Robust Backward Population Projections Made Possible: An Implication from the Theory of Demographic Potential

Dalkhat M. Ediev

Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography

Abstract. Based on formal results for population dynamics under varying fertility and mortality levels, this paper presents a new approach to backward population projection. Unlike other methods in the literature, the method presented here is robust and accurate in both the short and long run. The method and the theory behind it contribute to the knowledge about dynamic populations and may find applications in population modeling and reconstruction. Details of the backward projection method have been published recently in the Int J Forecasting but has not yet been well presented to the demographic audience. Here, I present those results as well as the underlying theory of dynamically stable populations.

For long time, the backward population projection has remained an unsolved problem. Formally speaking, the population projection matrix ($\mathbf{L}$) is singular, and therefore the usual matrix equation of forward population projection $\mathbf{P_{t+1} = LP_t}$, where $\mathbf{P_t}$ is the vector consisting of 1 January population numbers by single years of age, may not be inversed (as there is no $\mathbf{L^{-1}}$ for obtaining $\mathbf{P_t = L^{-1}P_{t+1}}$). An 'easy' way of overcoming this limitation by truncating the projection matrix at the last age at reproduction and then inversing the truncated matrix turned out not to work due to the instability of the results (Keyfitz, 1977).

The fundamental cause of this instability lies in the spectral properties of the population projection matrix (of the renewal operator in the continuous model): while the real eigenvalue (which determines the population intrinsic rate) dominates the spectrum of the forward projection matrix, it is dominated by all other (complex) eigenvalues in the backward projection.

Greville (1968), Greville and Keyfitz (1974) and Keyfitz (1977) proposed using a generalized inverse of $\mathbf{L}$, which (the inverse) has only three non-zero eigenvalues. However, this method only works for projection over short intervals, because the two remaining complex eigenvalues still dominate the real eigenvalue, which should determine the population dynamics in the long run.

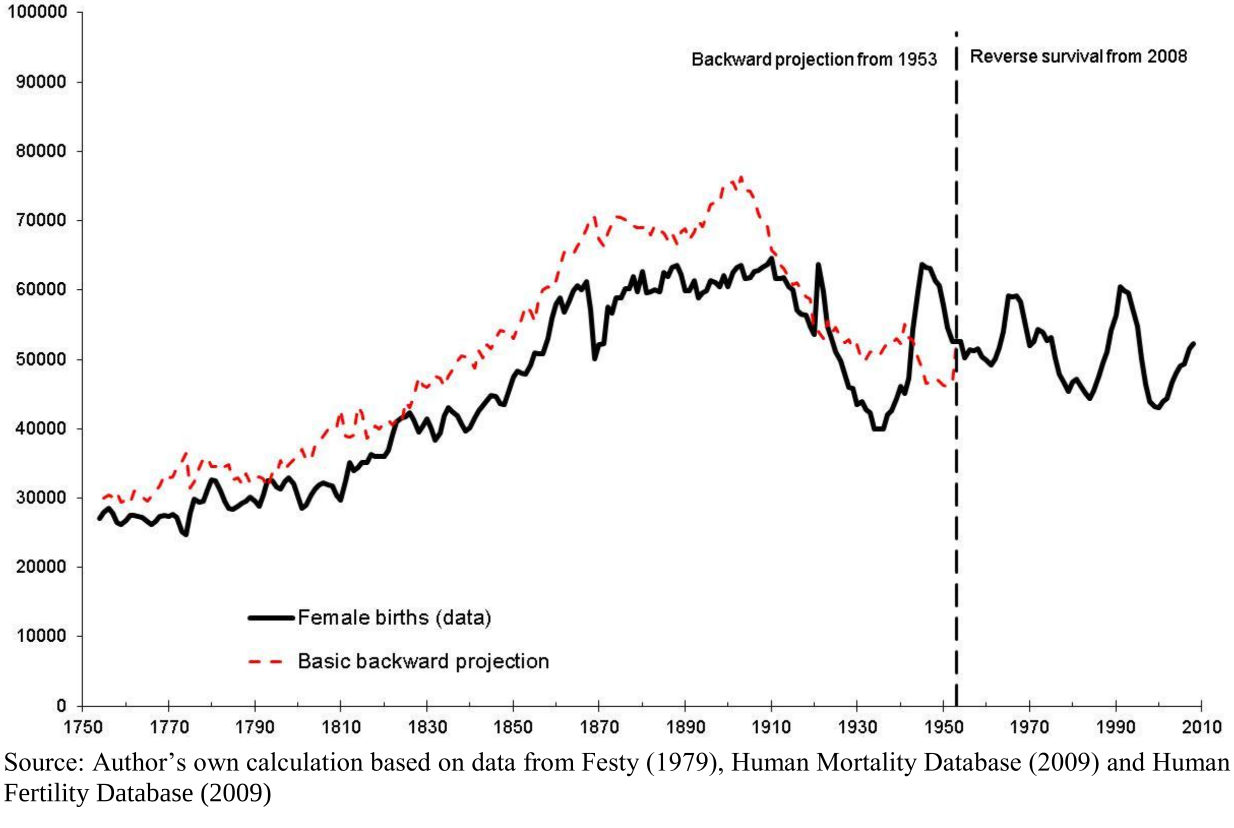
In our approach, we use results from the theory of demographic potentials and of dynamically stable populations. Unlike with the dynamics of births, the demographic potential of the newborn is determined by a renewal equation that is stable when applied to backward, not forward, projections. At the same time, in a class of populations we call *dynamically stable populations* the demographic potential of the newborn and number of births are inversely proportional. This property allows converting the (robust) backward renewal for the potentials into a simple model, *the basic backward projection model*, for the number of births and, hence, of the population:

