Agricultural Index Insurance Design: An Optimization Approach

José I. Velarde Morales

November 3, 2022

1 Introduction

Lack of access to credit and insurance is often cited as a significant factor hindering agricultural productivity in developing countries. Nearly two thirds of the world's poor are employed in agriculture, and addressing this problem could have significant welfare implications. Agricultural insurance is, even in the best circumstances, a hard problem. Many of the features one would want (independent units, uncorrelated risk, etc) are missing in this context. When considering insurance in developing countries, the problem becomes even harder because of verification costs. Traditionally, whenever an adverse event happens, the insured party contacts the insurer, and the insurer verifies the claim and issues a payout. However, agriculture in developing countries is often characterized by many small farmers spread out over hard to reach regions. This makes verification prohibitively costly. Additionally, the presence of correlated risks makes insurance more expensive because it makes large payouts more likely. Intuitively, if one farmer is affected by a drought, it is likely that other farmers were also affected. If large payouts are more likely, the insurer must have larger reserves in order to maintain solvency.

Researchers developed index insurance as a less costly way to offer insurance in developing countries. In index insurance, an index (or statistic) is created using easily observable quantities, and it is used to determine whether the insured party suffered an adverse event. In the past, indices have been constructed using rainfall, weather, and satellite images. If the index falls below a predetermined threshold, the insurance company automatically issues out payments to the insured. This allows the insurance company to circumvent the issue of verification, moral hazard, and adverse selection, since the actions of individual farmers cannot affect the index. Even though index insurance has proved to be a less costly way of providing insurance for small farmers, it has been difficult to scale up. There are several problems with index insurance. One of the main problems is low take up: farmers are often unwilling to purchase the insurance at market prices. Another problem, as previously mentioned, is the cost. The purpose of this project is to make this insurance less costly by improving the design of insurance contracts. The rest of this paper is organized as follows: Section 2 reviews the existing literature on index insurance, Section 3 describes our proposed approach, Section 4 describes our evaluation methods and results. We conclude with a brief discussion.

2 Literature Review

Impact of Index Insurance There are many studies that evaluate how access to index insurance impacts the behavior of farmers. Through a randomized evaluation in Northern Ghana, Karlan et al. (2014) found that farmers shifted their production to riskier but potentially more profitable crops when they had access to index insurance. Similarly, Cole et al. (2013) found that farmers in India that had access to insurance were more likely to produce cash crops. Mobarak and Rosenzweig (2013) conducted experiments in several states in India and found that insured farmers were more likely to grow high-yield varietes of rice. Overall, there is evidence that index insurance reduces reliance on detrimental risk-coping strategies, increases investment, and leads to riskier, but more profitable production decisions (Jensen and C. Barrett (2017)).

Demand for Index Insurance One of the largest barriers to the scale up and adoption of index insurance is low demand (Jensen and C. Barrett (2017)). Cole et al. (2013) found that demand for index insurance is highly sensitive to price and liquidity constraints. When offered discounts, over 60% of farmers opted to purchase the insurance product. They also found that cash grants made farmers more likely to purchase insurance. Cai, De Janvry, and Sadoulet (2020) found that subsidies and financial education increased take up of index insurance. Casaburi and Willis (2018) tested the effect of liquidity constraints on demand. They reduced liquidity constraints by collecting premiums at harvest time (when farmers have more cash), instead of the standard pay-up-front scheme. They found that this payment scheme increased take up to 72% from a baseline of 5%.

Design of Index Insurance There has been relatively little research done on the design of index insurance. In Chantarat et al. (2013), the authors describe the design of an index insurance for pastoralists in Northern Kenya. This insurance is based on a satellite based index, and is what is used in Kenya's Index Based Livestock Insurance (IBLI) program. In Flatnes, Carter, and Mercovich (2018) the authors propose augmenting a traditional index insurance contract with the option for an audit. In this augmented contract, the insured farmer has the option to request an audit if they believe a payout should have been issued but wasn't. In Jensen, Stoeffler, et al. (2019), the authors compare the welfare implications of using different satellite based indices for insuring pastoralists against drought. The method developed by Chantarat et al. (2013) is used in all of these studies. There are also numerous non-academic publications describing the implementation of index insurance programs in different parts of the world (Osgood et al. (2007),Bank (2011),Greatrex et al. (2015)). From these papers describing the implementation of these programs, it appears that there is no a standard methodology for developing index insurance products. As stated in Bank (2011), "The reader should be aware that there is no single methodology in this field ... [this paper] describes an approach that has been used in a number of index pilot activities undertaken by the World Bank and its parners."

Optimization Literature In this work, we will be drawing from the literature on chance constrained programs (Lagoa, Li, and Sznaier (2005); Charnes, Cooper, and Symonds (1958)). We also draw on the work on coherent risk measures (Artzner et al. (1999)), and work on the optimization of conditional value at risk by (Rockafellar, Uryasev, et al. (2000)). Additionally, we use the results on convex approximations of chance constrained programs by (Nemirovski and Shapiro (2007)).

3 Optimization Approach

3.1 Objective and Constraints

We conducted interviews with practitioners that had implemented index insurance programs in numerous countries (Malawi, Kenya, Senegal, Thailand, among others) to determine an appropriate objective for the model. Practitioners expressed that the ideal objective would be to minimize the probability that farmers' wealth would drop below a certain threshold. This is motivated/supported by the body of research on poverty traps. Poverty traps are typically defined as “any self-reinforcing mechanism which causes povery to persist” (Azariadis and Stachurski (2005)). The idea is that if a household’s wealth drops below a certain threshold, they can be stuck there for a long time. For example, when faced with a negative income shock, a farmer might be forced to sell some of their productive assets. Without these productive assets, it can be very difficult to accumulate enough savings to re-purchase the productive asset, and the farmer might be stuck at the lower income level for a long time. One of the main purposes of insurance is to prevent these kinds of situations.

