

# Evaluation Description

January 8, 2024

## 1 Overview

This document describes the initial evaluation I ran comparing our method to Chen’s method. I found that our method outperforms Chen’s method regardless of whether we use our definition of the premium or their definition of the premium. I also found that Chen’s method does worse than no insurance when we use our definition of the premium. I believe this is due to a difference in how we calculate the premium. I think that our method more accurately reflects practice. I describe the differences in section 3.3.

## 2 Data

To compare our method to Chen’s method, I use the same data sources they use. They get county-level annual corn production data for Illinois from the National Agricultural Statistics Service (NASS). They get climate and weather data from the PRISM Climate Group. Both datasets are publicly available.

**NASS Yield Data** Chen use annual yield data for Illinois from the National Agricultural Statistics Service (NASS). This provides county-level yield in units of bushels per acre from 1925-2018. We detrend this data using a second order polynomial fit with Thiel-Sen Regression, a robust regression method, as Chen describe in their paper. Chen don’t specify which robust regression method they used to fit this second order polynomial, so we tried several methods and Thiel-Sen was the one that yielded the most similar values to those found in Chen’s replication package. As Chen, we assume that the detrended yield data is both time and space homogenous.

**PRISM Weather Data** Chen use weather data from the PRISM Climate Group. This data is also available 1925-2018, and includes county-level monthly observations of precipitation, max/min temperatures, max/min vapor pressure deficit, and dew points.

## 3 Evaluation Procedure

### 3.1 Train/Test Split

Our dataset contains 7227 observations. First, we split the data into training, validation, and test sets. We use the same splits used by Chen in their replication package. The splits are as follows:

- **Train:** 1925 – 1991
- **Validation:** 1992 – 2003

- **Test:** 2004 – 2018

Chen’s method uses both the training set to train the neural network, and uses the validation set to determine when to stop training the network. However, the prediction algorithms we used for our method only require a training set, so we combine the training and validation set when using our method. In other words, we make sure that both methods have access to the same data when designing their contracts.

## 3.2 Design Contracts Using Both Methods

### 3.2.1 Our Method

First, we train a prediction model using the training and validation set. We select a model based on the performance metrics I described in the previous document. We then use the trained prediction model to generate predictions for the training set. These predictions as well as the true losses are then used as input for our optimization program to design the contracts. Once these contracts have been designed, we use the prediction model to generate predictions for observations in the test set, and simulate the payouts on the test set based on these predictions and the contracts the optimization program designed.

### 3.2.2 Chen’s Method

We implemented the neural network described by Chen et al 2023. We used the same architecture and parameter values we found in the replication package. We trained this network using the training set, and used the validation set to determine the best model. We then used this network to generate payouts for the test set.

## 3.3 Compare Outcomes

After the previous step, we now have losses and payouts for observations in the test set under both methods. Using this, we can compute the wealth and utility for each farmer under both methods. Farmer’s wealth is defined as in Chen et al 2023:

$$w = w_0 - y + I(X) - \pi(I(X)) \quad (1)$$

Where  $w_0 = 388.6$ ,  $y$  is the loss,  $I(X)$  is the payout, and  $\pi(I(X))$  is the premium for contract  $I$ . Here, the premium is computed using the training data. In other words, suppose we define the premium to be  $\pi = \mathbb{E}[I(X)]$ , then for our evaluation the premium would be the empirical expected value of the contract on the training set.

### 3.3.1 Note about premiums

The biggest modeling difference between our approach and Chen’s approach is in how we define the premium. We define the premium as

$$\pi = \mathbb{E}[I(X)] + c_\kappa (\text{CVaR}(I(X)) - \mathbb{E}[I(X)]) \quad (2)$$

This is consistent with Mapfumo et al 2017 and has the advantage of incorporating tail risk into the premium, which would make sense in practice.

Chen et al, on the other hand, define the premium as  $\pi = \lambda \mathbb{E}[I(X)]$ .  $\lambda$  is determined endogenously by their model, and they claim that their best performing model has  $\lambda = 1.2414$ .