$$B^{-1}(t) = \int_0^{\beta} l(y,t) f(y,t) B^{-1}(t+y) dy \, , \tag{1}$$

where *B(t)* is the intensity of births, *l(x,t)* and *f(x,t)* – survival and fertility functions of age *x* and birth cohort *t*; β is the upper age limit of childbearing.

The simple model produces rather accurate projections when the dynamic stability assumption is roughly adequate (Fig. 1).

Fig. 1. Backward projection of female births in Sweden from the 1935 population: reverse survival in 1880-1934 and basic backward projection prior to 1880



Source: Author's own calculation based on data from Festy (1979), Human Mortality Database (2009) and Human Fertility Database (2009)

Demographic content of this assumption may be interpreted as stability of timing of childbearing. Although this assumption was used in some studies on historical reconstruction, there may be many situations that violate the dynamic stability assumption. When there are substantial changes in the timing of childbearing, the model above may be further improved by introducing the effect of timing change on the births number (the tempo effect). To this end, we apply tempo-adjustment coefficients *R(t)* to the births' intensity in (1):

$$\left[\frac{B(t)}{R(t)}\right]^{-1} = \int_0^{\beta} l(y,t) f(y,t) \left[\frac{B(t+y)}{R(t+y)}\right]^{-1} dy . \qquad (2)$$

Based on analytical results for dynamical stability and decomposition of general population change into smoothing process and tempo effects (see Appendix), we present three adjustment procedures:

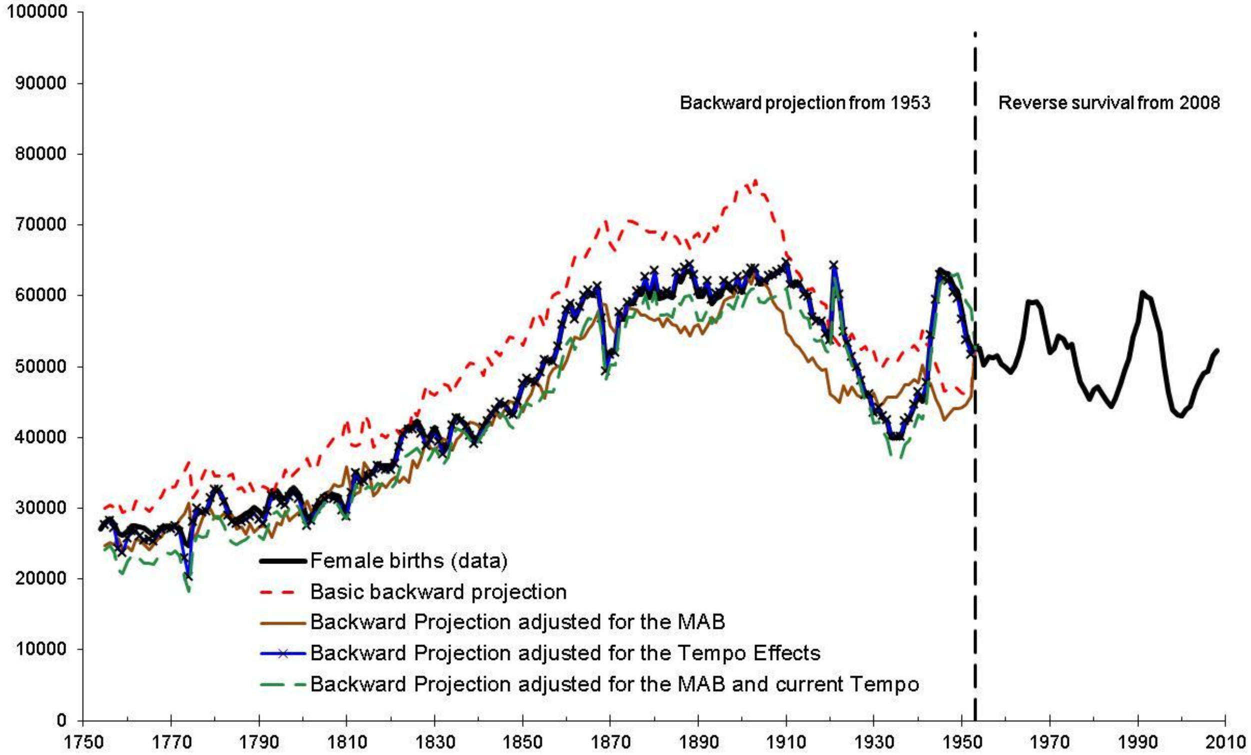
- rough approximate based on the mean age at childbearing, $R(t) = \frac{1}{MAB(t)}$;
- more accurate one calculated from the dynamics of demographic potentials, $$R(t) = \frac{\int_0^{\beta} l(x,t-x) f(x,t-x) R(t-x) dx}{\int_0^{\beta} l(x,t-x) f(x,t-x) dx} - U'(t), \quad R(t<t_0) \equiv 1,$$ where $U(t) = \int_x u(x,t-x)dx$ is the cross-sectional sum of remaining proportions of cohorts' demographic potentials (see the Appendix) and the first summand represents accumulation of past tempo-effects;
- and a method of intermediate complexity where cumulated past tempo-effects are approximated through the change in the mean age at childbearing and supplemented by the current tempo-effect, $R(t) = \frac{1-U'(t)}{MAB(t)}$.

Although approximate, the backward projection models with adjustment for the tempo-effect show good results when applied to real data (Fig. 2).

Further details on backward projection method: Ediev D.M. 2011. Robust backward population projections made possible. *Int J Forecasting* 27, 1241–1247.

Details on the dynamic stability are presented in the Appendix.

Fig. 2. Projection of female births in Sweden from 2008 backwards: basic backward projection and two backward projections adjusted for the deviation from dynamic stability



Source: Author’s own calculation based on data from Festy (1979), Human Mortality Database (2009) and Human Fertility Database (2009)

<u>Conclusion.</u> Apart from being important for the formal demography in showing the very possibility of robust backward projection, our results may have further applications in theoretical and practical studies. The theory of dynamic populations behind our method and the good fit of backward projections based on adjusting for the tempo effects indicates key role of those effects in population dynamics. It seems that, while the overall development in births is determined by general levels of fertility and mortality, all short-term fluctuations (but also long-term shifts in trend due to changing MAB) are caused by tempo effects. This observation, in its turn, may open up possibility to reconstruct the tempo effects and, using them, the historical change in the mean age at childbearing – a task that has not been resolved yet in reconstruction studies (Lee, 1974, 1985; Oeppen, 1993; and Wrigley and Schofield, 1982, assumed time-invariant MAB in their models). These same results may also help simplify population projections forward, which may be conducted by projecting the general (smooth) trends in births and tempo-distortions instead of the conventional components method. These possibilities deserve further study.

Appendix. On the dynamically stable population

1. Introductory results from the theory of demographic potential

Demographic potentials (Ediev, 2003a, 2007a, 2007b, 2009a) are defined through asymptotic relations between direct descent of individuals or sub-populations:

$$\frac{c_A}{c_B} \overset{def}{=} \lim_{t\to\infty} \frac{P_A(t)}{P_B(t)}, \tag{A.1}$$

where c_A and $P_A(t)$ are the demographic potential and population of direct descendants by time *t* for a (sub)population 'A' (that could denote any group of individuals or a single individual); same for the (sub)population 'B'.