The most important constraints in practice are budget constraints and payout frequency constraints (Osgood et al. (2007),Bank (2011)). Price constraints are important from both the demand and supply side. On the supply side, insurers have strong preferences with regards to the premium of the product they offer. From the demand side, demand for index insurance is highly price elastic (Jensen and C. Barrett (2017)). Payout frequency constraints are similarly from both the demand and supply sides. From the supply side, insurers don’t want payout frequency to be too high, becuase there are fixed costs associated with issuing payouts. From the demand side, farmers generally have low demand for an instrument that pays out very infrequently (Osgood et al. (2007)). Index insurance contracts normally define piecewise linear payout functions (Bank (2011),Chantarat et al. (2013)). Piecewise linear functions are popular because they are easy to explain, and have been found to work well in practice. Simplicity is an important feature, because low understanding of the product lowers demand for insurance (Cai, De Janvry, and Sadoulet (2020)).

3.2 Index Insurance Definition and Parameters

Index insurance generally involves an easily observable signal, θ, that is used to predict the loss, $\hat{\ell}(\theta)$, of some agricultural product. For example, θ could be rainfall, and $\hat{\ell}(\theta)$ could be livestock mortality. Index insurance contracts normally have the form: $I(\hat{l}(\theta)) = \min \left\{ \max \left\{ 0, a\hat{\ell}(\theta) + b \right\}, P \right\}$, where P is the maximum payout, and a, b are the contract parameters. For ease of notation, we will use $I(\theta)$ instead of $I(\hat{\ell}(\theta))$. The expected cost, $C(I(\theta))$ of an insurance contract, $I(\theta)$ for an insurer in a single period is: $C(I(\theta)) = \mathbb{E}[I(\theta)] + c_\kappa K(I(\theta))$, where c_κ is the cost of holding capital, and K is the amount of capital required to insure the contract. K is set by regulators, and is meant to ensure that insurers have enough capital to fulfill their contractual obligations with high probability. One commonly used formula for K is $K(I(\theta)) = CVaR_{1-\epsilon_P} \left(I(\hat{l}(\theta)) \right) - \mathbb{E}[I(\theta)]$ (Mapfumo, Groenendaal, and Dugger (2017)). ϵ_P is set by regulators, and commonly used values are $\epsilon_P = 0.01$ or $\epsilon_P = 0.05$.

3.3 CVaR Model

3.3.1 Motivation

We want a model that minimizes the probability that wealth drops below a certain threshold, subject to a budget constraint. However, probabilistic objectives are generally non-convex. The probability that wealth drops below a certain threshold is a measure of the risk farmers face; we want a convex measure that will capture this risk. We use the Conditional Value at Risk $CVaR$ of the loss net of insurance as our measure of risk.

Definition 1 *For a random variable z, representing loss, the $(1-\epsilon)$ Value at Risk (VaR) is given by*

$$VaR_{1-\epsilon}(z) := \inf \{ t : P(z \leq t) \geq 1 - \epsilon \}$$

Definition 2 *For a random variable z, representing loss, the $(1-\epsilon)$ Conditional Value at Risk ($CVaR$) is given by*

$$CVaR_{1-\epsilon}(z) := \mathbb{E}\left[z | z \geq VaR_{1-\epsilon}(z)\right]$$

Intuitively, the Conditional Value at Risk is the expected value of a loss given that it is above a certain threshold. It can also be thought of as measuring the worst outcomes. By minimizing the $CVaR$ of the net loss, we are focusing on improving farmers' wealth in the worst case scenarios, which is a key purpose of insurance. $CVaR$ has been extensively studied in the academic literature, and is also popular in practice (Rockafellar, Uryasev, et al. (2000),Rockafellar and Uryasev (2002),Artzner et al. (1999)). It also has the advantage of being convex, and thus amenable to optimization. This leads us to the following model:

$$\min_{a,b,\pi,K} \quad CVaR_{1-\epsilon}\left(\ell - I(\theta)\right) \tag{1}$$

$$\text{s.t. } I(\theta) = \min\{(a\hat{\ell}(\theta) + b)^{+}, P\} \tag{2}$$

$$\mathbb{E}\left[I(\theta)\right] + c_{\kappa} K \leq B \tag{3}$$

$$K = \left(CVaR_{1-\epsilon}\left(I(\theta)\right) - \mathbb{E}[I(\theta)]\right) \tag{4}$$

Our objective is the conditional value at risk of the farmers' loss net of insurance. The first constraint specifies the piecewise linear structure of the contract, and the second constraint is the budget constraint. Recall that the total cost of insurance consists of payouts plus the cost of capital. The last constraint is just the definition of the required capital.

3.3.2 Changes to constraints

To make the program convex, we replace $I(\theta)$ with conservative approximations where necessary, making sure that by doing so we still get a feasible solution. We replace $I(\theta)$ in the objective with $\min\{a\hat{\ell}(\theta) + b, K\}$. This will give us an lower bound on the performance of our contracts, since $\ell - \min\{a\hat{\ell}(\theta) + b, K\} \geq \ell - \min\{(a\hat{\ell}(\theta) + b)^{+}, P\}$. We replace $I(\hat{\ell}(\theta))$ in the budget constraint with $\max\left\{0, a\hat{\ell}(\theta) + b\right\}$. This is a conservative approximation of the constraint, since $\min\{(a\hat{\ell}(\theta) + b)^{+}, P\} \leq \max\left\{0, a\hat{\ell}(\theta) + b\right\}$. We also need approximations or proxies for $E[I(\theta)]$ in constraint 4. We use $\underline{\pi} \leq E[I(\theta)]$ or $\pi_{SQ} = E[I_{SQ}(\theta)]$, where I_{SQ} is the contract designed using the status quo method, as a proxy for $E[I(\theta)]$ in constraint 4.

3.3.3 Single Zone Model

Our single zone model minimizes the conditional value at risk of farmers' net loss (i.e. loss net of insurance), subject to an overall budget constraint. The total cost of the insurance consists of payouts and of costs associated with holding capital. We first describe the model parameters, and then describe the model. The model will output values for a and b, and the final insurance contracts will be $I(\theta) = \min\left\{\max\left\{0, a\hat{\ell}(\theta) + b\right\}, P\right\}$.

Model Parameters

- ϵ: This is the ϵ used for the $CVaR$ objective. $\epsilon = 0.1$ means that our objective is $E[\ell - I(\theta) | \ell - I(\theta) \geq VaR_{1-0.1}\left(\ell - I(\theta)\right)]$.

