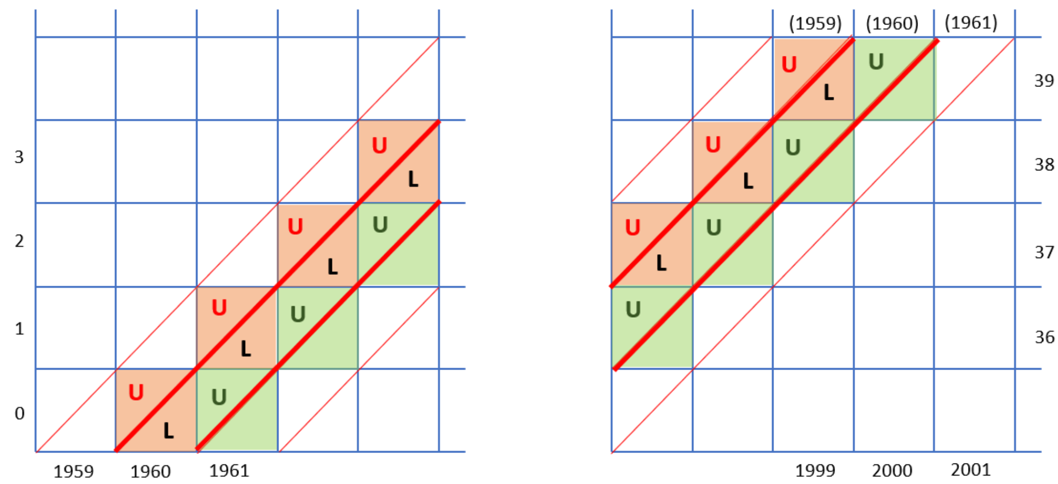
Figure 1: Lexis diagram for the 1960 birth cohort of women



Suppose that mortality is zero, and that there are N_{1960} women in the 1960 birth cohort. Define L_{xt} and U_{xt} as the number of births that happen in the lower and upper triangles, respectively, of the age-period square for age x, year t. The total births to the 1960 cohort before exact age 40 would then be

$$\begin{aligned} B_{1960} &= (L_{0,1960} + L_{1,1961} + \cdots + L_{39,1999}) \\ &\quad + (U_{0,1961} + U_{1,1962} + \cdots + U_{39,2000}) \\ &= L^*_{1960} + U^*_{1960} \end{aligned}$$

where we divide the births by the category of Lexis triangle that they fall into. With no mortality, true cohort completed fertility per woman would then be

$$F_{1960} = \frac{L^*_{1960} + U^*_{1960}}{N_{1960}}$$

If we used the orange age-period squares to approximate the fertility history there would be

$$\begin{aligned} B &= (L_{0,1960} + L_{1,1961} + \cdots + L_{39,1999}) + (U_{0,1960} + U_{1,1961} + \cdots + U_{39,1999}) \\ &= L^*_{1960} + U^*_{1959} \end{aligned}$$

and estimated completed fertility would be lifetime events / cohort size:

$$\hat{F}_{1960} = \frac{U^*_{1959} + L^*_{1960}}{\frac{1}{2}(N_{1959} + N_{1960})}$$

Simplifying Approximations

In order to develop a simple expression for "AP bias", assume that

1. Mortality is negligible below the highest age of childbearing

2. In a small neighborhood around our focal cohort c, cohort sizes change multiplicatively: $\frac{N_c}{N_{c-1}} = \frac{N_{c+1}}{N_c} = K$

3. In a small neighborhood around our focal cohort c, completed fertility changes additively: $(F_c - F_{c-1}) = (F_{c+1} - F_c) = \Delta$

4. Half of a cohort's lifetime births happen in lower Lexis triangles, half in upper triangles: $L_c^* = U_c^* \approx \frac{1}{2} N_c F_c$

Under these assumptions the "diagonal AP" estimator for completed fertility of the cohort born in year c is

$$\hat{F}_c \approx \frac{\frac{1}{2} N_{c-1} F_{c-1} + \frac{1}{2} N_c F_c}{\frac{1}{2}(N_{c-1} + N_c)}$$

If cohort sizes around c are changing multiplicatively with annual multiplier K, this implies that

$$\begin{aligned}
\hat{F}_c &\approx \frac{\frac{1}{2} N_{c-1} F_{c-1} + \frac{1}{2} K N_{c-1} F_c}{\frac{1}{2}(N_{c-1} + K \cdot N_{c-1})} \\
&= \frac{F_{c-1} + K F_c}{K+1} \\
&= \left(\tfrac{1}{K+1}\right) F_{c-1} + \left(\tfrac{K}{K+1}\right) F_c
\end{aligned}$$

That is, the AP diagonal estimator is is approximately equal to a weighted average of the previous and current cohort's completed fertility. The approximate additive error from using the AP diagonal estimator is therefore

$$\hat{F}_c - F_c \;=\; \left(\frac{1}{K+1}\right)(F_{c-1} - F_c) \;=\; -\left(\frac{\Delta}{K+1}\right)$$

and the approximate proportional error is

$$\frac{\hat{F}_c - F_c}{F_c} \;=\; -\left(\frac{1}{K+1}\right)\left(\frac{\Delta}{F_c}\right)$$

