"The new Interpolated Markov Chain software (IMaCh 0.99) - backward prevalence from Italian SILC and French HID surveys - time varying covariates from the American HRS survey"

REVES 2018. Univ. Michigan

Nicolas Brouard Yao-Chi Shih Cristina Giudici

French National Institute of Population Studies (France), National Chung Cheng University (Taiwan), Sapienza University (Italy)

REVES 2018, Wednesday, 30th May, Session 3. Developments in measures and methods , part 1

- Imach-064b (XIth REVES London 1999), Imach-096d (**Agnès Lièvre**'s thesis 2004), IMach 0.98q0 (2014) are numbers which sound "familiar" to members of the REVES microcosm since 1992 (REVES 1st, London)
- How will sound IMaCh 0.99? What's new in version 0.99? 3 new parameters:
  - Fixed Quantitative covariates (Years of education, attendance to religious offices etc), `nqv=`
  - Time varying Dummy covariates (`ntv=`)
  - `Time varying Quantitative covariates (BMI, U shape?) (nqtv=)`

  1 new computation: *backward prevalences*, not presented yesterday at the Preconference workshop.
- Demographic analysis
- Cross-sectional measurement of Duration of Life: C.A.L (Brouard, 1986)
- Prevalences: backward, cross-sectional, forward.

A *backward probability* can be defined as the probability to act today while being conditioned by an event which will occur in the future. A classical probability or forward probability is for example: the probability to have a child two years after marriage, but a demographer might be interested in studying fertilty by years before separation or divorce, because "probable" separation could explain lower fertility. *Forward probabilities* in Demography, Economics and Statistics are common and powerful.

Backward probabilities are clearly of less interest than classical probability but are they completely useless?

If computer programs allow us to look at backward probabilities and backward prevalences at the same cost than forward probabilities and forward prevalences, let us try the new IMaCh 0.99 `backcast=1` parameter line:

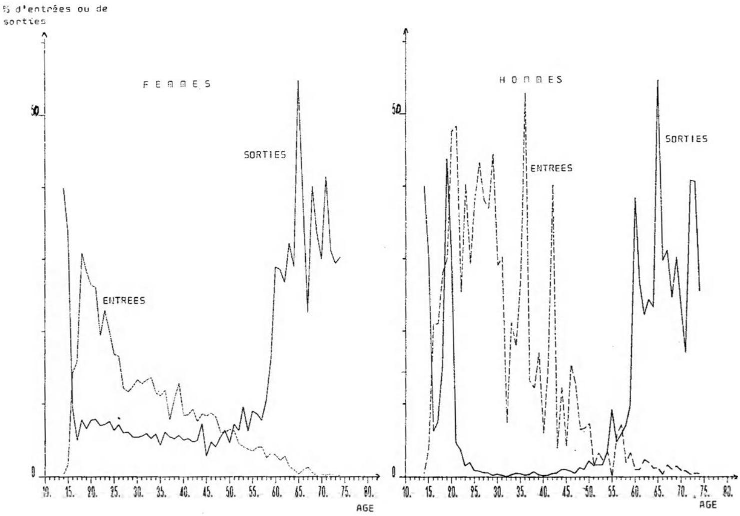
```
prevforecast=1 starting-proj-date=1/1/1998 final-proj-date=1/1/2020 mobil_average=0

backcast=1 starting-back-date=1/1/1998 final-back-date=1/1/1980 mobil_average=1
```

First chained Labor Force Surveys in France, 1977-1978, from which age specific flows (input/output matrices) can be extracted.

