentioned, the number of deaths at the age of 0 is known for each calendar year of the base period (2002-2018). The distribution of deaths for the remaining age groups was obtained by splitting the number of deaths by year of birth into the number of deaths by age according to this distribution in the urban area of the respective oblasts, as the distribution of deaths by both age and year of birth is available for the latter. Migration data are the least accurate. Therefore, the net migration by sex for the base period was computed according to the demographic growth-balance equation on the basis of the rest of available data.

The forecast of the population by sex and age is obtained by the cohort-component method. The difference from the traditional approach is that, as a result,

not a few, but many variants are obtained, sometimes several thousand trajectories of probable future development according to the future sample paths for each of these components: fertility, mortality, and migrations.

An important scientific principle is the possibility of independent verification of the obtained results. The advantage of the methodology developed by R. Hyndman et al. [14, 15, 16] is its reproducibility and replicability. Reproducibility is understood as obtaining the corresponding results using the same data, methods and calculation algorithm. Replicability means obtaining consistent results according to the same methodology using other data [17, p. 46]. Unlike the method of expert-based judgment, the parameters of the model are estimated entirely on historical data and do not require subjective assumptions [15, p. 338], although some judgments about the choice of model can be made. In particular, the algorithm [18] is quite flexible. Even in cases of obtaining unrealistic results using the default parameters, it is possible to choose another of the available model modifications implemented in the package, which will be mentioned below.

In any case, a careful indication of all the conditions of the study: the input data and the algorithms used make the study transparent and reproducible. That is, one of the most important scientific principles is fulfilled, which cannot be said about the methods involving expert judgment. There are serious doubts that even the expert himself will be able to reproduce his own research later.

Hyndman et al. in a series of publications have developed a holistic methodology which combines different approaches and models into one framework. According to the functional data paradigm, smoothing of mortality, fertility and migration rates [15, p. 324] using constrained regression splines is carried out before fitting the model. This makes it possible to immediately identify one of the three sources of randomness in the model: the randomness of the variation of births, deaths, or migrants. After fitting the model, the residuals not described by the model (the second source of randomness) are defined. The third source is the randomness inherent to time series [15, p. 326]. Smoothing also makes possible interpolation of birth rates from 5-year age intervals into one-year intervals, if they are not available. In this study, for all cities, fertility rates at 5-year age intervals were used for unification, although more detailed data are available for Kyiv.

R. Hyndman implemented his methodology in package “demography” for R programming language. This package provides functions for demographic analysis including: lifetable calculations; Lee-Carter modelling; functional data analysis of mortality rates, fertility rates, net migration numbers; and stochastic population forecasting (<https://www.rdocumentation.org/packages/demography/versions/1.22>).

The Lee-Carter model is used to forecast mortality [19]. The flexibility of the demography package [18] is in the implementation of several variants of the model. The researcher can choose not only, for example, the standard variant, but also the parameters of the variant itself. Thus, the original Lee-Carter me-

thod, after estimating the parameters of the model using singular value decomposition, involves the adjustment of the overall mortality level parameter to the number of actual deaths. However, this adjustment is optional and can be skipped. You can also choose to adjust the parameters of the model to the actual life expectancy at birth ( $e_0$ ), as in the Lee-Miller variant [20]. The Booth-Maindonald-Smith variant [21] differs in the criteria of parameter adjustment. The Hyndman-Ullah approach involves calculating more than one pair (up to 6) of the relative age-specific rates of mortality change parameter and overall mortality level parameter using the principal components decomposition.

Obviously, if the indicators of a process have a different growth rate for different objects in the base period, then, eventually, they will diverge or intersect. If the processes under study occur with related populations, such as males and females, it is logical to impose certain constraints to prevent this from happening. Since, in the absence of catastrophic events, the directions of change in mortality rates for males and females are not supposed to differ significantly, at least, in long term perspective. This non-divergence of subpopulation indicators is called coherence. Ensuring the non-divergence of life expectancy forecasts for both sexes in long-term was implemented in one of the modifications of the Lee-Carter model — Li-Lee variant [22]. The methodology proposed by Hyndman et al. is currently the most advanced. The product-ratio method developed by them consists in modelling not age coefficients, but square roots from their products and ratios. The advantage of this approach is that these values change roughly independently but have approximately the same variation [14, p. 263]. Confidence intervals are calculated using autoregressive integrated moving average (ARIMA) model [14, p. 265]. It can be noted apropos that DESA experts provide a non-divergence of life expectancy by sex in simpler way. They project the life expectancy for females and the gap in life expectancy between the sexes [10, p. 797].

