Alternative Choice-based Sampling Approach

I also tried choice-based sampling on both Y and X (cap ratio), which I will describe in this footnote. Since the population of triads is large (161,700), but the number of triads where Y=1 is comparatively small (approximately 120), this leaves a rather large population from which to draw my rather small random sample of triads where Y=0 (240). Given the small size of this random sample, it is quite possible to draw a random sample with little variation on **cap ratio**. In turn, this would limit the power of my test. I use the following procedures to stratify on X and Y:

1. Multiply the number of Y=1 triads by 2 to determine the number of Y=0 triads I need for my estimation. Call this value N_0

2. Divide N_0 by 4 to determine the number of observations I need to draw from each **cap ratio** quartile of the Y=0 triads. Call this value $\frac{N_0}{4}$

3. Starting with the first quartile of the Y=0 triads, take a random sample equal to size $\frac{N_0}{4}$.

4. Assign a weight to the observations in this sample. Assuming the first quartile has M observations, the weight, pi, is pi = $\frac{\frac{N_0}{4}}{M}$

5. Repeat for each quartile of triads.

6. Combine this sample of Y=0 triads with my Y=1 triads and estimate.

7. Repeat 500 times

This procedure lowers the bias from 1.27 to 1.13, lowers the Root Mean Squared Error from 3.59 to 3.49, and lowers Over Confidence from 7.34 to 6.55. Though these are reductions, I do not view them as substantial enough reductions to justify describing this procedure in the main text.