Age $x+1$ 1978	ACTIVE	INACTIVE	*Total*
Age x 1977			
ACTIVE	N_{11}	N_{12}	$N_{1.}$
INACTIVE	N_{21}	N_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	$N_{..}$

We can compute the age specific probability of exiting the labor force in one year $\hat{c}_x = \frac{N_{12}}{N_{1.}}$ as well as [re]entering the labor force $\hat{a}_x = \frac{N_{21}}{N_{2.}}$
And by multiplying the age specific matrices P_x, we can compute the probability to be out of the labor force after n years for somebody in the labor force at age x as well as the probability to be in the labor force after n years for somebody out of the labor force at age x, etc:

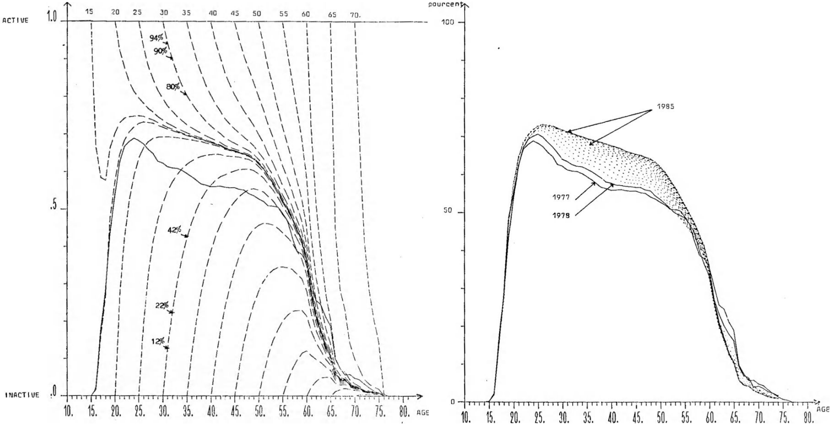
$$P_x = \begin{pmatrix} 1-\hat{c}_x & \hat{c}_x \\ \hat{a}_x & 1-\hat{a}_x \end{pmatrix} \qquad {}_nP_x = P_x P_{x+1} \cdots P_{x+n-1}$$



Probability to enter or exit the labor force in France (Labor force survey. 1977-1978. INSEE) by age and sex.

$$
\begin{array}{ccc} {}_2P_{30} & = P_{30} & P_{31} \end{array}
$$

$$
\begin{pmatrix} 90\% & 10\% \\ 22\% & 78\% \end{pmatrix} = \begin{pmatrix} 94\% & 6\% \\ 12\% & 88\% \end{pmatrix} \begin{pmatrix} 95\% & 5\% \\ 12.04\% & 87.96\% \end{pmatrix}
$$



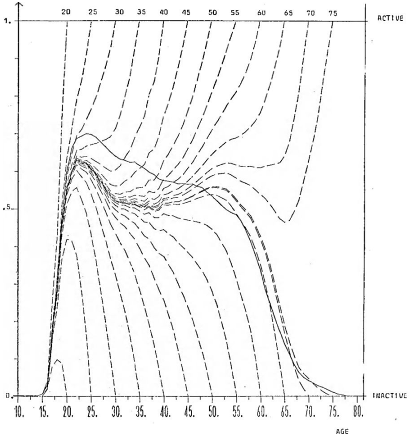
Cross-sectional vs period (stable) participation ratios. Forward projection of female participation rates (based on flows between two French labor force surveys 1977-78). Children do no more force women to leave the labor force. Weak ergodicity can then be observed by right multiplying matrices, inducing a convergence.

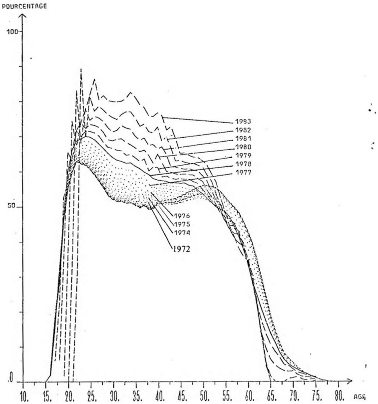
Age x 1978 / Age $x-1$ 1977	ACTIVE	INACTIVE	*Total*
ACTIVE	N_{11}	N_{12}	$N_{1.}$
INACTIVE	N_{21}	N_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	$N_{..}$

Now if we compute *backward* probability $b_x^{12} = \frac{N_{12}}{N_{.2}}$ defined as the probability to be active at age $x-1$ knowing that we will be inactive at age x and $b_x^{21} = \frac{N_{21}}{N_{.1}}$ as the probability to be inactive at age $x-1$ knowing that will be active at age x, we can make backward projections:

$$B_x = \begin{pmatrix} 1-b_x^{21} & b_x^{12} \\ b_x^{21} & 1-b_x^{12} \end{pmatrix} \quad {}_nB'_x = B'_x B'_{x-1} \cdots B'_{x-n+1}$$