The Lee-Carter model was generalized to forecast fertility [23]. The Hyndman-Ullah model implemented in the demography package is a successor in the sense that it also uses the principal component decomposition, but it uses more than one principal component, is more robust to random fluctuations during the base period [16], and does not require pre-specified constraints on the upper and lower bounds [23, p. 192] of total fertility rate (TFR), as proposed by Lee earlier.

The forecast of the net migration as compared to the forecast by inflows and outflows has its constraints, as it does not take into account the differentiation of migrants by main directions or countries which are the main migration partners. However, such an approach may be sufficient, given the incompleteness and inaccuracy of migration recording, as well as the variability of the process itself due to changes in legislation (including in other countries), labour market situations, political and environmental factors, and so on.

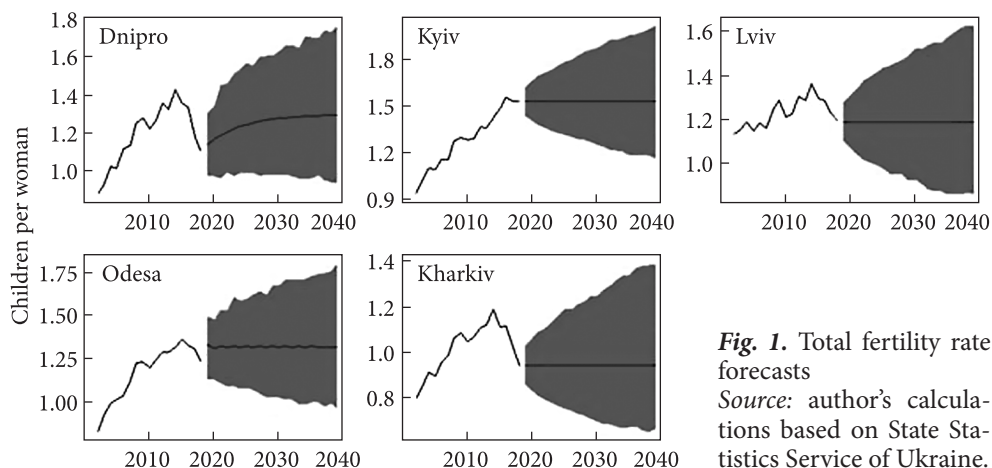
The described methodology is implemented in the software packages demography [18] and forecast [24]. The advantage of these packages compared to



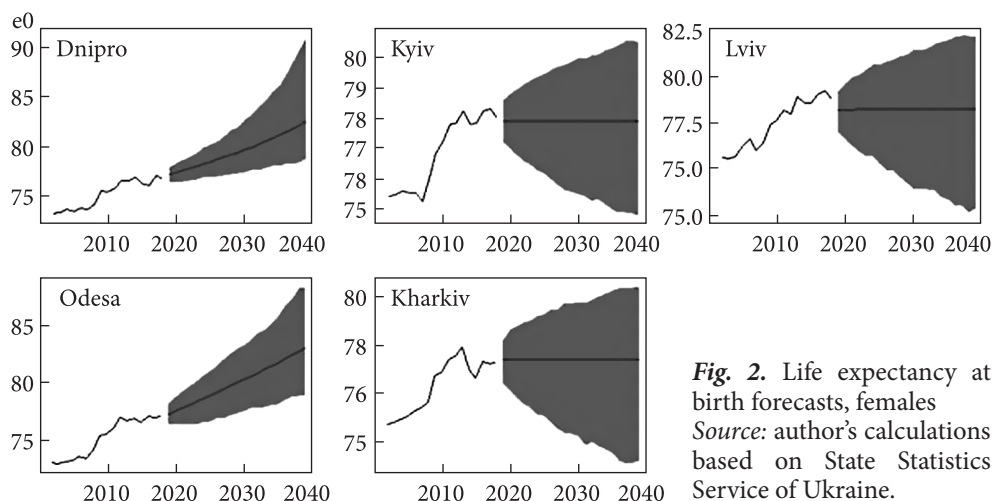
those used by DESA [12] is greater flexibility in detailing the input and output data. Thus, the computation can be carried out not only for 5-year time intervals and 5-year age groups, but for 1-year time intervals and 1-year or 5-year age groups. By default, the width of the confidence intervals is set at 80 % [18, p. 13–15], but in this study in all cases (fertility, mortality and migration) the level of 95 % was used. For every city 1,000 simulations were calculated.