At first, for purposes of the evaluation, I used our definition of the premium to evaluate both contracts. Let  $I^{\text{chen}}$  be the contract designed by Chen’s method and let  $I^{\text{VX}}$  be the contract designed by our method. When computing the farmer wealth under insurance contract  $I(X)$ , I would compute that contract’s premium as  $\pi = \hat{\mathbb{E}}[I(X)] + c_k \left( \text{CVaR}[I(X)] - \hat{\mathbb{E}}[I(X)] \right)$ , where  $\hat{\mathbb{E}}[I(X)]$  is the empirical expected value of  $I(X)$  on the training set, and  $\text{CVaR}[I(X)]$  is the empirical CVaR on the training set. However, when I did this, I noticed that farmer utility was lower with Chen’s insurance than under no insurance. In other words, Chen’s method wasn’t replicating. After further investigation, I realized that this was because during training, Chen’s model is assuming that the premium is  $\lambda \mathbb{E}[I(X)]$ , whereas during the evaluation, the premium was calculated differently. In other words, our model was outperforming Chen’s model at least in part because our model was trained using the same definition of the premium as was used in the evaluation. Chen’s method, on the other hand was being trained using a different definition than was used in the evaluation. It didn’t seem like a fair comparison.

As a result, I decided to run two experiments:

1. In the first experiment, I modified Chen’s method to use the same definition of the premium as our method. I did this by changing how the premium is defined in the loss function used for training the neural network. I then designed contracts using both methods and used our definition of the premium in the evaluation.
2. In the second experiment, I modified our method to use the same definition of the premium as Chen’s method. I then designed contracts using both methods and used Chen’s definition of the premium in the evaluation.

We can think of these experiments as seeing how the two methods perform under different assumptions. In the first experiment, we are assuming that our definition of the premium is the correct one. In the second experiment we are assuming Chen’s definition of the premium is the correct one.

### 3.3.2 Note about evaluation timing

I was a little skeptical about the fact that Chen’s method did not seem to improve utility when we used our definition of the premium, so I decided to look into the matter. As mentioned in the previous section, for the purposes of the evaluation, I calculate the premium using the training set. I do this because, in practice, insurers have to set the premium based on historical (i.e. the training) data. However, when looking at Chen’s code I discovered that in their evaluation, they calculate the premium based on the payouts in the test set (they don’t mention this in the paper). I think this is wrong. In practice, the insurer has to set the premium before they can observe the current period’s payouts. This discrepancy explains why Chen’s method does worse than no insurance when calculating the premium using our method. When I calculate the premium using Chen’s timing, (i.e. the premium for the evaluation is based on the payouts in the test set instead of the payouts in the training set), Chen’s method does better than no insurance. However, when I use the timing that I think most accurately reflects practice (i.e. the premium for the evaluation is based on payouts in the training set instead of payouts in the test set), Chen’s method does worse than no insurance.

Chen’s method seems to overfit to the training data. The average loss is very similar between the training set and the test set (821 and 824 respectively), however the average payouts from Chen’s

method are significantly higher in the training set than in the test set (137 and 99 respectively). In contrast, the difference in average payouts in the training set and test set for our method is much lower (83 vs 90).

## 4 Results

Our method outperforms Chen’s method, regardless of the definition of the premium that is used. Our method provides larger utility gains for farmers at a lower risk for the insurer. Additionally, it is more interpretable since it provides a simple linear contract between payouts and predicted losses. While we found that Chen’s method leads to a utility improvement when we use their definition of the premium, when we use our definition of the premium, farmers are worse off with Chen’s method than under no insurance. It appears that Chen’s network isn’t able to learn as well when the definition of the premium is changed. Note, the utilities shown in Chen et al 2023 are also negative. However, ours are more negative, I believe this comes from the fact that we use different methods to detrend the time-series. In the tables below, CEW is the certainty equivalent wealth, which is a measure of how much the farmer would be willing to pay for the insurance, based on the utility improvements provided by the insurance. A larger utility improvement leads to a larger CEW.

Method	Utility	CEW	Premium	Insurer Risk	Market Loading
No Insurance	-9574	-930	0.000	0.0	1.2414
Chen	-7701	-903	51.572	348.9	1.241
Our Method	-7050	-892	49.955	189.6	1.241

Table 1: Results using Chen’s definition of the premium:  $\pi = \lambda \mathbb{E}[I(X)]$ ,  $\lambda = 1.2414$

Method	Utility	CEW	Premium	Insurer Risk	Market Loading
No Insurance	-9574	-930	0.000	0.000	1.0
Chen	-10088	-937	165.951	327.436	1.0
Our Method	-7601	-902	99.993	201.679	1.0

Table 2: Results using our definition of the premium:  $\pi = \mathbb{E}[I(X)] + c_{\kappa}(\text{CVaR}(I(X)) - \mathbb{E}[I(X)])$

## 5 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:** We could compare the cost of insuring the US midwest (Illinois, Indiana, Missouri, Iowa) under both methods.
- **Data Shortening:** We could see how the performance of both methods is affected as we have fewer years of data available.
- **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