Meeting

February 9, 2024

1 Overview

- In our last meeting, I showed the results of our data shortening exercise, in which it appeared that the performance of our method seemed to get worse in when more data was available. I hypothesized that this was because the test set changed depending on the length. I changed the evaluation to make sure that the test set remained constant across different data lengths.

- After changing this, we still observe the same phenomenon. However, both methods seem to be affected by it.

- I also ran both methods on other midwestern states, our method outperformed Chen's in all other states and for all other lengths. However, there are times when neither method is better than no insurance.

2 "Selling" the method

- I think even without the multi-zone case this can be a significant contribution, at least to the index insurance literature.

- We can think of our method as being somewhere in between Chantarat's method, which yielded explainable, but not optimized contracts, and Chen's method, which yields possibly optimal, but completely unexplainable contracts.

- Our method yields contracts that are easily explainable, easily modified, and can accommodate constraints better than the two other methods.

- Additionally, our method performs better in low data settings, which is better for the developing country context.

- If we include the multi-zone case, then it can also be an advantage of our method, of managing risk better in those scenarios.

3 Results

3.1 Illinois

As before, our method outperforms Chen's method when there is less data available. More specifically, our method outperforms Chen's when the data length is less than or equal to 40, which corresponds to data being available since 1970. However, most satellite data is available from 1980.

Second, by looking at the cost plot we see that Chen's method seems to cost less. However, this is more from the model failing to recognize losses in the test set. The average payout in the training set tends to be higher than the average payout on the test set for Chen's model, even though average loss tends to be higher in the test set.
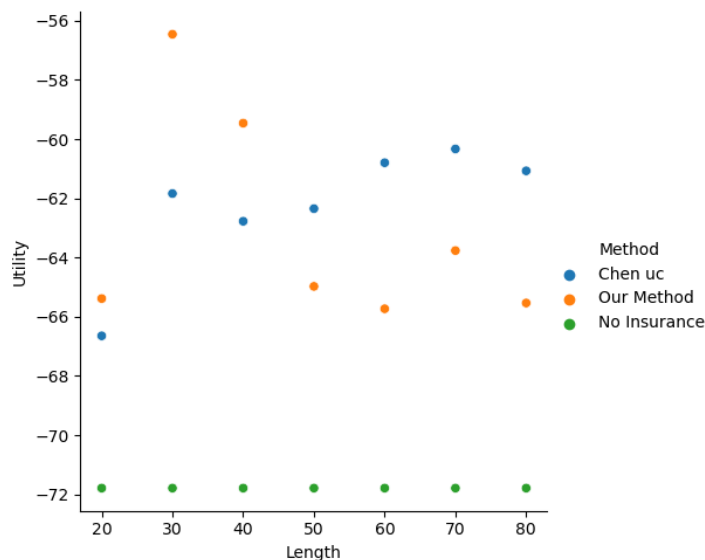
3.1.1 Our definition of the premium

Figure 1: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
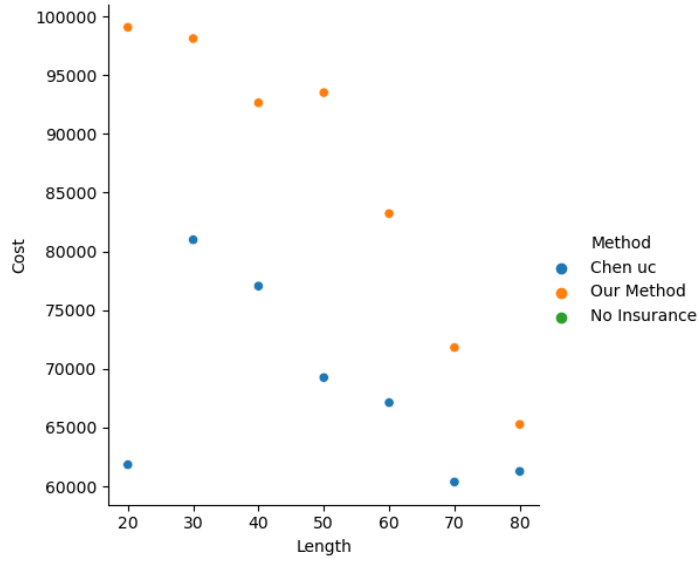
Figure 2: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available