**The main results of the research.** Coincidentally, during the base period, the fertility in Ukraine was mostly increasing. At the same time, in the studied cities the period of birth growth lasted 1–3 years longer than for the country as a whole. This may be due to both peculiarities of recording and deterioration of the accuracy of current population estimates. As for the peculiarities of recording, it is about the possibility since 2016 of recording a birth directly in a maternity hospital. Of course, some women can take advantage of this opportunity, even if they do not live in the locality. Births that occurred in the occupied territories are added to this. Thus, citizens living in the occupied territories can apply to any court and obtain a birth certificate. If this court is not located in the Donetsk or Luhansk oblasts, such a birth will be included in the statistical reporting of the administrative unit where the court is located. Due to more convenient transport accessibility, they can be large cities, such as Dnipro, Kyiv, Kharkiv. Obviously, this practice artificially increases the fertility in large cities. As for the deterioration in accuracy of the current population estimate, it is about giving the birth by those women who actually live in the city but are not registered in it. This error in the current population estimate has accumulated due to the long absence of a census. As all metropolises are attractive for migration, it is likely that their population size, especially those of reproductive age, is higher than according to statistics. Namely, this is the second source of population forecast errors, which was discussed at the beginning of this article. Accordingly, the fertility in large cities is actually lower than can be computed from official data.

Therefore, when computing the fertility forecast, such incorrect data lead to implausible results. Namely: the real growth of the fertility, which was really observed in the first decade of the 21st century, is amplified artificially due to the annual divergence of the denominator (number of women) of birth rates from the real situation. This leads to the ARIMA method extrapolating incorrect data and therefore giving an implausible result. As a result, TFR was obtained in 2039 from almost 2 children per woman in Lviv and Kharkiv to 3 in Kyiv. At present, such a level seems unlikely for Ukraine, and even more so for large cities. Therefore, when developing fertility forecasts for these cities was used random walk model, which is also available in the forecast package [24]. For Dnipro and Odesa, the default method gave plausible results directly. Interestingly, in the case of Odesa, both ARIMA and the random walk model give very similar results. It is possible that in this city the underestimation of the departed population number is compensated by the underestimation of the arrived population num-



**Fig. 1.** Total fertility rate forecasts  
Source: author's calculations based on State Statistics Service of Ukraine.



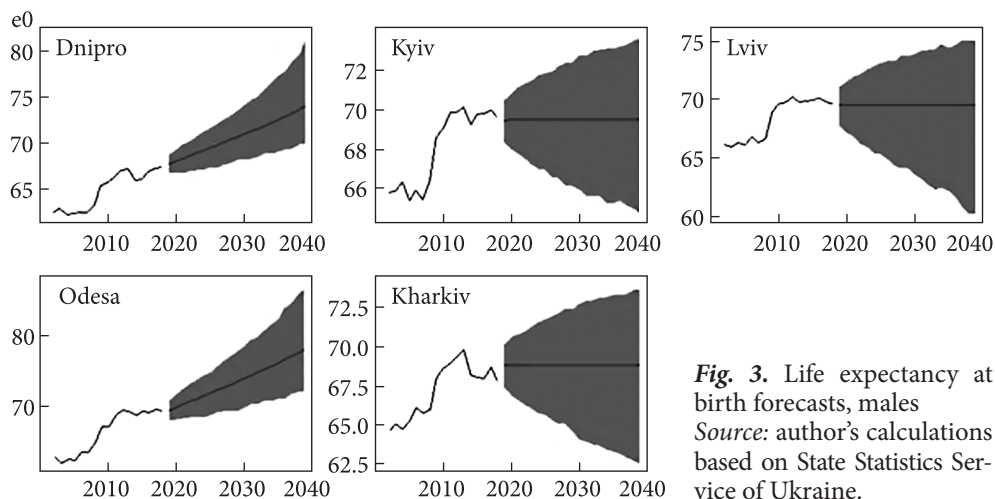
**Fig. 2.** Life expectancy at birth forecasts, females  
Source: author's calculations based on State Statistics Service of Ukraine.

ber, which leads to the relevance of the current fertility rates (namely their denominators) and, consequently, the consistency of the forecast results regardless of the method.

