On the benefits of index insurance in US agriculture: a large-scale analysis using satellite data

Matthieu Stigler*, David Lobell

Stanford University, Center for Food Security and the Environment

November 12 2020 (latest version here)

Abstract

Index insurance has been promoted as a promising solution for reducing agricultural risk compared to traditional farm-based insurance. By linking payouts to a regional factor instead of individual loss, index insurance reduces monitoring costs, and alleviates the problems of moral hazard and adverse selection. Despite its theoretical appeal, demand for index insurance has remained low in many developing countries, triggering a debate on the causes of the low uptake. Surprisingly, there has been little discussion in this debate about the experience in the United States. The US is an unique case as both farm-based and index-based products have been available for more than two decades. Furthermore, the number of insurance zones is very large, allowing interesting comparisons over space. As in developing countries, the adoption of index insurance is rather low — less than than 5% of insured acreage. Does this mean that we should give up on index insurance?

In this paper, we investigate the low take-up of index insurance in the US leveraging a field-level dataset for corn and soybean obtained from satellite predictions. While previous studies were based either on county aggregates or on relatively small farm-level dataset, our satellite-derived data gives us a very large number of fields (close to 1.8 million) comprised within a large number of index zones (600) observed over 20 years. To evaluate the suitability of index insurance, we run a large-scale simulation comparing the benefits of both insurance schemes using a new measure of farm-equivalent risk coverage of index insurance. We make two main contributions. First, we show that in our simulations, demand for index insurance is unexpectedly high, at about 30% to 40% of total demand. This result is robust to relaxing several assumptions of the model and to using prospect theory instead of expected utility. Second, we examine the spatial determinants of the suitability of index insurance across the 600 counties in our dataset. Our results indicate that the choice of metric to assess the suitability of insurance can lead to opposite results. When assessed against no insurance, index insurance is most beneficial in the counties with highest temporal variability. When assessed against farm-level insurance instead, index insurance is now the least beneficial in those same counties. Taken together, our results contribute towards improved policy design by shedding a more optimistic light on the overall usefulness of index insurance, and by deepening our understanding of the spatial factors constraining its spread.

*Corresponding author, `Matthieu.Stigler@gmail.com`. We thank participants at the AAEA Summer Meeting, UC Davis Big Ag Data Conference, and the Agricultural Policy Conference 2020 for useful comments. We thank also Michael Carter, Elinor Benami, Andrew Hobbs, Jon Einar Flatnes, Jisang Yu, Zara Khan and Sylvain Coutu for helpful feedback and comments.

1 Introduction

Risk is ubiquitous in agriculture. Weather has an important influence on production, yet remains difficult to predict. Likewise, agricultural prices are typically very volatile, as experienced for example during the price spike in 2007-2008. This risk has several negative consequences on farmers. In presence of risk, farmers reduce output, and opt for low-yielding low-risk technologies. Further, in developing countries with missing credit markets, risk affects negatively farmer's ability to smooth consumption, and reduces both demand and supply of credit (Boucher et al., 2008; Karlan et al., 2014).

Agricultural insurance is an important tool to reduce the risk faced by farmers. Historically, initial insurance instruments focused on indemnity-based schemes, where payouts are triggered when yields on a given field fall below a certain percentage of the field's expected yield. This *field-based* scheme suffers however from multiple issues: 1) moral hazard, where being insured leads to taking undue risk, 2) adverse selection, where the possible under and over-evaluation of individual risk leads to adverse sorting of farmers, and finally 3) high monitoring costs due to the requirement of assessing damage and the data needed for pricing individual premiums. As a response to these issues, index insurance offers an interesting alternative. Index insurance links the insurance payout to low realizations of an external index, which is often defined based on output (average yields in a given area) or inputs (weather variables such as rainfall, temperature, etc). The advantages of index insurance are reduced costs as monitoring individual fields is no longer necessary, absence of moral hazard since farmers individual actions have no influence on the index, and potentially reduced adverse selection.[1] These advantages of index insurance over traditional indemnity-based insurance have led to the implementation of several schemes throughout the world, in particular in developing countries, and to a sustained interest in the literature (see the reviews by Barnett and Mahul, 2007; Miranda and Farrin, 2012; Carter et al., 2017).