3.1.2 Their definition of the premium

Here, we observe a similar phenomenon as in the previous case, both the performance of both models tends to deteriorate as more data is added. Note that this doesn't contradict Chen's results, since in the full data case we get a similar utility improvement to what they reported in the paper (13% vs 14%).
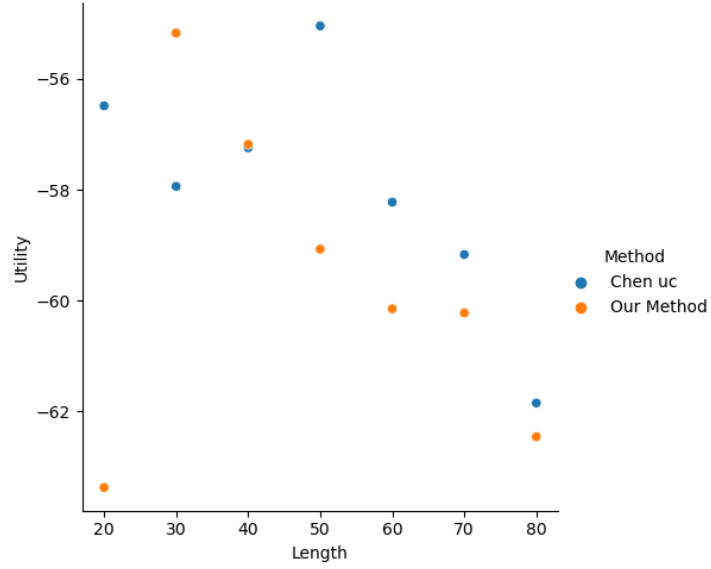
Figure 3: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
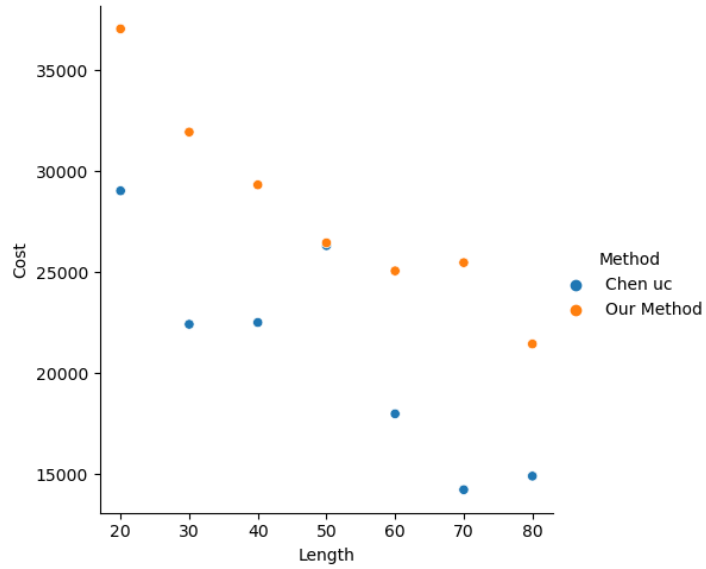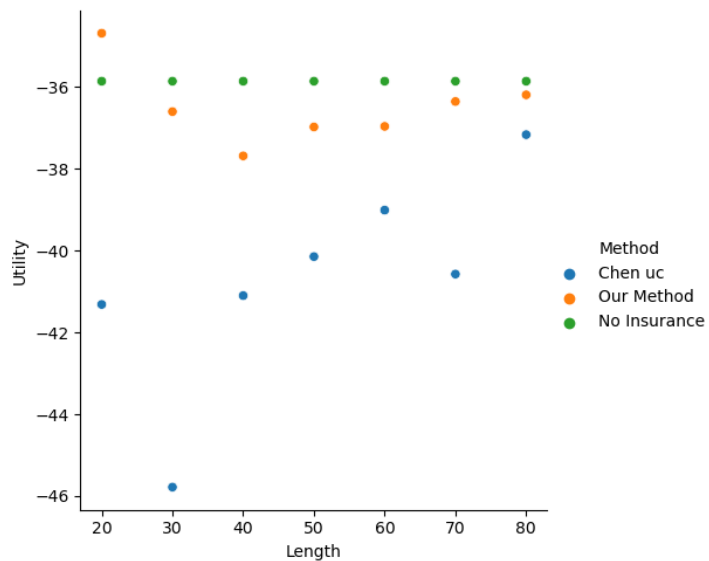