- ϵ_K: This is the epsilon used in the formula for required capital. Recall that the required capital $K = CVaR_{1-\epsilon_K}(I(\theta)) - E[I(\theta)]$.
- B: This is the budget constraint for the total cost of the insurance.
- π_{SQ}: This is the average payout made by the status quo contract, $E[I_{SQ}(\theta)]$. This is also called the premium of the status quo contract.
- P: This is the maximum insured amount.
- c_κ: This is the cost of capital.

Model In the model below, our objective is the conditional value at risk of the farmer's loss net of insurance. The first constraint is the budget constraint and the last constraint is the definition of the required capital.

$$\min_{a,b,K} CVaR_{1-\epsilon}\left(\ell - \min\left\{(a\hat{\ell}(\theta) + b), P\right\}\right) \tag{5}$$

$$\text{s.t. } \mathbb{E}[\max\left\{0, a\hat{\ell}(\theta) + b\right\}] + c_\kappa K \leq B \tag{6}$$

$$K = CVaR_{1-\epsilon_P}\left(\max\{0, a\hat{\ell}(\theta) + b\}\right) - \pi_{SQ} \tag{7}$$

We reformulated the problem as a linear program using the results from Rockafellar, Uryasev, et al. (2000). In the model below, p^j is the probability of event j, and j indexes the possible realizations of θ, ℓ. N is the total number of samples.

$$\min_{a,b,\gamma,\gamma_K,\alpha,t,t_K} \quad t + \frac{1}{\epsilon}\sum_j p^j \gamma^j \tag{8}$$

$$\text{s.t. } \gamma^j \geq \ell^j - \min\left\{(a\hat{\ell}(\theta^j) + b), P\right\} - t, \forall j \tag{9}$$

$$\gamma^j \geq 0, \forall j \tag{10}$$

$$B \geq \frac{1}{N}\sum_j \alpha^j + c_\kappa K \tag{11}$$

$$t_K + \frac{1}{\epsilon_K}\sum_j p^j \gamma_K^j \leq K + \pi_{SQ} \tag{12}$$

$$\gamma_K^j \geq \alpha^j - t_K, \forall j \tag{13}$$

$$\gamma_K^j \geq 0, \forall j \tag{14}$$

$$\alpha^j \geq a\hat{\ell}(\theta^j) + b, \forall j \tag{15}$$

$$\alpha^j \geq 0, \forall j \tag{16}$$

3.3.4 Multiple Zone Model

The multiple zone model is very similar to the single zone model. In the objective, we minimize the maximum conditional value at risk of farmers' net loss across all zones, z. We minimize the maximum $CVaR$ across all zones to avoid situations where one zone has a contract that is significantly worse than other zones. The other change is that the budget constraint includes the payouts of all zones, and the required capital is determined using the sum of payouts across all zones.

Model Parameters

- ϵ: This is the ϵ used for the $CVaR$ objective. $\epsilon = 0.1$ means that our objective is $E[\ell - I(\theta)|\ell - I(\theta) \geq VaR_{1-0.1}(\ell - I(\theta))]$.
- ϵ_K: This is the epsilon used in the formula for required capital. Recall that the required capital $K(I(\theta)) = CVaR_{1-\epsilon_K}(I(\theta)) - E[I(\theta)]$.
- Z: number of insured zones.
- c_κ: cost of capital.
- P_z: maximum payout for zone z.
- π_{SQ}: This is the average payout made by the status quo contract, $E[I_{SQ}(\theta)]$. This is also called the premium of the status quo contract.

Model In the model below, our objective is the maximum conditional value at risk of the net loss across all zones. The second constraint is the budget constraint, which now includes the sum of payouts across all zones. The formula for required capital was also changed to include the sum of payouts across all zones.

$$\min_{a,b,K} \max_{z} \quad CVaR_{1-\epsilon}\left(\ell_z - \min\left\{(a_z\hat{\ell}_z(\theta_z) + b_z), P_z\right\}\right) \tag{17}$$

$$\text{s.t. } \mathbb{E}\left[\sum_z \max\left\{0, a_z\hat{\ell}_z(\theta_z) + b_z\right\}\right] + c_\kappa K \leq B \tag{18}$$

$$K + Z\pi_{SQ} \geq CVaR_{1-\epsilon_K}\left(\sum_z \max\left\{0, a_z\hat{\ell}_z(\theta_z) + b_z\right\}\right) \tag{19}$$

We reformulated the problem as a linear program using the results from Rockafellar, Uryasev, et al. (2000). In the model below, p^j is the probability of event j, and j indexes the possible realizations of θ, ℓ. N is the total number of samples.

$$
\begin{aligned}
\min_{a,b,\alpha,\gamma,t,m,K^P} \quad & m && (20)\\
\text{s.t.} \quad & t_z + \frac{1}{\epsilon}\sum_j p^j \gamma_z^j \leq m, \forall z && (21)\\
& \gamma_z^j \geq \ell^j - \min\Big\{(a_z \hat{\ell}_z(\theta_z^j) + b_z), P_z\Big\} - t_z, \forall j, \forall z && (22)\\
& \gamma_z^j \geq 0, \forall j, \forall z && (23)\\
& B \geq \frac{1}{N}\sum_j \sum_z \alpha_z^j + c_\kappa K && (24)\\
& t_K + \frac{1}{\epsilon_K}\sum_j p^j \gamma_K^j \leq K + Z\pi_{SQ} && (25)\\
& \gamma_K^j \geq \sum_z \alpha_z^j - t_K, \forall j && (26)\\
& \gamma_K^j \geq 0, \forall j && (27)\\
& \alpha_z^j \geq a_z \hat{\ell}_z(\theta_z^j) + b_z, \forall j, \forall z && (28)\\
& \alpha_z^j \geq 0, \forall j, \forall z && (29)
\end{aligned}
$$

4 Evaluation

4.1 Baseline Approach

We evaluate our method by comparing its performance with the method developed by Chantarat et al. (2013). This method is the standard method used in academic publications describing the design of index insurance contracts (see Flatnes, Carter, and Mercovich (2018); Jensen, Stoeffler, et al. (2019)). It is also what was used to design Kenya's Index Based Livestock Insurance (IBLI) program. In what follows, we will be calling the smallest unit of observation a "location." In practice, this could be a village or a group of villages. When designing index insurance, these smaller units of observation are assigned to larger groups, or zones.