Despite the theoretical appeal of index insurance, success of the various schemes implemented is rather limited, as summarized by Binswanger-Mkhize (2012) provocative title, *Is There Too Much Hype about Index-based Agricultural Insurance?* In general, take-up is found to be very low, even at subsidized premium rates, questioning the sustainability of such schemes without public subsidies (Cole et al., 2013; Cole and Xiong, 2017). The main culprit lies in the principle itself of index insurance: by de-linking payouts from individual losses, index insurance introduces *basis risk*, i.e. the probability that a farmer experiences a loss whereas the index does not lead to a payout. Ultimately, basis risk is a function of the index accuracy, and hence depends on whether aggregate yields (for outcome-based indices) or specific rainfall variables (for input-based indices) predict well individual yields. While basis risk is widely acknowledged as the main issue of index insurance, few studies yet have been able to measure it in practice. Among the few of those, Jensen et al. (2016) analyze a livestock index insurance program in Kenya using four years of data, and conclude with a cautionary note, finding a substantial basis risk.

In this paper, we take advantage of satellite data techniques to construct a large dataset of field-level yields for corn and soybeans in the Corn Belt area of the United States of America. This enables us to conduct an in-depth analysis of basis risk, and compare the suitability of index insurance over a large number of zones. The US Corn Belt offers an interesting case study for two reasons. Firstly, its large and rather uniform fields offer a particularly favorable setting for satellite data, and accuracy of the satellite predictions is

[1] Note that adverse selection due to spatial or temporal variations in the accuracy of the index is still possible, see Jensen et al. (2018).

currently higher than in many other countries. Second, the US hosts one of the largest and possibly oldest index insurance scheme, based on county average yields. Interestingly —and somehow underappreciated in the literature— lessons from this scheme are not very encouraging as take-up is very low compared to demand for the traditional indemnity-based schemes also offered in the US. Obviously, many explanations for this low take-up of index insurance pertain to peculiarities of the US context, yet we believe that the lessons from the US case have a larger relevance in the global discussion on index insurance. In particular, the US case provides probably an upper-bound for the suitability of index insurance in general, as its relatively homogeneous production system makes it well-suited for an output-based index insurance. We would expect basis risk to be higher in developing settings characterized by larger heterogeneitydue to disparities in access to technology, information and credit.

Using satellite data provides us with a very rich dataset compared to any other study. Our dataset contains the majority of fields in each of the close to 600 counties in nine states within the US Corn Belt.[2] We observe corn and soybeans yields over a fairly long period of 20 years (2000-2019), which is long enough to comprise normal cropping years as well as exceptional events such as the 2012 drought. Keeping only fields for which we have a high classification accuracy as well as at least eight years planted to corn or soybeans, we end up with 1.8 million fields, representing 2.8 M field-crop pairs. Previous studies in the US used much shorter dataset, ranging from a few hundred fields in Miranda (1991), Smith et al. (1994), Carriker et al. (1991) to above one thousand fields in Deng et al. (2007). Barnett et al. (2005) have to our knowledge the largest number of fields in the literature —60'000 corn producers— yet these are spread out over ten states and contain hence only few fields per county, making it difficult to conduct a comprehensive basis risk analysis. In developing countries, dataset are even smaller, and contain typically one to two thousand households, covering shorter time periods given the later implementation of index insurance schemes (Jensen et al., 2016; Flatnes et al., 2018).

To analyze the suitability of index insurance, we proceed in two steps. We first run an analysis at the individual level, comparing for every field the expected utility of 1) no insurance, 2) field-level insurance and 3) index insurance. In a second step, we aggregate these measures at the county level, and using Miranda (1991)'s framework, we compute county-level measures of basis risk. We compare these various county-level metrics of index insurance suitability to county characteristics such as temporal and spatial variance, seeking to predict which counties are the most suited for index insurance. Later on, we take advantage of the spatially-explicit nature of our data to investigate whether we can redesign zones in a more natural way instead of relying on arbitrary administrative boundaries.