Figure 4: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available

4

3.2 Iowa

Here, our method outperforms Chen for all dataset lengths. Chen's method was much

Figure 5: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
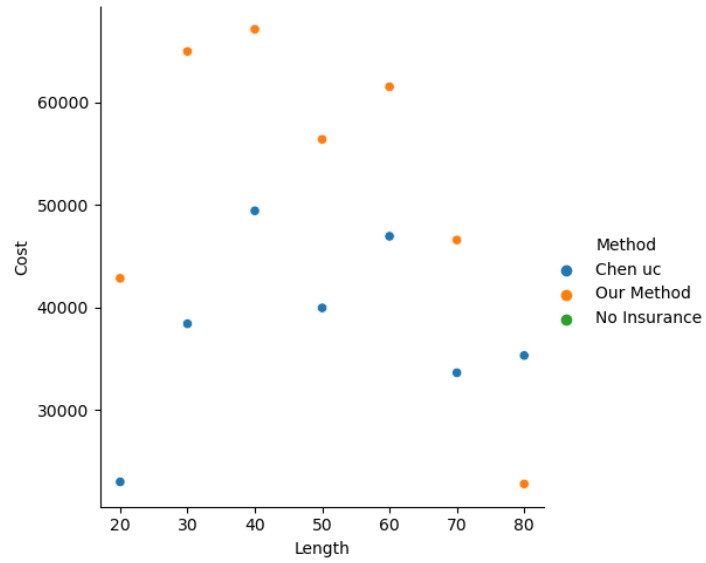
Figure 6: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
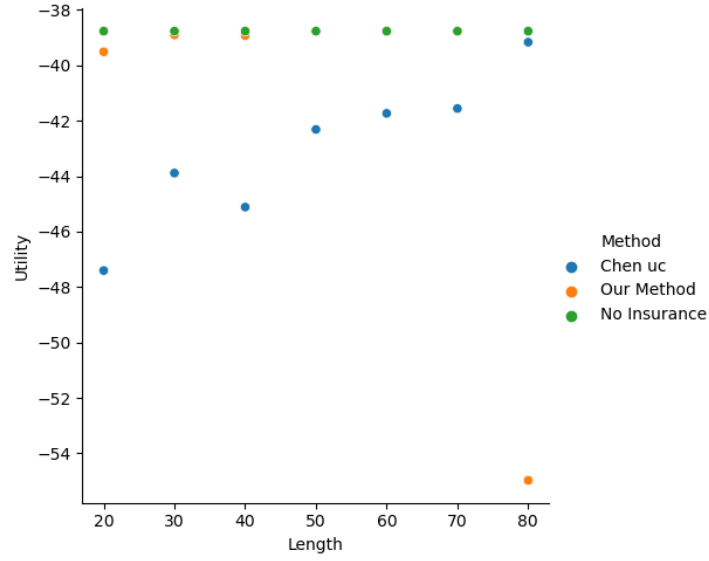
3.3   Indiana

Figure 7: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
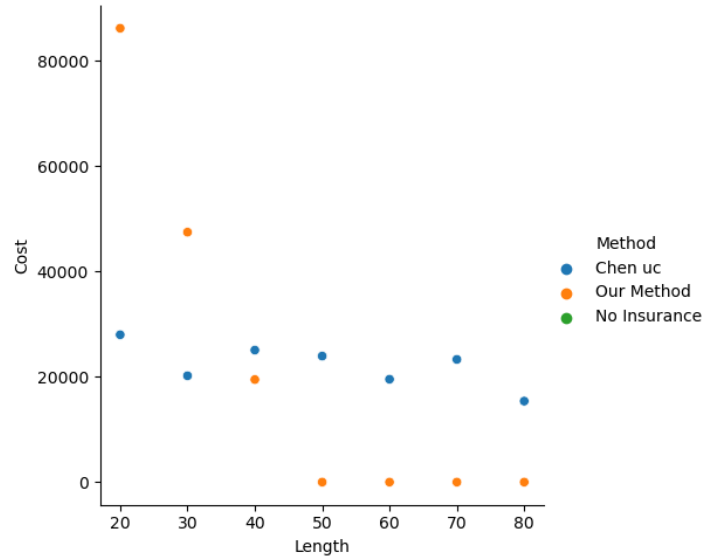