4.1.1 Method description

- First, locations are assigned to zones. Assignments are either taken as exogenously given, or a clustering algorithm is used to group locations into zones. This clustering is usually based on historical weather data.
- Historical data is then used to fit a linear regression model to predict losses in each cluster. A different model is estimated for each cluster.
- Contracts are of the form: $I(\theta) = \max(\hat{\ell}(\theta) - \ell^*, 0) \times TIU \times P_{IU}$ where $\hat{\ell}(\theta)$ is the predicted loss rate, ℓ^* is the strike value, TIU is the total number of insured agricultural units, and P_{AU} is the price per insured agricultural unit. In other words, their contract pays farmers for the full predicted loss beyond a threshold, ℓ^*. This threshold, ℓ^* is the contract's strike value.
- The next step is to set the strike value. The method chooses the strike value that would explain the highest share of insured losses in the historical data. Specifically, the method

runs the following regression: $y_s = \beta_s \hat{y_s} + \epsilon$ where y_s is the actual insured losses at strike value s and $\hat{y_s}$ is the predicted insured losses at strike value s. For example, suppose that $TIU = 100$ (ie there are 100 insured units), and that $P_{IU} = 25$ (ie each unit is worth 25), and that $\ell^* = 0.25$ (ie contract starts paying out once the predicted loss rate exceeds 25%). If the actual loss rate is 0.5, then actual insured losses would be $y_{25} = \max(\ell - \ell^*, 0) \times TIU \times P_{IU} = (0.5 - 0.25) \times (100) \times (25)$. If the predicted mortality rate in that scenario was 0.4, the predicted insured losses, $\hat{y_{25}} = \max(\hat{\ell}(\theta) - \ell^*, 0) \times TIU \times P_{IU} = (0.4 - 0.25) \times (100) \times (25)$. The method uses historical data to calculate $y_s, \hat{y_s}$, and then runs the following regression: $y_s = \beta_s \hat{y_s} + \epsilon$. The method chooses the strike value $s = \arg\max_s \beta_s$. The goal of choosing the strike value that explains the largest share of the losses is to minimize the basis risk, which is the probability that a loss occurs but that the insurance contract doesn't pay out. This takes into account the fact that the prediction model, $\hat{\ell}(\theta)$ might be better at predicting some losses better than others. For example, we could have a prediction model that is good at predicting large losses, but bad at predicting small losses.

4.2 Simulation Set Up

In this section, we describe how we set up the simulation used to evaluate our method. We describe the data generating process, the scenarios we test, and the simulation itself. We test our model on a toy example consisting of two zones. The goal of this exercise is to provide a fair comparison of the two methods. As a result, we will make sure that the two methods have the same budget constraint.

4.2.1 Data Generating Process

For the two zone example, we generate samples from two models: a linear model and a non-linear model. We choose a quadratic model for the non-linear model because it is the simplest non-linear model. Since our prediction model is a linear model, this will allow us to evaluate how the two models perform in cases where the prediction model is misspecified. The data generating processes are as follows:

- Linear DGP: $\ell = \beta\theta + \epsilon$
- Nonlinear DGP: $\ell = \beta\theta^2 + \epsilon$

In both cases we have: $\theta \sim \mathcal{N}((5, 5), \Sigma), \beta = diag(1.5, 1.5), \epsilon \sim \mathcal{N}(0, I)$.

4.2.2 Optimization Model Parameters

We use the following parameters for our optimization model in the simulations:

- $\epsilon = 0.2$ We picked this because it focuses on minimizing the CVaR of the 80^{th} percentile of the loss distribution, which roughly corresponds to once in every 5 year events, which is the desired frequency of insurance payouts.
- $\epsilon_P = 0.01$ This is a commonly set value by regulators.
- $K_z = 8, \forall z$ This is the maximum payout given by the status quo method.
- $c_k = 0.15$ This is an estimate from the literature (Kielholz (2000)).

4.2.3 Scenarios to be tested

We are interested in how the two models behave in three basic scenarios. The first scenario is when there is no correlation between the losses in the two insured zones. The second scenario is when the losses in the two zones are positively correlated. The last scenario is when the losses in the two zones are negatively correlated. We test both DGPs for each scenario.

No correlation Case This is the baseline case where the losses of the two zones are uncorrelated.

- $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

Positive correlation Case This is the case where the losses of the two zones are positively correlated. This is an important scenario to test, because it is highly common in agricultural insurance. Covariate risk is one of the reasons that agricultural insurance is so difficult, since it increases the risk of catastrophic losses for the insurer. Intuitively, if one farmer was affected by a drought, it is likely that many others were affected as well. This can happen when two insured areas are geographically close to each other.

- $\Sigma = \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}$

Negative correlation Case This is the case where the losses of the two zones are negatively correlated. This is a common feature of large scale climate processes. For example, certain El Nino-Southern Oscillation states are associated with increased rainfall in the Greater Horn of Africa, but with increased drought probabiliy in Southern Africa (C. B. Barrett et al. (2007)).

- $\Sigma = \begin{bmatrix} 2 & -1.6 \\ -1.6 & 2 \end{bmatrix}$

4.2.4 Simulation details

For each scenario we draw 300 samples for training, 50 samples for parameter selection (which are used to select the strike value in the baseline model), and 100 samples for evaluation. We run 1000 of these scenarios and compute the metrics for each one. We then report the median, 5^{th}, and 95^{th} percentile values of each performance metric across the 1000 simulations.

1. Draw samples from model, samples will be of the form (ℓ, θ) where ℓ is loss and θ is the predictor.

2. Train linear prediction model. We run $\ell = \beta_0 + \beta_1 \theta + \epsilon$. Use model to generate predictions $\hat{\ell}(\theta)$ for training and test data.