Our whole analysis is based on a stylized insurance scheme, where we compute ex-post fair premiums and compute the benefits of insurance assuming the farmer takes the product every single year. By doing so, we rule out moral hazard and adverse selection, and abstract from the real-world intricacies of the Federal Crop Insurance system. We do not seek either to factor in the differences in costs between the different schemes, which are hard to measure in practice. These abstractions are useful as they allow us to focus on the core question, that of basis risk and suitability of index insurance per se. This also alleviates the need to make strong assumptions and settle for specific models of moral hazard or adverse selection. Taken all together, we are probably under-estimating the benefits of index insurance from the insurer perspective, as we are not modeling the benefits of lower monitoring costs and absence of moral hazard.

[2] The states are Iowa, Indiana, Illinois, Ohio, Michigan, Minnesota, Missouri, South Dakota and Wisconsin.

The paper is organized as follows: in Section 2, we describe the Federal Crop Insurance Program, and present our modeling approach. Section 3 presents the dataset, its construction and validation. Finally, in Section 4, we show our main results.

2 Context and conceptual model

2.1 The US Federal Crop Insurance Program

The US Federal crop insurance program has become since its inception in 1938 one of the largest programs of the Farm Bill, costing close to $8 billions a year, second only to the nutrition program. These large costs can be explained by the generous nature of the program: the government covers all operational costs, and subsidizes a large share of the premiums (40-60%). These high subsidy rates are deemed necessary to induce farmers to participate into the program, given the relatively low initial participation rates in early years. Participation is now high, with about 86% of eligible acres covered in 2015.

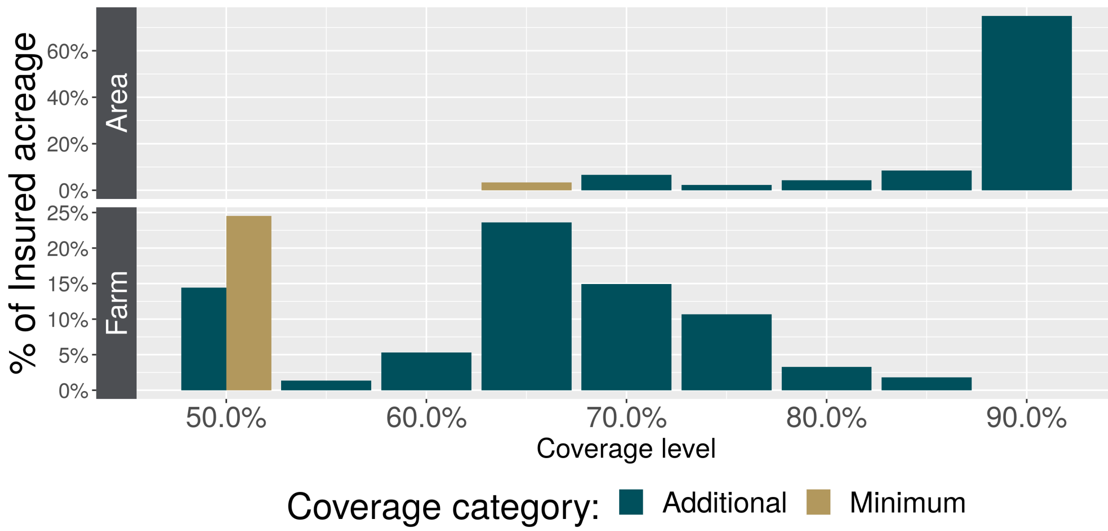
The Risk Management Agency (RMA) has offered a plethora of insurance plans throughout the years, with evolving names and specificities. In a nutshell, these can be classified into plans insuring yields or revenue, and into plans insuring at the farm-level or at the county-level. Yield insurance at the farm level was historically the standard insurance plan. Area-based plans were introduced in 1993 under the initial name of Group Risk Plan. This area-based plan is an index scheme, where the index is the average county yield as measured by official statistics collected by the US Department of Agriculture (USDA). The general idea behind all these plans is that indemnities are triggered whenever actual (farm or area) yield is below a certain percentage of its expected value. This *trigger* level (called somehow ambiguously coverage level in RMA terms) is offered at various levels, ranging from 50% to 85% for farm-level, and 65% to 90% for the area-based product. Premiums are heavily subsidized, at an average rate of 60%, with the rate decreasing for higher levels of trigger (see Table A.1). Figure 1 shows the trigger levels selected by the farmers for the farm- and area-based insurance over the 2011-2019 period averaging over corn and soybeans. The figure shows also the so-called catastrophic trigger (CAT) which comes at lower cost yet delivers lower indemnity. Strong differences appear between the farm and area-yield trigger selected. For the area-based scheme, the vast majority chooses the maximum trigger level, 90%. On the other side, for the farm-based product, farmers choose either the lowest trigger at 50%, or an intermediate value of 65%, while very few opt for the maximum coverage at 85%.[3] This difference between the trigger choice for farm- or area-based coverage suggests that area-based provides only a partial protection due to the basis risk.