The change over time of individual demographic potentials is described by renewal equations resembling those for the population renewal, with reversed time flow, e.g., for the potential of newborn:

$$c_{0,t} = \sum_x \frac{L_{x,t}}{L_{0,t}} F_{x,t} c_{0,t+x+1}\,, \tag{A.2a}$$

$c_{x,t}$ is demographic potential of a person at age x born in year t; $L_{x,t}$ is the life table population for the birth cohort t; and $F_{x,t}$ is the fertility rate at age *x* for the birth cohort t. Or, in continuous form:

$$c(0,t) = \int_0^\beta l(x,t) f(x,t) c(0,t+x)dx\,, \tag{A.2b}$$

here β is the maximum childbearing age. Renewal Eq. (A.2a), Eq. (A.2b) is robust and shares ergodic property similar, in reversed time, to the ergodic property of population renewal (Arthur, 1982) under the same conditions (Ediev, 2007b).

The vector of age-specific demographic potentials $\mathbf{c}_t = \begin{pmatrix} c_{0,t} & c_{1,t-1} & \dots & c_{\omega,t-\omega} \end{pmatrix}^T$ (ω is the lifespan limit; superscript '*T*' denotes transpose) in period *t* is described by matrix relation resembling the population projection equation involving the time-dependent population projection (Leslie) matrix **L**:

$$\mathbf{c}_t^T = \mathbf{c}_{t+1}^T \mathbf{L}_t \, . \tag{A.3}$$

(Similar equation may be derived in the continuous case, using matrix of survival and fertility *operators*, see Ediev, 2007b.)

The total demographic potential of the population closed to migration is time-invariant. In discrete case, that follows immediately from (A.3) and population renewal:

$$C(t) = \mathbf{c}_t^T \mathbf{P}_t = \mathbf{c}_t^T \mathbf{L}_{t-1} \mathbf{P}_{t-1} = \mathbf{c}_{t-1}^T \mathbf{P}_{t-1} = C(t-1), \tag{A.4}$$

where $\mathbf{P}_t = \begin{pmatrix} P_{0,t} & P_{1,t} & \dots & P_{\omega,t} \end{pmatrix} = \mathbf{L}_{t-1}\mathbf{P}_{t-1}$ is the vector of age-specific population numbers at time *t*. Proof for the continuous case may be found in Ediev (2007a, 2007b).

For further analysis it will also be useful to note the following general formal relation between demographic potentials, survival and fertility (Ediev, 2010):

$$F_{x,t} = \frac{c_{x,t}}{c_{0,t+x+1}} - \frac{L_{x+1,t}}{L_{x,t}} \frac{c_{x+1,t}}{c_{0,t+x+1}} = \frac{L_{0,t}}{L_{x,t}} \frac{c_{0,t}}{c_{0,t+x+1}} \left(u_{x,t} - u_{x+1,t} \right), \tag{A.5a}$$

where $u_{x,t} \overset{def}{=} \frac{L_{x,t}}{L_{0,t}} \frac{c_{x,t}}{c_{0,t}}$ is the proportion remaining by age *x* of demographic potential of the birth cohort *t* (alternative interpretation could be: the expected relative future potential of newborn). In the continuous case:

$$f(x,t) = -\frac{1}{l(x,t)} \frac{c(0,t)}{c(0,t+x)} \frac{\partial u(x,t)}{\partial x}, \tag{A.5b}$$

where $u(x,t) \overset{def}{=} l(x,t) \frac{c(x,t)}{c(0,t)}$.

Age pattern of the proportions *u(x)* is one of the most stable characteristics of population reproduction, although it is sensitive to changes in timing of reproduction (Ediev, 2007b, 2009b). The latter property is evident from the link between these quantities and the cohort mean age at reproduction of the demographic potential (from Eq. (A.2a) and Eq. (A.5a)):

$$A_t \overset{def}{=} \frac{\sum_x x \frac{L_{x,t}}{L_{0,t}} F_{x,t} c_{0,t+x+1}}{\sum_x \frac{L_{x,t}}{L_{0,t}} F_{x,t} c_{0,t+x+1}} = \frac{\sum_x x \frac{L_{x,t}}{L_{0,t}} F_{x,t} c_{0,t+x+1}}{c_{0,t}} = \sum_x x\left(u_{x,t} - u_{x+1,t}\right) = \sum_x u_{x,t} \ . \tag{A.6a}$$

Similarly:

$$A(t) \overset{def}{=} \frac{\int_x x l(x,t) f(x,t) c(0,t+x) dx}{\int_x l(x,t) f(x,t) c(0,t+x) dx} = -\int_x x \frac{\partial u(x,t)}{\partial x} dx = \int_x u(x,t) dx \, . \tag{A.6b}$$