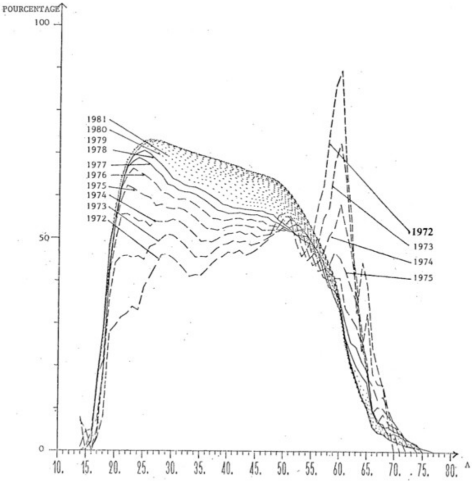
and because of weak ergodicity we get convergence into the past.





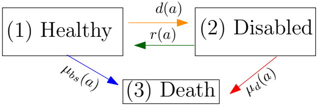
Backward projection of female participation rates (France 1977-78).
Stable curve roughly corresponds to the situation before the 1968th wave.

Please, notice that backward convergence into the past is very different from the so-called *retroprojection* method. Retroprojection consists in the inversion of the projection matrix (for example Leslie matrices). Therefore the second and higher order eigen values are higher than 1 and the back projection is diverging.

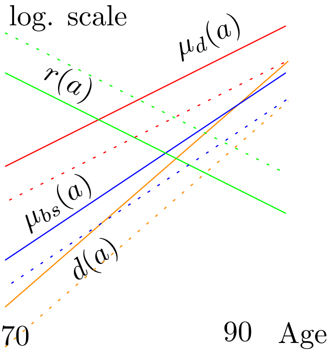


Backward probability with differential mortality

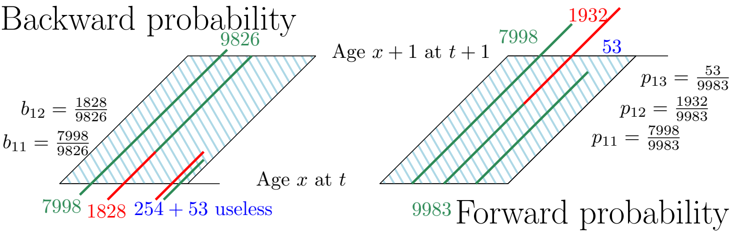
- More live states, 2 or 3 (healthy, mild disability, severe disability)



- Absorbing states like Death



- Differential mortality



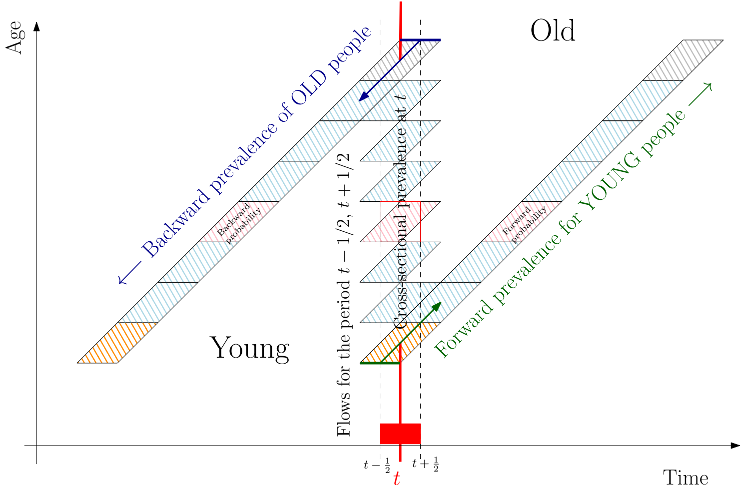
Lexis diagram of forward and backward approaches with DIFFERENTIAL MORTALITY. Green for Healthy, red for Disabled, blue for Death.