3. Determine the parameters for baseline contracts using method described in section 4.1.

4. Once the baseline contracts have been determined, use training data to determine cost of baseline method on the training data. This gives us B, π_{SQ} for our model.

5. Use $\hat{\ell}$ from step 2, training data, and B, π_{SQ} from step 4 as input into optimization model. Use optimization model to determine contract parameters.

6. Given test data, generate predictions and use predictions to calculate payouts from baseline and from optimal contract.

7. Calculate performance metrics on test data.

4.3 Performance Metrics

The following are the metrics we calculate on the test set. Below, N is the sample size.

Probability of loss exceeding threshold We report this metric because it is of particular interest to practitioners. This metric is motivated by the literature on poverty traps, which shows that negative shocks can have very long lasting effects for individuals, especially if these shocks bring the individual's wealth below a certain threshold. We calculate this in the following way. For each sample $\left\{\ell_1^i, \theta_1^i, \ell_2^i, \theta_2^i\right\}_{i=1}^N$ in the test set, we calculate the net loss, $\Delta \ell_j^i \triangleq \ell_j^i - I_j(\theta_j^i)$. We then create an indicator variable $p_j^i = \mathbb{1}\{\Delta \ell_j^i > \bar{\ell}\}$. The probability of loss exceeding a certain threshold is then: $P(\Delta \ell > \bar{\ell}) = \frac{1}{N}\sum_{i=1}^N (p_1^i \vee p_2^i)$. We set $\bar{\ell}$ to be the 60^{th} percentile of the loss variable, ℓ.

Difference in average net loss This is the difference in average net loss in the two insured zones. For each sample $\left\{\ell_1^i, \theta_1^i, \ell_2^i, \theta_2^i\right\}_{i=1}^N$ in the test set, we calculate the net loss, $\Delta \ell_j^i \triangleq \ell_j^i - I_j(\theta_j^i)$. We then take the average of this quantity by zone, and take the absolute value of the difference: $\left|\frac{1}{N}\sum_{i=1}^N \Delta \ell_1^i - \Delta \ell_2^i\right|$.

Required Capital We report this measure because we think it is one of the comparitive advantages of our method, and it has implications for the insurer. Higher capital requirements for the insurance translate to higher costs for the insurer, and it is cost that is not necessarily benefitting the farmers. The formula for required capital is: $K(I(\theta)) = CVaR_{1-\epsilon_P}(\sum_z I_z(\theta)) - \mathbb{E}[\sum_z I_z(\theta)]$. For the $CVaR_{1-\epsilon_P}$, we first calculate the sum of all payouts in every scenario. We use these sums to calculate the empirical $VaR_{1-\epsilon_P}(\sum_z I_z(\theta))$. We then calculate the average of all sums greater than or equal to this quantity. We calculate $\mathbb{E}[\sum_z I_z(\theta)]$ using the empirical mean. We set $\epsilon_P = 0.01$ because it is a commonly used value by regulators.

Average Cost of Insurance We report this measure to ensure that the two methods have the same (or very similar costs). This will make it easier to compare the methods. We define this to be $\frac{1}{N}\sum_{i=1}^N \sum_z I_z(\theta_z^i) + c_\kappa K$. This is the empirical average of the cost of the insurance in every scenario in the test set plus the cost of capital

4.4 Results

The performance of our contracts was sensitive to whether or not the underlying prediction model was correctly specified. When the prediction model was correctly specified, the contracts designed by our model consistently outperform the baseline contracts in two areas: equity and required capital. The contracts designed by our model tend to lead to more equitable outcomes in terms of average net loss in each zone. Furthermore, the contracts designed by our model consisently require less capital to fund than the baseline contracts. Our model provided similar protection as the baseline contracts against catastrophic loss.

A fair comparison of the two methods is harder when the prediction model is misspecified, because it becomes harder to enforce the budget constraint. Misspecification is when the model doesn't match the data generating process. Here, our prediction model is linear, so when the true data generating process is nonlinear, the prediction model is misspecified. As a result, the overall cost of the two methods differs considerably more than when the prediction model is correctly specified. In general, in the misspecified case, the contracts designed by our model offer a similar level of protection at a lower cost, and with lower capital requirements.

4.4.1 Correctly specified prediction model

No correlation Case The insurance contracts designed by our model have a slightly lower probability of staying below the specified threshold as the baseline model. Moreover, the difference in average loss between the two zones is considerably smaller in our model, and the insurance contracts designed by our model require about 12% less capital than the baseline model. We believe this happens because our model takes into account the cost of capital, whereas the baseline method ignores it. The contracts developed by our model also have a slightly lower average cost.

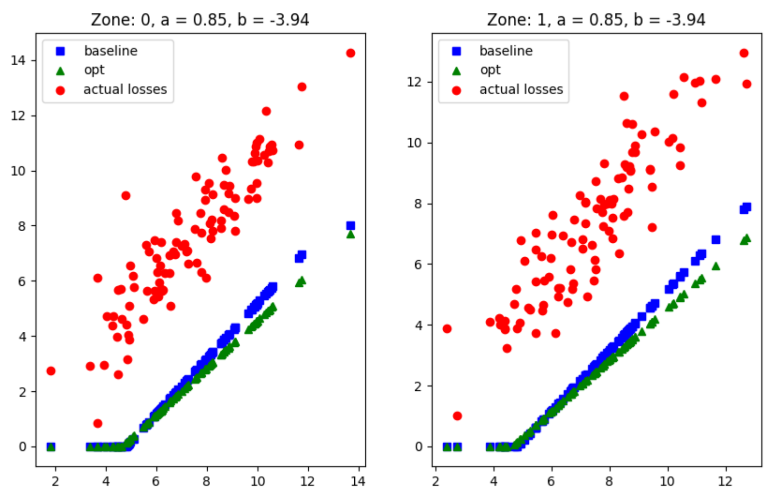

Figure 1: No Correlation, model correctly specified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled “opt”. The blue payment schedule is for the baseline contracts, labeled “baseline”. The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Model	Max CVaR	Max VaR	Max SemiVar	$\|VaR_2 - VaR_1\|$	Required Capital	Average Cost
Baseline	8.98	9.06	1.6	0.2	6.83	6.06
	[8.63, 9.4]	[8.8, 9.37]	[0.95, 2.94]	[0.02, 0.55]	[5.21, 9.01]	[4.16, 7.56]
Opt	9.17	9.03	1.24	0.16	6.2	5.8
	[8.78, 9.63]	[8.79, 9.31]	[0.74, 2.28]	[0.01, 0.47]	[4.7, 8.22]	[4.11, 7.13]
Pct Diff	-1.82	0.51	18.81	18.18	8.97	1.71
	[-6.47, 0.96]	[-1.79, 2.09]	[-5.96, 48.0]	[-566.26, 91.44]	[0.14, 18.01]	[-0.97, 17.79]