2. Dynamically stable population. Decomposition of births into tempo effects and smoothing process.

Time-invariance of the total demographic potential indicates that the scales of population size and individual demographic potentials should be in inverse relation. Indeed, in a stable population, where the annual number of births is an exponential function of time, $B(t) \propto \lambda^t$, λ being the (Lotka's) stable population growth rate, the demographic potential of newborn is exponential function with inversed time, $c_0(t) \propto \lambda^{-t}$, so that the total potential of births in a given year is time invariant. Although the exponential growth may not be extended to general dynamic populations, some important analytical results may be obtained for populations with postulated time-invariance of the total potential of newborns:

$$c_{0,t} P_{0,t} \equiv const \tag{A.7a}$$

or, in continuous form:

$$c(0,t) B(t) \equiv const \, . \tag{A.7b}$$

We term the population satisfying this property the *dynamically stable* population. The population may be dynamically stable either permanently or over a given period of time when Eq. (A.7a) or Eq. (A.7b) applies. An important implication of dynamic stability is possibility of robust backward renewal that follows from Eq. (A.2a) or Eq. (A.2b) (Ediev, 2011):

$$\left(P_{0,t}\right)^{-1}=\sum_x \frac{L_{x,t}}{L_{0,t}} F_{x,t}\left(P_{0,t+x+1}\right)^{-1}, \tag{A.8a}$$

$$B^{-1}(t)=\int_0^{\beta} l(x,t) f(x,t) B^{-1}(t+x) dx. \tag{A.8b}$$

To find out demographic conditions for dynamic stability, the following decomposition of population change may be used that follows from Eq. (A.5a), Eq. (A.5b) and usual population renewal equations:

$$c_{0,t}P_{0,t}=\sum_x c_{0,t-1-x}P_{0,t-1-x}\left(u_{x,t-1-x}-u_{x+1,t-1-x}\right)$$

$$=\sum_y\left(u_{y,t-1-y}-u_{y+1,t-1-y}\right)\frac{\sum_x c_{0,t-1-x}P_{0,t-1-x}\left(u_{x,t-1-x}-u_{x+1,t-1-x}\right)}{\sum_y\left(u_{y,t-1-y}-u_{y+1,t-1-y}\right)}$$

$$=\left(1+\sum_y u_{y,t-1-y}-\sum_y u_{y,t-y}\right)\sum_x c_{0,t-1-x}P_{0,t-1-x}\frac{\left(u_{x,t-1-x}-u_{x+1,t-1-x}\right)}{\sum_y\left(u_{y,t-1-y}-u_{y+1,t-1-y}\right)}$$

$$=\left(1+U_{t-1}^{p}-U_t^{p}\right)\sum_x w_{x,t-1}c_{0,t-1-x}P_{0,t-1-x}, \tag{A.9a}$$

where $w_{x,t-1}=\frac{u_{x,t-1-x}-u_{x+1,t-1-x}}{\sum_y\left(u_{y,t-1-y}-u_{y+1,t-1-y}\right)}$ are non-negative weights that are proportional, in each age, to the net maternity in respective cohort and sum up to unity, they determine smoothing process for the total demographic potential of newborns; that process, however, is continuously distorted by the multiplier before summation in Eq. (A.9a) where quantities $U_t^p=\sum_y u_{y,t-y}$ are related to the timing of reproduction. In continuous model:

$$c(0,t)B(t)=-\int_x c(0,t-x)B(t-x)\frac{\partial u(x,t-x)}{\partial x}dx$$

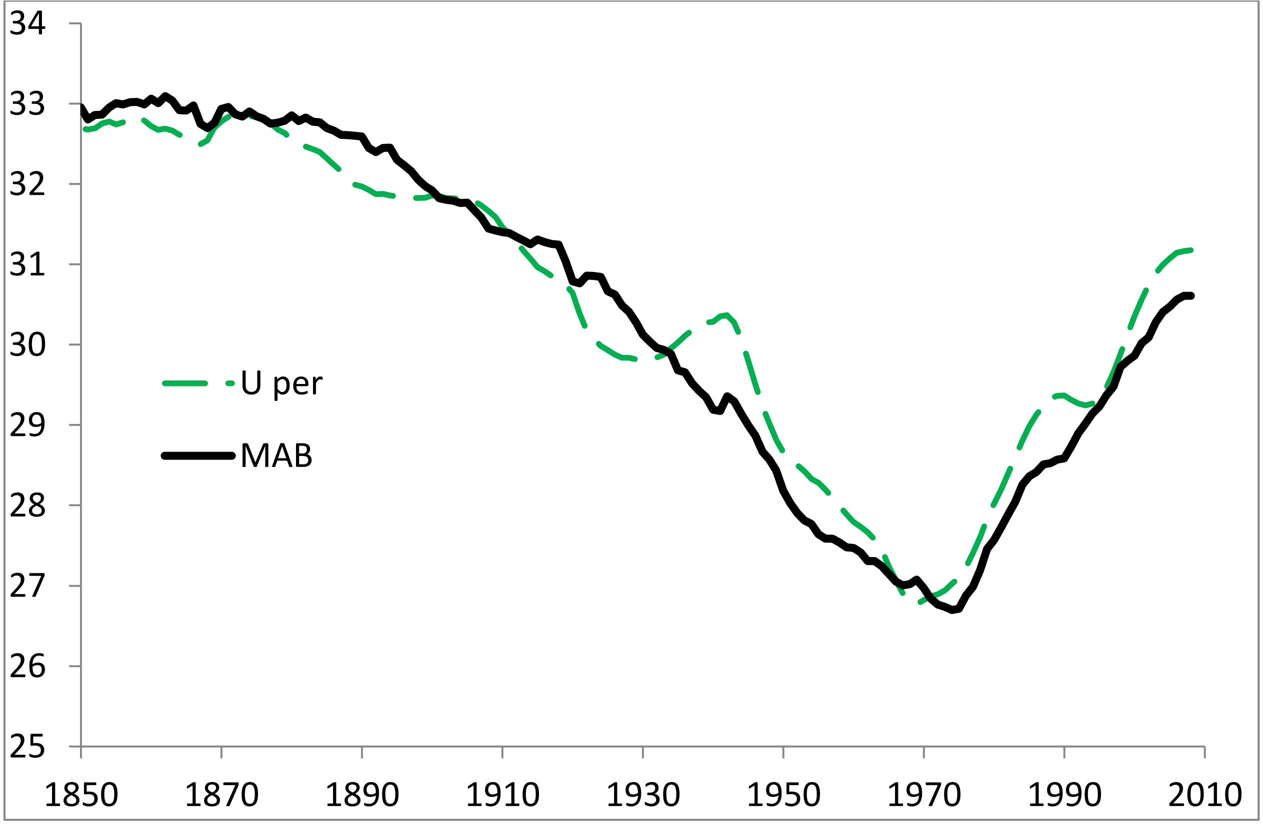
$$=\int_y\left(-\frac{\partial u(y,t-y)}{\partial y}\right)dy\cdot\frac{\int_x c(0,t-x)B(t-x)\left(-\frac{\partial u(x,t-x)}{\partial x}\right)dx}{\int_y\left(-\frac{\partial u(y,t-y)}{\partial y}\right)dy}$$

$$=\left(1-\frac{dU^p(t)}{dt}\right)\cdot\int_x w(x,t-x)c(0,t-x)B(t-x)dx, \qquad \text{(A.9b)}$$

where nonnegative function $w(x,t-x)=\frac{\left(-\frac{\partial u(x,t-x)}{\partial x}\right)}{\int_y\left(-\frac{\partial u(y,t-y)}{\partial y}\right)dy}$ integrates over age to unity and $U^p(t)=\int_y u(y,t-y)dy$.