		Hthy	*Dis.*	*Dead*
Healthy	9983	7998	1932	53
Disabled	6607	1828	4525	254
Total	16590	9826	6457	307

The classical forward approach of a Lexis diagram is on the right of the figure, showing the future of a cohort of healthy people at age x with their corresponding statuses at age $x+1$: still healthy, disabled and dead. In the backward approach, we are looking at the probability to be healthy a year before, knowing the status this year.

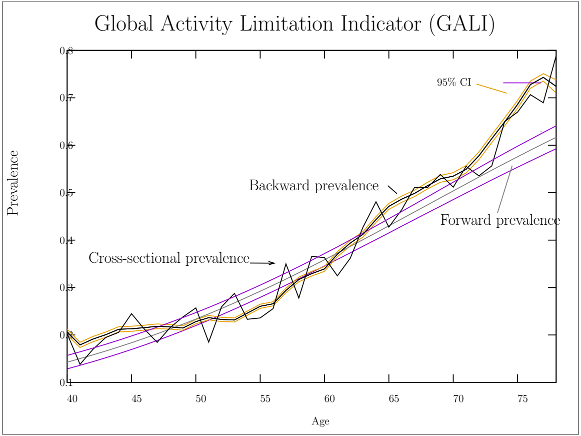
The probability to be healthy today knowing that we will be healthy next year is different from the probability to be healthy next year knowing that we are healthy today.

Demographic analysis and Lexis Diagram

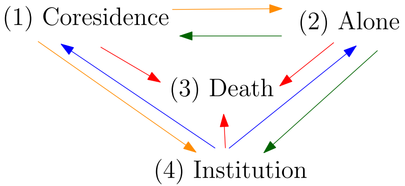


Did the 2008 crisis in Italy impact the Health Expectancies?

Results from 2013 longitudinal release of the Italian "Statistics on Income and Living Condition" survey (EU-SILC) (Giudici et al., 2017) shows that the cross-sectional, forward as well as backward prevalences of disability measured either by the Gali or self-rated health or chronic diseases are almost identical, proving that the crisis hasn't had a major impact on health expectancies. Also, younger generation will be less disabled at very old ages.

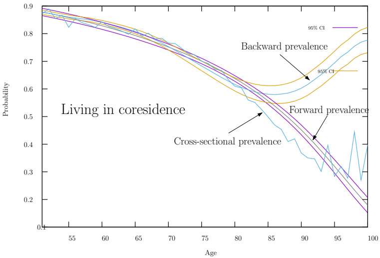


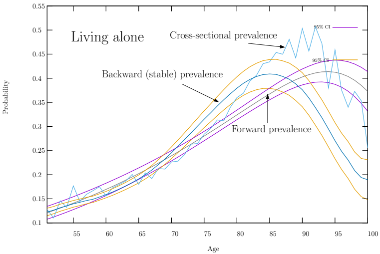
Change in living arrangement: the HR Study

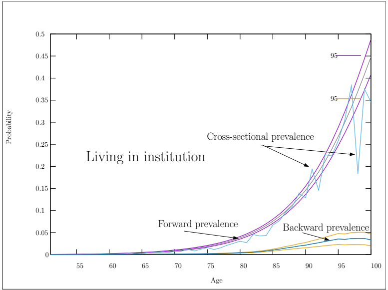


- Biennial Living Arrangements 1998 to 2012, US HRS ;
- Three live states: 1 co-residence, 2 alone, 3 in institutions.
- The 4th state is death.
- 8 waves;

We explored thed properties of this proposed index using the latest US Health and Retirement Study with eight waves (1998-2012). Also, we are interested not only in Health Expectancy but in how the living arrangements are changing in relation with the disability statuses while a person is aging(Shih, 2016).