Table 1: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

Positive Correlation Case Under the insurance contracts designed by our model, farmers have a slightly lower probability of having a net loss that exceeds the specified threshold. Furthermore, the difference between the average losses between the two zones is significantly smaller relative to the baseline. The contracts designed by our model require 14% less capital than the baseline contracts. The costs of the two contracts are basically equivalent.

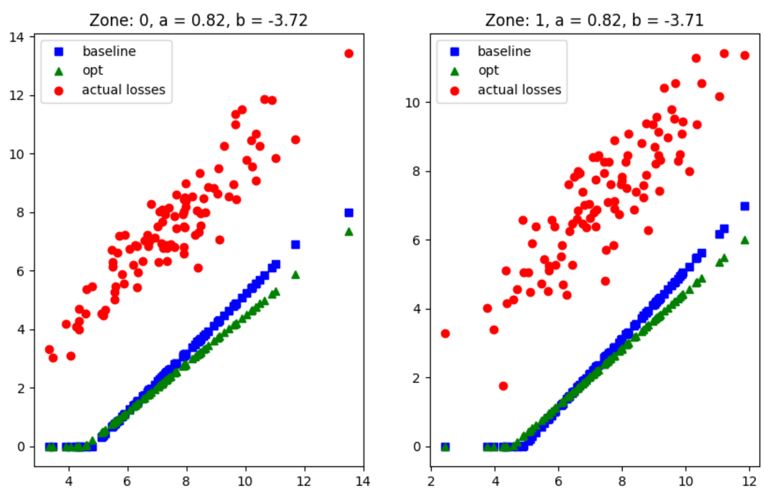

Figure 2: Positive Correlation, model correctly specified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled “opt”. The blue payment schedule is for the baseline contracts, labeled “baseline”. The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Model	Max CVaR	Max VaR	Max SemiVar	$\|VaR_2 - VaR_1\|$	Required Capital	Average Cost
Baseline	9.12	9.23	1.38	0.17	9.19	6.53
	[8.74, 9.55]	[8.93, 9.53]	[0.8, 2.5]	[0.01, 0.45]	[7.17, 10.85]	[4.64, 8.02]
Opt	9.3	9.19	1.13	0.16	8.44	6.28
	[8.9, 9.7]	[8.89, 9.47]	[0.67, 2.03]	[0.02, 0.45]	[6.48, 10.42]	[4.62, 7.58]
Pct Diff	-1.73	0.45	16.12	4.05	8.14	1.78
	[-5.09, 0.64]	[-1.5, 2.13]	[-5.47, 38.74]	[-774.0, 88.63]	[-1.58, 15.99]	[-0.98, 13.72]

Table 2: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

Negative Correlation Case Under the insurance contracts designed by our model, farmers have a slightly lower probability of having a net loss that exceeds the specified threshold. Furthermore, the difference between the average losses between the two zones is significantly smaller relative to the baseline. The contracts designed by our model require 11% less capital than the baseline contracts. The costs of the two contracts are practically equivalent.

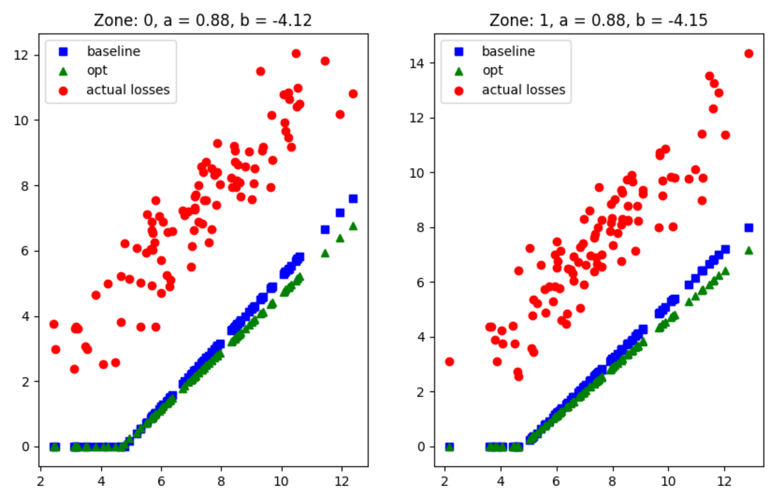

Figure 3: Negative Correlation, model correctly specified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled “opt”. The blue payment schedule is for the baseline contracts, labeled “baseline”. The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Table 3: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

Model	Max CVaR	Max VaR	Max SemiVar	$\lvert VaR_2 - VaR_1 \rvert$	Required Capital	Average Cost
Baseline	8.69	8.79	2.29	0.21	3.27	5.5
	[8.34, 9.09]	[8.56, 9.11]	[1.28, 3.86]	[0.02, 0.63]	[2.5, 4.29]	[3.85, 6.92]
Opt	8.94	8.8	1.63	0.16	3.11	5.22
	[8.55, 9.42]	[8.57, 9.11]	[0.88, 3.0]	[0.02, 0.47]	[2.39, 3.96]	[3.79, 6.55]
Pct Diff	-2.51	0.04	22.23	20.4	5.68	3.08
	[-7.96, 0.6]	[-2.57, 1.87]	[-1.76, 57.28]	[-735.88, 92.36]	[-18.8, 22.49]	[-0.39, 19.49]

4.4.2 Incorrectly specified prediction model

No correlation Case The insurance contracts designed by our model provide a similar level of protection as the baseline contracts, at a cost that is 6% lower. However, the average losses are slightly less equitable. The capital requirements for our contracts are 4.9% lower than the baseline.

