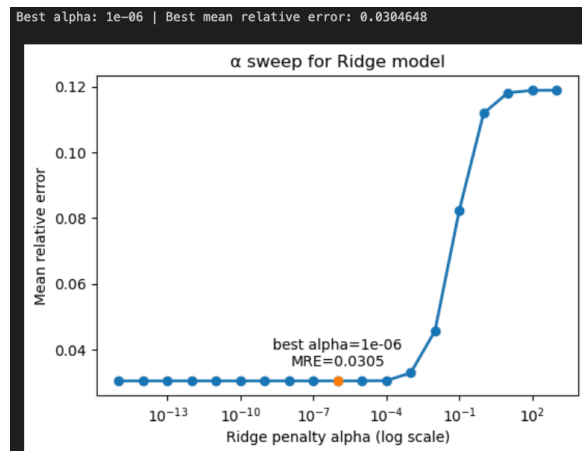


Census Prediction Minproject Report

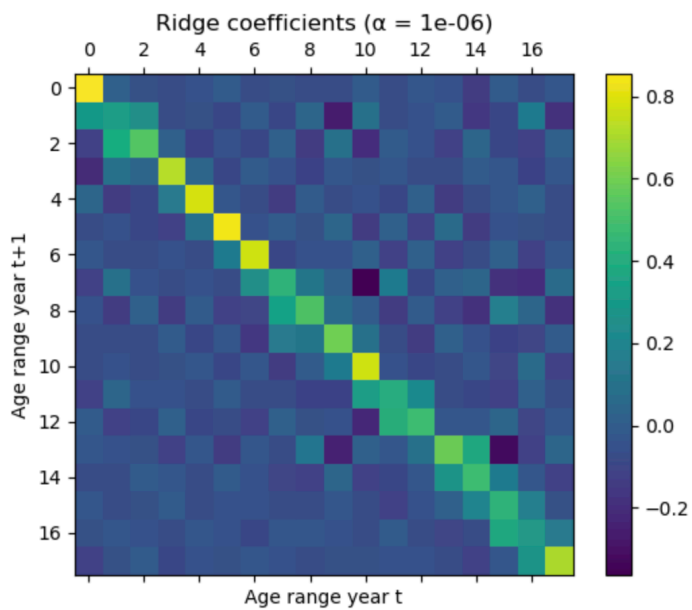
1) Optimal alpha and test mean relative error (only using power of 10's):

- $\alpha = 10^{-6}$
- Test mean relative error: 0.0304648



(fig 1)

2) Regression coefficient visualization:

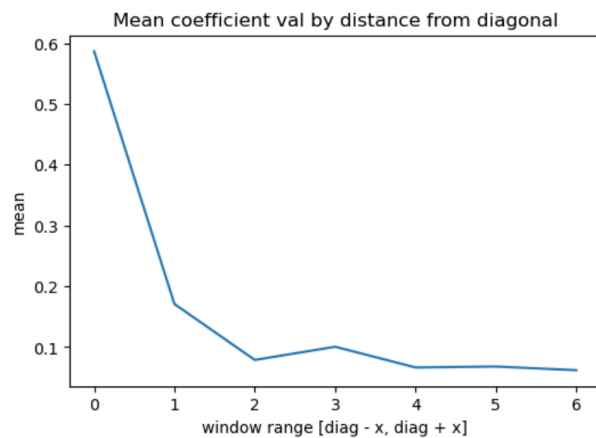


(fig 2)

3) Explanation:

The 18x18 grid comparison represents a transition matrix from year t to year $t+1$. There is an obvious pattern of concentration on the diagonals, which represents a similar coefficient from one year to the next.

I made an additional graph to quantify this pattern:



(fig 3)

This image plots the mean values on different windows from the diagonal of the graph in (fig 2). The average mean absolute value along the diagonal is 0.587. Taking 1 step out in both directions [diag - 1, diag + 1], the value shrinks to 0.171, and taking an additional step [diag - 2, diag + 2], we get an even lower 0.079. This supports the pattern of concentration around the diagonal band.

This pattern makes logical sense, as with 5 year age brackets and 1 year step size, most people would remain within their bracket from year t to $t+1$. There is a weaker diagonal concentration at the beginning and end, which would represent more variability from year to year in birth and death rates.

4) Code submission (also attached in email):

https://github.com/cpdiprete/Census_Statistical_Analysis