So, the application of different methods to different cities made possible to obtain satisfactory results of the fertility forecasts. Thus, during the forecast horizon TFR with a probability of 95 % will be in the range of 0.93-1.75, 1.16-2.00, 0.88-1.63, 0.97-1.73 and 0.63-1.37 children per woman in Dnipro, Kyiv, Lviv, Odesa and Kharkiv, respectively. The mean values will be 1.29, 1.52, 1.19, 1.32 and 0.94 in Dnipro, Kyiv, Lviv, Odesa and Kharkiv, respectively (Fig. 1).

Mortality forecasts were not particularly difficult, and the results were plausible and consistent with the dynamics of mortality rates being observed and the





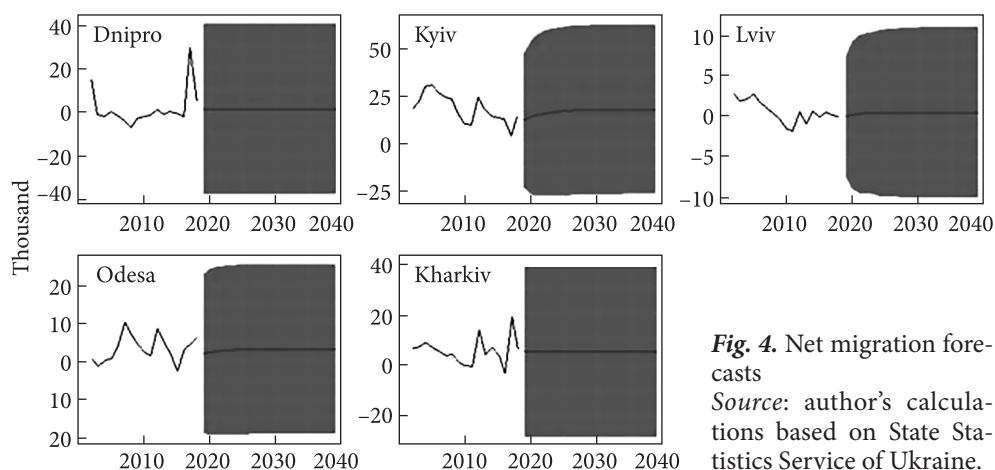
**Fig. 3.** Life expectancy at birth forecasts, males  
Source: author's calculations based on State Statistics Service of Ukraine.

quality of current data. The same default method was used for all cities. Interestingly, the model captured different trends for Dnipro and Odesa here too, compared to Kyiv, Lviv and Kharkiv. Thus, according to the available data and the model, in Dnipro and Odesa, the decline in mortality will continue, and e0 for females will increase to 82.4 and 83.0 years in 2039 with confidence intervals from 78.7 to 91.9 and 78.9 to 89.0 years respectively (Fig. 2).

As the methodology provides coherent forecasts for both sexes, a similar trend will be observed for males in these cities. Thus, e0 for males will increase to 73.4 in Dnipro and to 77.8 in Odesa in 2039. Confidence intervals will be from 69.7 to 81.0 and from 72.1 to 86.5 for males in Dnipro and Odesa, respectively (Fig. 3).

The sharp increase of e0 in 2008-2013 was replaced by stagnation in the following years. It is this stagnation that is extrapolated by the model to the other three metropolises (Figs. 2-3). The functional time series approach involves a geometrically decreasing weight for data more distant in the past [25, p. 200]. This is quite logical in the case of the availability of long time series. In cases with data from Kyiv, Lviv and Kharkiv, stagnation after 2013 proved to be more significant than the decrease in mortality in previous years. As a result, forecasts were obtained with indicators close to those achieved in 2018. Over time, uncertainty will increase, so the confidence intervals are gradually expanding until target year (Figs. 2-3). In 2039, the e0 for females in Kyiv, Lviv and Kharkiv will be in the range of 74.9-80.6, 72.5-82.1 and 72.6-80.2 (Fig. 2). For males, e0 in 2039 will be 64.7-73.4, 61.3-74.9 and 62.1-73.5 years, respectively (Fig. 3).