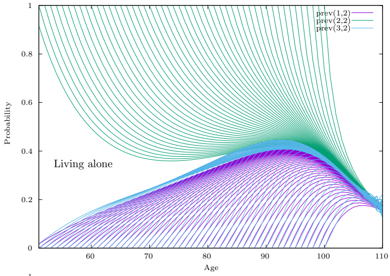


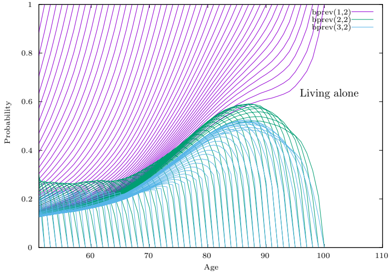


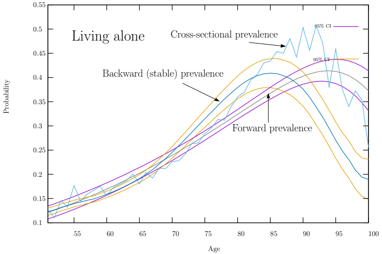
If cross-sectional analysis already shows a move from coresidence to living alone and institutions with the ageing process. We can postulate, for example, that living alone will be postponed because of the decline in mortality. Institutionalization and coresidence will not move dramatically. Finer analyses, controlling by disability status, are forecasting increases in institutionalization!

Confidence intervals for backward prevalences are surprisingly

very small.

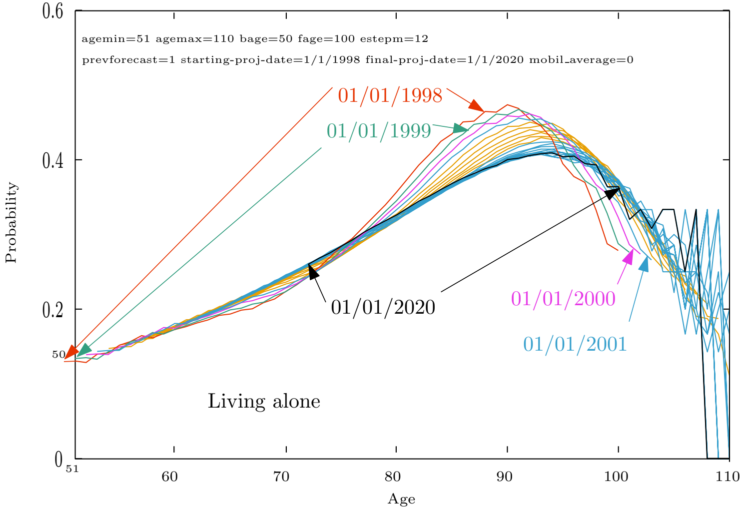






Focus on Living alone shift:

- Convergence to forward and backward prevalence of living alone (HRS).
- Shift to older ages (backward, cross-sectional, forward).



(Forward) Projection of prevalence of living alone from 1998 to 2020 (HRS)

Adding time varying covariates to the model: influence of disability

$$\operatorname{logit} p_x^{ij} = a^{ij} + b^{ij} * age + c^{ij} * Disabled \quad (1)$$

$$model = 1 + age + V_3 \quad (2)$$

V_3 being a time varying dummy variable (IADL disabled)

	If non disabled at each wave, expected years to be spent in:			
Status at 65	Coresidence	Alone	Institution	Total
Coresidence	14.9	4.5	0.7	20.0
Alone	6.2	12.3	0.8	19.4
Institution	5.4	3.0	3.4	12

	If disabled at each wave, expected years to be spent in			
Status at 65	Coresidence	Alone	Institution	Total
Coresidence	8.9	1.5	0.8	11.25
Alone	3.8	6.7	0.94	11.40
Institution	1.08	0.43	3.93	5.44

Disability reduces life expectancy by half. At age 65, people initially in coresidence have almost the same life expectancy than people initially living alone: 20 years if non disabled and 11 if disabled. A person institutionalized at age 65 will spent 5.4 years in coresidence, 3 years alone and 3.4 years in an institution if he or she is non disabled; but if he/she is disabled he or she will survive only 5.4 years, mostly in an institution (4 years).

Limitations and conclusion

Estimations of forward as well as backward prevalences over an age range, for example [70:100], requires an extrapolation of the the logistic model over a much broader age range, [40:120], in order to get convergence. Even if statistical confidence intervals are widening on one or another side, we are still considering the validity of these extrapolations.
Concept of time varying covariates: disabled at the beginning of each wave?