The multipliers distorting the smoothing process in Eq. (A.9a) and Eq. (A.9b) resemble, not incidentally, the *Timing Index* (*TI*) of Butz and Ward (1979, 1980) and Ryder (1980). Those authors defined the *TI* as a cross-sectional sum, over age, of proportions of cohort completed fertility that fall within the calendar period of interest. When cohorts share a similar timing of fertility (but not necessarily a similar level of completed fertility), the *TI* is time-invariant and equals one. As can be seen from Eq. (A.5a) and Eq. (A.5b) (and also from the very definition of quantities $u_{x,t}$ and $u(x,t)$), decrements $u_{y,t-1-y}-u_{y+1,t-1-y}$ in Eq. (A.9a) or $-\frac{\partial u(y,t-y)}{\partial y}$ in Eq. (A.9b) are proportions of respective cohort's demographic potential reproduced in the calendar period *t*. Summing them up cross-sectionally produces quantity similar to the *TI*. To see that this resemblance is not incidental, note that *TFR* (explained through the *TI*) assumes no mortality and stationary age composition of the synthetic cohort. Under no mortality and stationary population, proportion of cohort fertility and proportion of cohort's demographic potential reproduced at a given age are equal. Despite that the two indices may differ in realistic conditions, we term the multiplier in Eq. (A.9a) and Eq. (A.9b) *the Timing Index (TI)* assuming it refers to the tempo effects in the reproduction of demographic potential. Change in *TI* indicates presence of tempo-effects that distort the smoothing process of the births' total demographic potential. Link of the *TI* to the timing of reproduction may also be evident in the final expression for the multiplier that features the change over time of U_t^p and $U^p(t)$. As we saw earlier (Eq. (A.6a), Eq. (A.6b)), *u(x)* integrated over age in a birth cohorts equal the mean age at reproduction of demographic potential in the cohort. When integrated cross-sectionally, a similar result applies only when cohorts share similar timing of reproduction and *TI* equals one. More generally, U_t^p is in the same relation to the mean ages at reproduction as the relation of the cross-sectional average length of life, CAL (Brouard, 1986; Guillot, 2003), to the life expectancy at birth. Similar to the CAL (Goldstein, 2006), U_t^p may be shown to approximate the mean age at reproduction in the cohort currently at such age. Empirically, the cross-sectional U_t^p corresponds well to the period mean age at childbearing (e.g., Figure A1).

Figure 1. Change over time of the period mean age at childbearing (MAB) and the cross-sectional sum of remaining proportions of cohorts' demographic potentials (U per). Sweden, 1850-2008. (Estimates of U per for the most recent periods are based on assumption of constant future fertility rates.)



Summarizing the description of the decomposition Eq. (A.9a), Eq. (A.9b), the demographic potential of newborns is described by a smoothing process that may, however, be interrupted by tempo effects when *TI* deviates from unity. That would also imply deviation from the dynamical stability, would the population be dynamically stable before the tempo effect occurred. The decomposition also indicates that, once the tempo effects cease to affect the population dynamics, fertility being nonnegative (hence, $u_x(t)-u_{x+1}(t)\geq 0$) and meeting the usual "no-common-divisor" condition of ergodicity (Arthur, 1982), the smoothing process will lead to convergence of the total annual demographic potential of newborns to a constant level and of population to the dynamically stable trend.

When, in particular, the age pattern of reproduction $u_{x,t}\equiv u_x$, $u(x,t)\equiv u(x)$ is time-invariant (this is a generalization to the quadratic hyperstable or metastable model by Kim and Schoen (1996), Schoen and Jonsson (2003), and Schoen (2006), see Ediev (2009b)), the mean age at childbearing observed in the dynamically stable population equals U and the cohort mean age at reproduction of demographic potential Eq. (A.6a), Eq. (A.6b) and is time-invariant:

$$A_t^p=\frac{\sum_x x\frac{L_{x,t-1-x}}{L_{0,t-1-x}}P_{0,t-1-x}F_{x,t-1-x}}{\sum_x \frac{L_{x,t-1-x}}{L_{0,t-1-x}}P_{0,t-1-x}F_{x,t-1-x}}=\frac{\sum_x x\frac{L_{x,t-1-x}}{L_{0,t-1-x}}\frac{1}{c_{0,t-1-x}}F_{x,t-1-x}}{\sum_x \frac{L_{x,t-1-x}}{L_{0,t-1-x}}\frac{1}{c_{0,t-1-x}}F_{x,t-1-x}}=\frac{\sum_x xc_{0,t}\left(u_{x,t-1-x}-u_{x+1,t-1-x}\right)}{\sum_x c_{0,t}\left(u_{x,t-1-x}-u_{x+1,t-1-x}\right)}$$

$$=\sum_x x\left(u_x-u_{x+1}\right)\equiv U\,. \tag{A.10a}$$

$$A^p(t)=\frac{\int_x xl(x,t-x)B(t-x)f(x,t-x)dx}{\int_x l(x,t-x)B(t-x)f(x,t-x)dx}=\frac{\int_x xc(0,t)\left(-\frac{\partial u(x,t-x)}{\partial x}\right)dx}{\int_x c(0,t)\left(-\frac{\partial u(x,t-x)}{\partial x}\right)dx}$$

$$=\int_x x\left(-\frac{\partial u(x)}{\partial x}\right)dx\equiv U\,. \tag{A.10b}$$