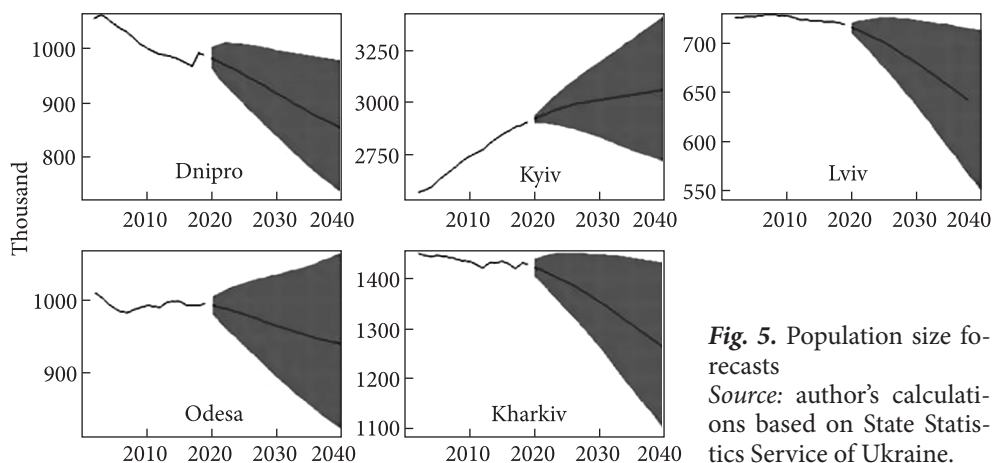
The width of the confidence intervals is larger for males than for females and smaller for larger cities than for smaller ones (Figs. 2-3). That is, the model fairly estimates a greater degree of uncertainty for males and smaller populations, which also indicates its robustness.



**Fig. 4.** Net migration forecasts  
Source: author's calculations based on State Statistics Service of Ukraine.

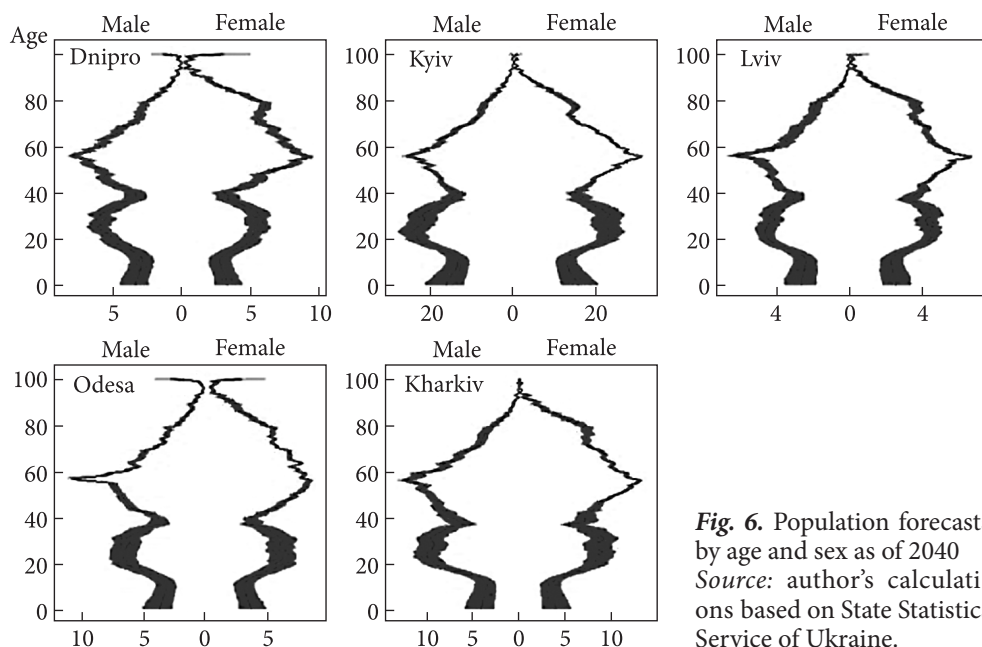
The difference of mortality forecasts for Dnipro and Odesa from Kyiv, Lviv and Kharkiv, as in the case of fertility, may be due to the closer correspondence of the current estimate of the number and sex-age structure of the population to the real situation in the first two cities. Significant stagnation of life expectancy in 2013-2018 in Kyiv and Kharkiv, as well as males in Lviv (Figs. 2-3) may be a consequence, as in the case of fertility, of divergence of the real population size from the current estimate. If the actual population size of these cities is higher than the current data, it generates a relatively larger number of deaths, which artificially reduces life expectancy, computed taking into account the registered population number only. Therefore, it is possible that life expectancy in these cities continues to increase and, unless current data had been distorted due to the long absence of census, the model would have detected this trend, and life expectancy forecasts would have shown further growth, as for Dnipro and Odesa.