From these relationships we learn that

- bias from using the AP diagonal estimator is negative when completed fertility is rising with cohort birth year ($\Delta > 0$),and negative wieh completed fertility is falling
- bias from using the AP diagonal estimator is smaller when cohort size is increasing more quickly between consecutive birth years (larger K implies additive and relative bias closer to zero)

averaging AP diagonals

If we try estimating F_{1960} by averaging $\hat{F}_{1960}$ and $\hat{F}_{1961}$ we get **[assume that cohort size ratios stay constant]**

$$\begin{aligned}\tilde{F}_{1960} = &\ \tfrac{1}{2}\left[\left(\tfrac{1}{K+1}\right)F_{1959} + \left(\tfrac{K}{K+1}\right)F_{1960}\right] \\ &+ \tfrac{1}{2}\left[\left(\tfrac{1}{K+1}\right)F_{1960} + \left(\tfrac{K}{K+1}\right)F_{1961}\right] \\ =&\ \tfrac{1}{2}F_{1960} + \tfrac{1}{2}\left[\left(\tfrac{1}{K+1}\right)F_{1959} + \left(\tfrac{K}{K+1}\right)F_{1961}\right]\end{aligned}$$

If change in F_c across cohorts is linear, $F_c - F_{c-1} = \Delta$, then

$$\begin{aligned}\tilde{F}_{1960} =&\ \tfrac{1}{2}F_{1960} + \tfrac{1}{2}\left[\left(\tfrac{1}{K+1}\right)(F_{1960}-\Delta) + \left(\tfrac{K}{K+1}\right)(F_{1960}+\Delta)\right] \\ =&\ F_{1960} + \tfrac{1}{2}\left(\tfrac{K-1}{K+1}\right)\Delta\end{aligned}$$

old derivation

So the error from summing along the diagonal of the age-period squares is

$$\hat{F}_{1960} - F_{1960} = \frac{L^*_{1960} + U^*_{1959}}{\frac{1}{2}(N_{1959}+N_{1960})} - \frac{L^*_{1960} + U^*_{1960}}{N_{1960}}$$

Defining

$$K = \frac{N_{1959}}{N_{1960}}$$

as the size of the 1959 cohort relative to the 1960 cohort, we can rewrite the error as

$$
\begin{aligned}
\hat{F}_{1960} - F_{1960} &= \frac{L^*_{1960} + U^*_{1959}}{\frac{1}{2}(K \cdot N_{1960} + N_{1960})} - \frac{L^*_{1960} + U^*_{1960}}{N_{1960}} \\
&\approx \frac{\frac{1}{2}F_{1960} \cdot N_{1960} + \frac{1}{2}F_{1959} \cdot K \cdot N_{1960}}{\frac{1}{2}(K \cdot N_{1960} + N_{1960})} - F_{1960} \\
&= \frac{F_{1960} + F_{1959} \cdot K}{K+1} - F_{1960} \\
&= F_{1960}\left(\frac{1 + \frac{F_{1959}}{F_{1960}} \cdot K}{K+1} - 1\right) \\
&= F_{1960}\left(\frac{1 + \frac{F_{1959}}{F_{1960}} \cdot K}{K+1} - \frac{K+1}{K+1}\right) \\
&= F_{1960}\left(\frac{K \cdot \left(\frac{F_{1959}}{F_{1960}} - 1\right)}{K+1}\right) \\
&= F_{1960}\left(\frac{K \cdot}{K+1}\right)\left(\frac{F_{1959}}{F_{1960}} - 1\right) \\
&= \left(\frac{K \cdot}{K+1}\right)(F_{1959} - F_{1960})
\end{aligned}
$$

$$
\begin{aligned}
\hat{F}_{1960} - F_{1960} &= \frac{L^*_{1960} + U^*_{1959}}{\frac{1}{2}(K \cdot N_{1960} + N_{1960})} - \frac{L^*_{1960} + U^*_{1960}}{N_{1960}} \\
&= \frac{1}{N_{1960}}\left[\frac{L^*_{1960} + U^*_{1959}}{\frac{1}{2}(K+1)} - (L^*_{1960} + U^*_{1960})\right] \\
&= \frac{1}{N_{1960}}\left[\left(\frac{L^*_{1960}}{\frac{1}{2}(K+1)} - L^*_{1960}\right) + \left(\frac{U^*_{1959}}{\frac{1}{2}(K+1)} - U^*_{1960}\right)\right] \\
&= \frac{1}{N_{1960}}\left[\left(\frac{L^*_{1960}}{\frac{1}{2}(K+1)} - L^*_{1960}\right) + \left(\frac{U^*_{1959}}{\frac{1}{2}(K+1)} - U^*_{1959}\right) + (U^*_{1959} - U^*_{1960})\right] \\
&= \frac{1}{N_{1960}}\left[\left(\frac{1}{\frac{1}{2}(K+1)} - 1\right)(L^*_{1960} + U^*_{1959}) + (U^*_{1959} - U^*_{1960})\right] \\
&= \frac{1}{N_{1960}}\left[\underbrace{\left(\frac{1-K}{1+K}\right)(L^*_{1960} + U^*_{1959})}_{\text{cohort size effect}} + \underbrace{(U^*_{1959} - U^*_{1960})}_{\text{rate trend effect}}\right]
\end{aligned}
$$

where the first term in square brackets is a cohort size effect, and the second is a rate-change effect.

If $K = 1$ the error simplifies to $\hat{F}_{1960} - F_{1960} = \frac{U^*_{1959} - U^*_{1960}}{N_{1960}}$.