Time-invariance of the population total demographic potential allows obtaining a general result for the annual total potential of newborns in the dynamically stable population, where $C=\sum_x \frac{L_{x,t-1-x}}{L_{0,t-1-x}} c_{x,t-1-x} P_{0,t-1-x} = \sum_x c_{0,t-1-x} u_{x,t-1-x} P_{0,t-1-x} = c_{0,t} P_{0,t} \sum_x u_{x,t-1-x} = c_{0,t} P_{0,t} U$, that is:

$$c_{0,t} P_{0,t} = \frac{C}{U}. \tag{A.11a}$$

Continuous case: $C=\int_x l(x,t-x)c(x,t-x)B(t-x)dx = \int_x c(0,t-x)u(x,t-x)B(t-x)dx = c(0,t)B(t)U$, that is:

$$c(0,t)B(t)=\frac{C}{U}. \tag{A.11b}$$

For an initially not dynamically stable population, that would be the asymptote in the absence of new tempo effects. These relations show that it is possible to extend to the dynamically stable case the more narrow result that the births in the stable population equal the total reproductive value ($V(t)=C/c(0,t)$, Ediev, 2007a, 2009a) divided by the mean age at childbearing in the population as well as the related results for the population momentum.

References

Arthur, W.B., 1982. The ergodic theorems of demography: a simple proof. *Demography* 19, 439-445.

Brouard, N., 1986. Structure et dynamique des populations. La pyramide des années à vivre, aspects nationaux et exemples régionaux [Structure and dynamics of population: the pyramid of living years, national aspects and regional examples], *Espaces, Populations, Sociétés* 2(14-15), 157-168.

Butz, W.P. and Ward, M.P., 1979. Will U.S. fertility remain low? A new economic interpretation. *Population and Development Review* 5, 663-688.

Butz, W.P. and Ward, M.P., 1980. Completed Fertility and Its Timing. *Journal of Political Economy* 88(5), 917-940.

Ediev, D.M., 2003. The Concept of Demographic Potential and Its Applications. *Matematicheskoye Modelirovanie (Mathematical modeling)* 15(12), 37-74.

Ediev, D.M., 2007a. On an extension of R.A. Fisher's result on the dynamics of the reproductive value. *Theoretical Population Biology* 72(4), 480-484. (doi:10.1016/j.tpb.2007.03.001)

Ediev, D.M., 2007b. Demographic potentials: theory and applications. Max-PRESS, Moscow.

Ediev, D.M., 2009a. On the definition of the reproductive value: Response to the discussion by Bacaër and Abdurahman. *Journal of Mathematical Biology* 59(5), 651-657.

Ediev, D.M., 2009b. On the reproductive value and the spectrum of population projection matrix with implications for dynamic population models. *Theoretical Population Biology* 78(2), 67-70.

Ediev D.M. 2011. Robust backward population projections made possible. *International Journal of Forecasting* 27, 1241–1247.

Goldstein, J.R., 2006. Found in Translation? A cohort perspective on tempo-adjusted life expectancy. *Demographic Research* 14(5), 71-84.

Guillot, M., 2003. The cross-sectional average length of life: A cross-sectional mortality measure that reflects the experience of cohorts, *Population Studies* 57(1), 41-54.

Kim, Y.J., Schoen, R., 1996. Populations with quadratic exponential growth. *Mathematical Population Studies* 6, 19-33.

Ryder, N.B., 1980. Components of temporal variations in American fertility. In: R.W. Hiorns (ed), *Demographic Patterns in Developed Societies*. London: Taylor and Francis, 15-54.

Schoen, R., 2006. *Dynamic Population Models*. Springer, Dordrecht.

Schoen, R., Jonsson, S.H., 2003. Modeling Momentum in Gradual Demographic Transitions. *Demography* 40(4), 621-635.

Lee, R.D., 1974. Estimating series of vital rates and structures from baptisms and burials: a new technique, with application to pre-industrial England. *Population Studies* 28, 495-512.

Lee, R.D., 1985. Inverse projection and back projection: a critical appraisal and comparative results for England 1539-1871. *Population Studies* 39, 233-248.

Oeppen, J., 1993. Back Projection and Inverse Projection: Members of a Wider Class of Constrained Projection Models. *Population Studies* 47, 245-267.

Wrigley, E.A., and Schofield, R.S., 1982. *The Population History of England, 1541-1871*. Edward Arnold and Harvard University Presses.