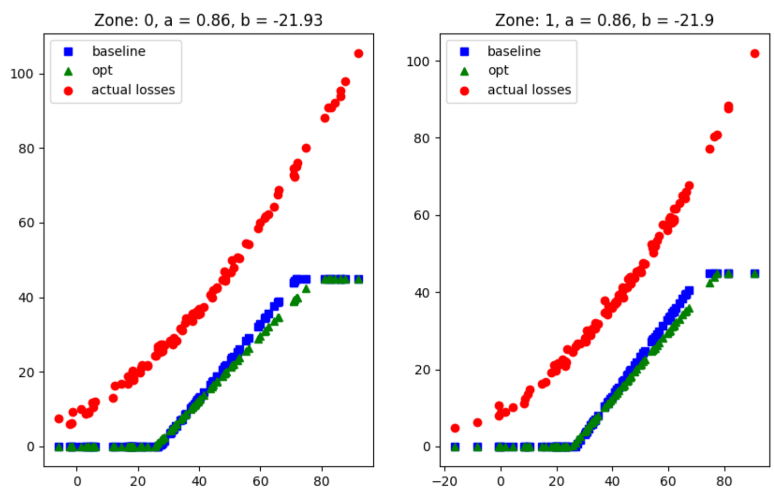

Figure 4: No Correlation, model misspecified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled "opt". The blue payment schedule is for the baseline contracts, labeled "baseline". The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Model	Max CVaR	Max VaR	Max SemiVar	$\lvert VaR_2 - VaR_1 \rvert$	Required Capital	Average Cost
Baseline	55.53	47.48	174.58	1.18	54.16	40.85
	[51.75, 60.85]	[45.75, 49.42]	[83.1, 384.27]	[0.11, 3.58]	[44.8, 59.89]	[36.53, 45.24]
Opt	54.95	46.43	217.05	1.27	51.48	40.49
	[50.63, 60.1]	[44.46, 49.15]	[101.08, 519.26]	[0.13, 3.74]	[43.14, 67.07]	[35.7, 46.32]
Pct Diff	0.91	2.38	-13.4	-2.57	3.8	0.77
	[-4.42, 7.57]	[-3.7, 5.29]	[-167.22, 16.21]	[-1041.76, 88.64]	[-31.96, 17.55]	[-5.9, 3.78]

Table 4: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

Positive Correlation Case The insurance contracts designed by our model provide a slightly worse level of protection as the baseline contracts, at a cost that is 11% lower. However, the capital requirements for our contracts are 5.7% higher than the baseline.

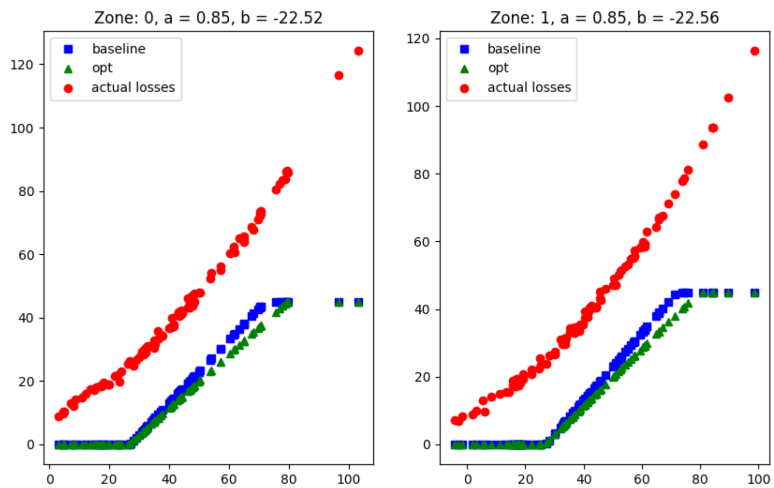

Figure 5: Positive Correlation, model misspecified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled "opt". The blue payment schedule is for the baseline contracts, labeled "baseline". The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Model	Max CVaR	Max VaR	Max SemiVar	$\|VaR_2 - VaR_1\|$	Required Capital	Average Cost
Baseline	55.19	47.52	167.32	0.77	57.29	41.25
	[51.2, 60.76]	[45.4, 49.55]	[68.24, 363.61]	[0.06, 2.23]	[50.85, 62.95]	[36.38, 46.77]
Opt	54.22	47.46	236.3	1.08	58.17	41.32
	[50.16, 59.25]	[45.16, 49.75]	[108.33, 468.21]	[0.11, 2.91]	[52.54, 65.93]	[36.19, 48.35]
Pct Diff	1.89	0.24	-37.33	-39.14	0.0	0.0
	[-2.63, 6.08]	[-3.97, 3.67]	[-139.79, 3.69]	[-1699.59, 88.14]	[-24.41, 8.99]	[-4.27, 2.05]

Table 5: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

Negative Correlation Case The insurance contracts designed by our model provide a similar level of protection as the baseline contracts, at a cost that is 7% lower. The equity of the losses is similar with both contracts. The capital requirements for our contracts are 10% lower than the baseline.

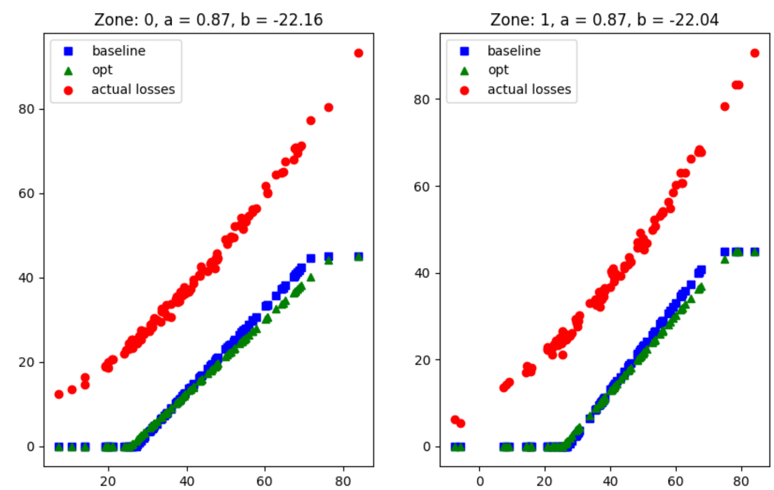

Figure 6: Negative Correlation, model misspecified. These are graphs of the payment schedules of the insurance contracts for the two zones. The green payment schedule is for the contract designed by our method, labeled “opt”. The blue payment schedule is for the baseline contracts, labeled “baseline”. The red dots are the farmer losses. a is the slope of the payout of our contract, and b is the deductible of our contract.

