Data Cleaning Documentation

For the 2001, 2004, and 2008 SIPP surveys some work was needed to isolate the observations where someone lost their job and the ones where they acquired a new one. First, I kept observations with start dates that happened during the time periods covered by the surveys and job end dates were the reason for separation was either due to being discharged/fired or when due to the employer being bankrupt. Then using identification variables in the survey, I kept those who are observed to have both a job end date and a job start date.

The main thing to from here is to link the correct job start and end dates. The surveys observations are in person months and the job end date is listed for each month this happened in one survey wave (a survey wave is 4 months). So, to link them I did a cartesian join of the start dates and end dates. From there I calculated a difference between and observations start date and end date and only kept those with a positive start date. From there I only kept observations with the lowest which minimized this date difference for each person’s current jobless spell. Finally, I filtered the last of the person observations down to just those for the month that they lost their job and the month in which they got their new job.

From here I calculate the percent change in their wage from their new job compared to their old job and calculate some summary statistics. I do find an increase in monthly earnings from their new job. I am wondering if some of this might be due to me using the month when they are lose their job and the month in which they start their new job to calculate their monthly earnings. I am considering either using the month before to calculate their monthly earnings or an average of previous monthly earnings from that job. Otherwise the standard deviation on wages is quite large.