Fitting Demographics Parameters with Polynomials

1 Mortality Hazard Rates

We have data on mortality hazard rates from age 68 to age 90 for past, present, and future years.

We will fit these to the following univariate polynomial function.

$$\ln \rho_{ist} = \sum_{j=1}^{J} \beta_{ji}^{s} \left(\frac{s}{100}\right)^{j} + \beta_{ji}^{t} \left(\frac{t}{100}\right)^{j} \tag{1.1}$$

where i indexes the country, s is the age, and t is the year (an integer value that could be positive or negative, zero corresponding to the base year). Note that even though we have data only up to age 90, we are assuming that agents live until age 100 now. Mortality rates beyond age 90 will be inferred from the polynomial fit.

We fit each country separately in different regressions, but use a panel dataset for each country over age and year.

The vector of beta coefficients, B_i, can be estimated using OLS.

$$B_i = (X'X)^{-1}X'Y_i \tag{1.2}$$

where X is an $23 \times (J+1)$ matrix of ages and time periods raised to various powers, and Y_i is an $23T \times 1$ vector of the natural log of mortality hazard rates for country i for all ages and years.

B_i will thus be a $2(J+1)$ vector of coefficients with the first $J+1$ terms corresponding to the β^s_{ji} coefficients and the last $J+1$ terms corresponding to the β^t_{ji} coefficients

We fit these polynomials and save only the regression coefficients, the B_i's, to pass to the program.

In our Python program, to generate mortality hazard rates for agents that live for S periods we use the regression equation above replacing 100 with S. Note this gives us the one-year hazard rate at various age intervals. To adjust for changes in the length of the period we must do further adjustments as shown below.

$$\rho_{ist} = 1 - \left(1 - e^{\left[\sum_{j=1}^{J} \beta^s_{ji}\left(\frac{s}{S}\right)^j + \sum_{j=1}^{J} \beta^t_{ji}\left(\frac{t}{S}\right)^j\right]}\right)^{\frac{80}{S}} \tag{1.3}$$

2 Fertility Rates

We have data on fertility rates from age 23 to age 45 for past, present, and future years.

We will fit these to the following univariate polynomial function.

$$\ln f_{ist} = \sum_{j=1}^{J} \gamma^s_{ji} \left(\frac{s}{100}\right)^j + \gamma^t_{ji} \left(\frac{t}{100}\right)^j \tag{2.1}$$

where i indexes the country, s is the age, and t is the year.

As above, we fit each country separately in different regressions, but use a panel dataset for each country over age and year.

The vector of beta coefficients, B_i, can be estimated using OLS as above

$$\Gamma_i = (X'X)^{-1} X'Y_i \tag{2.2}$$

Now X is an $23 \times (J+1)$ matrix of ages and time periods raised to various powers, and Y_i is an $23T \times 1$ vector of the natural log of mortality hazard rates for country i for all ages and years.

Γ_i will thus be a $2(J+1)$ vector of coefficients with the first $J+1$ terms corresponding to the γ^s_{ji} coefficients and the last $J+1$ terms corresponding to the γ^t_{ji} coefficients

We fit these polynomials and save only the regression coefficients, the Γ_i's, to pass to the program.

In our Python program, to generate mortality hazard rates for agents that live for S periods we use the regression equation above replacing 100 with S. Note this gives us the one-year hazard rate at various age intervals. To adjust for changes in the length of the period we must do further adjustments as shown below.

$$\rho_{ist} = \left(e^{\left[\sum_{j=1}^{J} \gamma^s_{ji} \left(\frac{s}{S}\right)^j + \sum_{j=1}^{J} \gamma^t_{ji} \left(\frac{t}{S}\right)^j\right]} \right)^{\frac{80}{S}} \tag{2.3}$$

References