Model	Max CVaR	Max VaR	Max SemiVar	$\|VaR_2 - VaR_1\|$	Required Capital	Average Cost
Baseline	53.5	45.55	200.35	1.53	25.62	36.46
	[49.98, 58.97]	[44.34, 47.34]	[111.35, 409.29]	[0.12, 4.57]	[18.35, 36.23]	[33.05, 40.77]
Opt	53.08	44.39	241.91	1.43	31.59	37.35
	[48.65, 58.12]	[43.01, 46.67]	[123.16, 575.09]	[0.14, 4.08]	[19.36, 57.88]	[33.52, 43.67]
Pct Diff	1.01	2.78	-11.37	7.88	-19.55	-2.02
	[-4.69, 8.37]	[-2.72, 5.78]	[-152.03, 21.22]	[-1088.85, 90.55]	[-123.71, 19.39]	[-12.92, 2.0]

Table 6: Performance Metrics. The values shown correspond to the median value of the metric across 1000 simulation. The intervals shown are the 5^{th} and 95^{th} percentile values of the metrics.

5 Discussion

In this paper we proposed a standardized approach to design index insurance products. Our proposed method provides more equitable and cost efficient protection, and can scale to more areas. We think there are many benefits to having a standardized, optimization based approach. For one, it can be easier to incorporate supply side constraints with regards to risk and overall cost. It can also make it easier to incorporate constraints regarding the quality of the product. Both of these are significant considerations when thinking about the large scale adoption of index insurance (Jensen and C. Barrett (2017)). In the future, we hope to better incorporate uncertainty into the model, by using robust optimization. One of the factors that drives up the cost of index insurance is that there is often very little historical data for the insured regions, as a result, insurers add an uncertainty markup to protect themselves. Reinsurers behave similarly. We hope that by incorporating this data uncertainty into the problem, we can make a product that is less costly and more attractive to insurers and reinsurers.

References

Charnes, Abraham, William W Cooper, and Gifford H Symonds (1958). "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil". In: *Management science* 4(3), pp. 235–263.

Artzner, Philippe et al. (1999). "Coherent measures of risk". In: *Mathematical finance* 9(3), pp. 203–228.

Kielholz, Walter (2000). "The cost of capital for insurance companies". In: *The Geneva Papers on Risk and Insurance. Issues and Practice* 25(1), pp. 4–24.

Rockafellar, R Tyrrell, Stanislav Uryasev, et al. (2000). "Optimization of conditional value-at-risk". In: *Journal of risk* 2, pp. 21–42.

Rockafellar, R Tyrrell and Stanislav Uryasev (2002). "Conditional value-at-risk for general loss distributions". In: *Journal of banking & finance* 26(7), pp. 1443–1471.

Azariadis, Costas and John Stachurski (2005). "Poverty traps". In: *Handbook of economic growth* 1, pp. 295–384.

Lagoa, Constantino M, Xiang Li, and Mario Sznaier (2005). "Probabilistically constrained linear programs and risk-adjusted controller design". In: *SIAM Journal on Optimization* 15(3), pp. 938–951.

Barrett, Christopher B et al. (2007). "Poverty traps and climate risk: limitations and opportunities of index-based risk financing".

Nemirovski, Arkadi and Alexander Shapiro (2007). "Convex approximations of chance constrained programs". In: *SIAM Journal on Optimization* 17(4), pp. 969–996.

Osgood, Daniel E et al. (2007). "Designing Weather Insurance Contracts for Farmers in Malawi, Tanzania and Kenya: Final Report to the Commodity Risk Management Group, ARD, World Bank".

Bank, World (2011). *Weather index insurance for agriculture: guidance for development practitioners.* World Bank.

Chantarat, Sommarat et al. (2013). "Designing index-based livestock insurance for managing asset risk in northern Kenya". In: *Journal of Risk and Insurance* 80(1), pp. 205–237.

Cole, Shawn et al. (2013). "Barriers to household risk management: Evidence from India". In: *American Economic Journal: Applied Economics* 5(1), pp. 104–35.

Mobarak, Ahmed Mushfiq and Mark R Rosenzweig (2013). "Informal risk sharing, index insurance, and risk taking in developing countries". In: *American Economic Review* 103(3), pp. 375–80.

Karlan, Dean et al. (2014). "Agricultural decisions after relaxing credit and risk constraints". In: *The Quarterly Journal of Economics* 129(2), pp. 597–652.

Greatrex, Helen et al. (2015). "Scaling up index insurance for smallholder farmers: Recent evidence and insights". In: *CCAFS Report.*

Jensen, Nathaniel and Christopher Barrett (2017). "Agricultural index insurance for development". In: *Applied Economic Perspectives and Policy* 39(2), pp. 199–219.

Mapfumo, Shadreck, Huybert Groenendaal, and Chloe Dugger (2017). *Risk modeling for appraising named peril index insurance products: A Guide for practitioners.* World Bank Publications.

Casaburi, Lorenzo and Jack Willis (2018). "Time versus state in insurance: Experimental evidence from contract farming in Kenya". In: *American Economic Review* 108(12), pp. 3778–3813.

Flatnes, Jon Einar, Michael R Carter, and Ryan Mercovich (2018). "Improving the quality of index insurance with a satellite-based conditional audit contract". In: *Unpublished Working Paper.*

Jensen, Nathaniel, Quentin Stoeffler, et al. (2019). "Does the design matter? Comparing satellite-based indices for insuring pastoralists against drought". In: *Ecological economics* 162, pp. 59–73.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2020). "Subsidy policies and insurance demand". In: *American Economic Review* 110(8), pp. 2422–53.