Figure 8: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
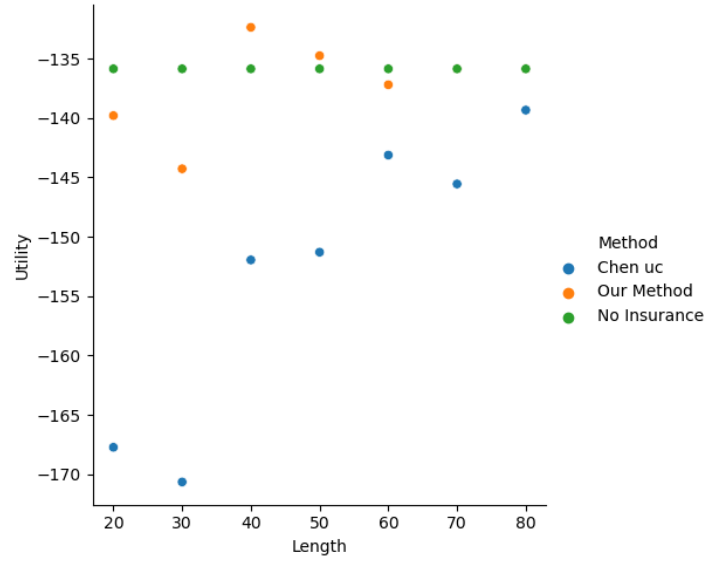
3.4 Missouri

Figure 9: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available
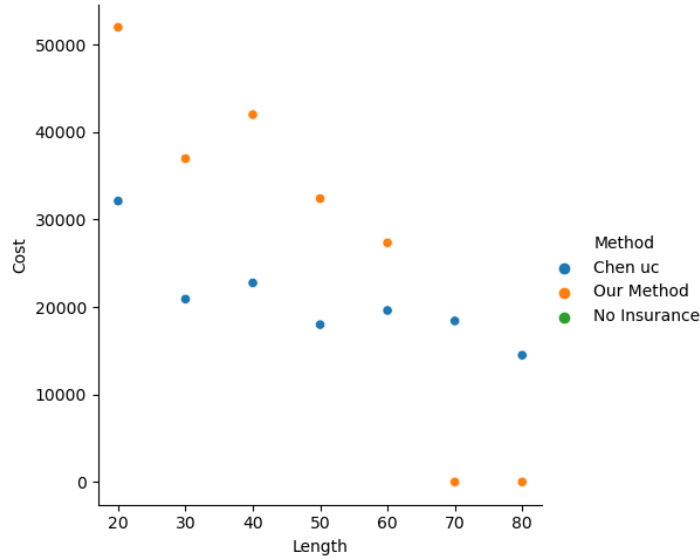
Figure 10: Data shortening exercise using our definition of the premium. Here, Length is dataset length, or number of years of data available

4 Next Steps

There are a couple of things we can add, I'm listing them in order of how much I think they would add to the paper.

- **Multiple Zones:** This is almost done, it'll take another day to complete.

- **Robustness to Misspecification:** I presented this at an econ workshop, and someone suggested that for the evaluation, I give each farmer a randomly chosen utility function, and that I compare how the two methods handle this.

- **Descriptive Analysis:** We could create plots to illustrate the difference between the two methods:

  1. Plot of payouts vs losses
  2. Distribution of farmer utility/wealth
  3. Distribution of insurer costs
  4. Distribution of insurer profits
  5. Plot illustrating in what cases farmer's are worse off with insurance