Migration forecasts were also computed by default (except for the replacing 80% confidence intervals with 95 %) without any problems. The model averaged the age structure of the net migration for the base period and extrapolated to the target year, wrapping within the confidence interval. It should be noted that all studied cities have a similar age structure of the net migration, as they are powerful educational centres and attract student youth. Therefore, the most vivid feature of all age structures is the “bulge” at about 18 years, which corresponds to the arrival of former schoolchildren in order to continue their education, and “deepening” at about 22-24 years connected with their departure. The mean of the confidence intervals of the net migration forecasts is 1.1 thousand for Dnipro, 17.9 thousand for Kyiv, 0.4 thousand for Lviv, 3.0 thousand for Odesa and 4.8 thousand people annually for Kharkiv (Fig. 4). The confidence intervals for net migration are wide: from -38.0 to +40.2 thousand for Dnipro, -26.2 to +62.0 thousand for Kyiv, from -10.1 to 11.0 thousand for Lviv, from -19.2 to



**Fig. 5.** Population size forecasts

Source: author's calculations based on State Statistics Service of Ukraine.



**Fig. 6.** Population forecasts by age and sex as of 2040

Source: author's calculations based on State Statistics Service of Ukraine.

+25.3 thousand for Odesa, from –28.9 to 38.6 thousand for Kharkiv (Fig. 4). It must be noted that after the closest census these values will be revised thorough.

Computation showed that the population size of Dnipro, Lviv and Kharkiv as of January 1 2040 is most likely to be below the level reached in early 2019. Namely, with a probability of 95 % population of Dnipro will be from 738,000 to 983,000; the population of Lviv — from 550,000 to 715,000; and Kharkiv — from 1.098.000 to 1.432.000 people (Fig. 5). Some chances for population growth remain in Odesa, where the number of inhabitants at the beginning of 2040

may be from 825.000 to 1.060.000 people. Kyiv is likely to have a larger population size of up to 3.421.000 although it may return to 2,717,000 people (Fig. 5).

The age distribution of the population in all cities in 2040 looks quite similar, as it is formed by demographic processes of approximately the same intensity and direction. In particular, the number of people over 40 years of age has less uncertainty, as mortality in the absence of catastrophic events varies slowly, and the intensity of migration at this age is not so high (Fig. 6). At the age of 20 to 40 years, the uncertainty is much higher. This is a consequence of the uncertainty of youth migration during the forecast horizon of 2019-2039, because in 2040 those who were students in 2020 will reach the age of 40 and can stay in the big city or leave. Uncertainty of forecasts of the number of persons under 20 years of age (Fig. 6) is formed from two sources: uncertainty of fertility forecasts and uncertainty of the number of reproductive cohort, i.e. those 20-40-year-olds, who have just been mentioned.

**Conclusions.** The demography and forecast packages correspond to the details of the statistics available in Ukraine, and provide a fairly detailed output that can be used in practice for urban development planning. Another advantage is objectivity in the sense that the computation is based entirely on statistics and does not require any additional arbitrary hypotheses. At the same time, the set of models in the packages is quite wide and provides means for obtaining a satisfactory result even under conditions of a relatively short base period and, possibly, incorrect input data.