- Demographic analysis: backward, cross-sectional, forward prevalences
- Examples of stationarity (Italy-SILC), important shifts in "Living alone" US-HRS, importance of disability in living arrangements movements and durations in each state.
- Interpolation of Markov Chain programs (like IMaCh 0.99) are mandatory for estimating chained data from modern cross-longitudinal surveys with many waves and missing interviews.
- Move the C sources from INED CVS to Github. Changes since versions 0.99r14 of June 2017?: http://sauvy.ined.fr/cgi-bin/cvsweb.cgi/imach/src/imach.c.diff?r1=1.276;r2=1.288;f=h

References I

Brouard, N. (1980). Espérance de vie active, reprises d'activité féminine : un modèle. *Revue économique*, 31(6):1260–1287. (special issue on Economy and Demography).

Brouard, N. (1986). Structure et dynamique des populations. la pyramide des années à vivre, aspect nationaux et exemples régionaux. *Espace Populations Sociétés*, (2):157–168. (transl. by T. Riffe in 2018 on SocArxiv).

Giudici, C., Brouard, N., Coppola, L., and Trappolini, E. (2017). Disability trends among italian adult population: a multistate approach. In *IUSSP XXVII Cape Town*. IUSSP. poster 56.

Hoem, J. and Fong, M. (1976). *A Markov Chain Model of Working Life Tables. Report No. 1. A new method for the construction of tables of working life.* Laboratory of Actuarial Mathematics. University of Copenhagen, Denmark.

Shih, Y.-C. (2016). *Living alone and subsequent living arrangement transitions among older Americans*. PhD thesis, University of Massachusetts Boston, Boston, MA.

Limitations and conclusion

Thank you for your attention!

Appendix

$$\begin{bmatrix} N_{1.} & 0 & 0 \\ 0 & N_{2.} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{matrix} N_{1.} \\ N_{2.} \\ N_{..} \end{matrix} \quad \begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \end{bmatrix} \\ \quad\quad N_{.1} \quad N_{.2} \quad N_{.3}$$

$$\begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} w_1 p_{11} & w_1 p_{12} & w_1 p_{13} \\ w_2 p_{21} & w_2 p_{22} & w_2 p_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{N_{1.}}{N_{..}} \frac{N_{11}}{N1.} & \frac{N_{1.}}{N_{..}} \frac{N_{12}}{N1.} & w_1 p_{13} \\ w_2 p_{21} & w_2 p_{22} & w_2 p_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \end{bmatrix} / N_{..}$$

$$b_{ij} = \frac{N_{ij}}{N.j} = \frac{w_i p_{ij}}{\sum_i w_i p_{ij}} = \frac{N_{i.}}{N_{..}} \frac{N_{ij}}{N_{i.}} / \sum_i$$

and then we multiply the matrices to get the backward prevalence limit!!!!

$$B_{x+1} = \begin{bmatrix} \frac{w_1 p_{11}}{w_1 p_{11} + w_2 p_{21}} & \frac{w_1 p_{12}}{w_1 p_{12} + w_2 p_{22}} \\ \frac{w_2 p_{21}}{w_1 p_{11} + w_2 p_{21}} & \frac{w_2 p_{22}}{w_1 p_{12} + w_2 p_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{w_1 p_{11} + w_2 p_{21}} & 0 & 0 \\ 0 & \frac{1}{w_1 p_{12} + w_2 p_{22}} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_{x+1} = \mathrm{Diag}(w_x) P_x \, \mathrm{Diag}(\sum_i w_x^i p_x^{ij})$$

$$ {}_{n}B_x = B_{x-(n-1)} B_{x-(n-2)} \cdots B_{x-1} B_x $$

to be compared with

$$ {}_{n}P_x = P_x P_{x+1} \cdots P_{x+n-1} $$

In order to higlight the period prevalence at exact age x as the limit when $n \to \infty$, it can be rewritten

$$ \begin{aligned} {}_{n}P_{x-n} &= P_{x-n} P_{x-(n-1)} \cdots P_{x-1} \\ &= P_{x-n} \cdot {}_{n-1}P_{x-(n-1)} \end{aligned} $$