Due to possibly incorrect current data, it is appropriate to review these forecasts after receiving the results of the closest census. Namely, it will be interesting to check whether the current population estimate in Dnipro and Odesa is really close to the actual situation. Of course, the forecast for Kyiv will need urgent revision, as its population size probably has a significant divergence from the current estimate. Also, in future studies it is expedient to re-compute the forecasts for the data already revised after the census and to find out whether close results will be obtained.

Obtaining after the census refined data will allow developing better demographic forecasts for the regions of Ukraine. In particular, longer time series are available for them. Therefore, the development of probabilistic forecasts for them, firstly, will give more reliable results, and secondly, preliminary testing of the models on historic data will be possible in order to choose the best model and its parameters.

Promoting probabilistic demographic forecasts by disseminating them alongside with the traditional ones will gradually teach consumers that such forecasts are necessary. It is known that demand does not always shape supply. For example, in the 19th century there was no demand for phones. Only later, when consumers had appreciated all the benefits, new products became items of mass consumption. Dissemination of probabilistic forecasts will help ensure that the

public (media) does not take an exclusively “pessimistic” variant, and decision makers — exclusively “optimistic” one, as is usually the case, but are informed about the degree and bounds of uncertainty. After all, everyone has long been accustomed to the fact that the weather forecast is not formulated deterministically, but within certain limits. A study in Australia [7] showed that the transition of users to the probabilistic forecasts is easy: who needs one variant, can take the mean, and others for their own purposes may consider the probability distribution.

## REFERENCES

1. Keilman, N. (2008). European demographic forecasts have not become more accurate over the past 25 years. *Population and Development Review*, 34(1), 137-153.
2. Khan, H.T.A., W. Lutz (2007). How well did past UN Population Projections anticipate demographic trends in six Southeast Asian countries? *Working paper WP-507*. Oxford Institute of Ageing.
3. Socioeconomic Impact of HIV/AIDS in Ukraine (2006). Washington. The World Bank
4. Keilman, N., Pham, D.Q. & Hetland, A. (2002). Why population forecasts should be probabilistic - illustrated by the case of Norway. *Demographic Research*, Vol. 6, Art. 15, 409-453.
5. Keilman, N. (2018). Probabilistic demographic forecasts. *Vienna Yearbook of Population Research*. Vol. 16, 25-35. <http://dx.doi.org/10.1553/populationyearbook2018s025>
6. *Recommendations on Communicating Population Projections* (2018). Geneva: UN
7. Wilson, T., & Shalley, F. (2019). Subnational population forecasts: Do users want to know about uncertainty? *Demographic Research*, Vol. 41, Art. 13, 367-392. <http://dx.doi.org/10.4054/DemRes.2019.41.13>
8. Lutz, W., Sanderson, W.C., & Scherbov, S. (1998). Expert-based probabilistic population projections. W. Lutz, J. W. Vaupel and D. A. Ahlburg (Eds.), *Frontiers of population forecasting, supplement to population and development review* (139-155).
9. Bijak, J., & Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies* 70(1), 1-19. <http://dx.doi.org/10.1080/00324728.2015.1122826>
10. Raftery, A. E., Lalic, N., & Gerland, P. (2014). Joint probabilistic projection of female and male life expectancy. *Demographic Research* 30(27), 795-822. <http://dx.doi.org/10.4054/DemRes.2014.30.27>
11. R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
12. Sevcikova, H., Raftery, A.E. & Gerland, P. (2013). Bayesian Probabilistic Population Projections: Do It Yourself. *Joint Eurostat/UNECE Work Session on Demographic Projections*, WP 13.2, October. 19 p.
13. Alkema, L., Raftery, A.E., Gerland, P., Clark, S.J., Pelletier, F., Buettner, T. & Heilig, G.K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography*, 48(3), 815-839. <http://dx.doi.org/10.1007/s13524-011-0040-5>
14. Hyndman, R. J., Booth, H. & Yasmeen, F. (2013). Coherent Mortality Forecasting: The Product-Ratio Method with Functional Time Series Models. *Demography*, 50, 261-283. <http://dx.doi.org/10.1007/s13524-012-0145-5>
15. Hyndman, R.J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24, 323-342. <http://dx.doi.org/10.1016/j.ijforecast.2008.02.009>

16. Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51, 4942-4956.
17. *Reproducibility and Replicability in Science*. (2019). National Academies of Sciences, Engineering, and Medicine 2019. Washington, DC: The National Academies Press, 256 p.
18. Hyndman, R. J., with contributions from H. Booth, L. Tickle, J. Maindonald (2019). *Packge demography: forecasting mortality, fertility, migration and population data. R package, Version 1.22*
19. Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87, 659-671.
20. Lee, R. D., & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38, 537-549.
21. Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56, 325-336.
22. Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3), 575-594. <http://dx.doi.org/10.1353/dem.2005.0021>
23. Lee, R. D. (1993). Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. *International Journal of Forecasting*, 9, 187-202.
24. Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software, Volume 27, Iss. 3*, 1-22.
25. Hyndman, R.J., H.L. Shang (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38, 199-221. <http://dx.doi.org/10.1016/j.jkss.2009.06.002>

Стаття надійшла до редакції журналу 22.06.2020

П.Є. Шевчук, канд. екон. наук, старш. наук. співроб.

Інститут демографії та соціальних досліджень ім. М.В. Птухи НАН України

01032, м. Київ, бул. Т. Шевченка, 60

E-mail: pavlo-shevchuk@ukr.net

ORCID 0000-0003-1158-4438

О.В. Єгорченков, канд. техн. наук, доцент

Київський національний університет імені Тараса Шевченка

04116, м. Київ, вул. Б. Гаврилишина, 24

E-mail: alexee@ukr.net

ORCID 0000-0003-1390-5311

#### ЙМОВІРНІСНІ ДЕМОГРАФІЧНІ ПРОГНОЗИ ДЛЯ МЕТРОПОЛІСІВ УКРАЇНИ

Невизначеність є невід'ємною характеристикою демографічних процесів. Ще більшою мірою це стосується майбутнього. Точні детерміністичні прогнози принципово неможливі. Це визначає необхідність кількісної оцінки майбутньої невизначеності. Мета дослідження — розробка ймовірнісних демографічних прогнозів для мегаполісів України й аналіз результатів. Вперше розроблені ймовірнісні демографічні прогнози для деяких міст України. Дослідження проводилось за допомогою методології функціональних даних, яка включає широкий набір демографічних методів і моделей, реалізованих у кількох пакетах мови програмування R. Обрана методологія базується повністю на статистичних даних і не потребує введення додаткових довільних гіпотез. У трьох випадках (а саме щодо народжуваності у Києві, Львові та Харкові)



метод за замовчуванням (*ARIMA*) показав неправдоподібні результати, які могли бути спричинені ненадійними поточними даними. У цих випадках використовувались модель випадкового блукання. Для Одеси обидві моделі дають подібні результати. Цілком можливо, що в цьому місті недооцінка вибулого населення компенсується заниженням числа прибулих, що призводить до реалістичності поточних показників народжуваності (а саме їх знаменників) і, отже, узгодженості результатів прогнозу незалежно від методу. Прогнози смертності відповідають динаміці спостережуваних показників смертності та якості поточних даних. Використана модель екстраполювала тенденції до збільшення тривалості життя для Дніпра та Одеси і стагнацію для інших міст, що може бути спричинене невідповідністю знаменника для повікових коефіцієнтів смертності в останніх. Розрахунок показав, що чисельність населення Дніпра, Львова та Харкова у 2040 році, швидше за все, буде нижчою за рівень, досягнутий на початку 2019 року. Деякі шанси на зростання чисельності населення залишаються в Одесі, а Київ, швидше за все, матиме більшу людність. Віковий розподіл населення в усіх містах у майбутньому виглядає подібно. Кількість людей старше 40 років має найменшу невизначеність. У віці від 20 до 40 років невизначеність значно більша, що є наслідком невизначеності міграції молоді протягом горизонту прогнозу на 2019—2039 роки, оскільки всі ці міста є потужними освітніми центрами та приваблюють студентів. У 2040 р. ті, хто був студентом у 2020 році, досягнуть 40 років і можуть залишитися у великому місті чи виїхати. Невизначеність кількості осіб до 20 років впливає з двох джерел: невизначеності прогнозів народжуваності та невизначеності чисельності репродуктивних когорт, тобто цих 20—40-річних. Після отримання результатів найближчого перепису є необхідним перегляд цих прогнозів.

**Ключові слова:** ймовірнісний прогноз, функціональні дані, невизначеність, велике місто.