Do farmers prefer area- or farm-based insurance? Figure 2 shows the percentage of each scheme in terms of total acreage covered, both for the yield and revenue types. The demand for index insurance is very small, not more than 5% in each case compared to traditional indemnity-based insurance. These results do not appear very encouraging for index insurance, casting doubt as to whether index insurance should be promoted at all. One should bare in mind however that it is difficult to compare directly the two products. Not only are subsidy rates different, but there are plenty subtle differences that we sidestepped for the sake of clarity.[4] Interestingly, a similar product with both a farm- and area-based option is offered

[3] The fact that farmers select only intermediate coverage for the farm-value has been discussed in various papers, see **?**Babcock (2015); Feng et al. (2020)

[4] Most notably, we did not discuss here the details related to the *protection price* for area-based insurance, nor the *enterprise units* for

Figure 1: Demand for insurance at various trigger levels



Source: Own computation from Risk Management Agency's Summary of Business

by the USDA Farm Service Agency, and there the conclusion is reversed: the area-based product is largely preferred over the farm-based one (Schnitkey et al., 2015). This observation motivates our approach below to evaluating the benefits of index insurance using a stylized representation abstracting from many institutional peculiarities.

2.2 Conceptual model

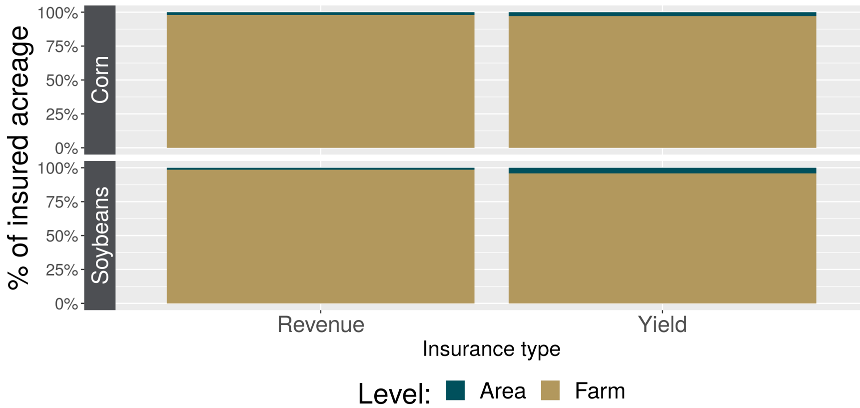
We follow here the model of Miranda (1991) measuring the benefit of area-based index insurance. We write y_{ict} as the yield for field i in county c at time t, and write the annual county average yield as $\bar{y}_{\cdot ct}$, the long-term county average yield as $\bar{\bar{y}}_{\cdot c \cdot}$, where the $\cdot$ notation indicates over which dimension the averaging is done.[5] The county-level payout is triggered whenever actual county yields $\bar{y}_{\cdot ct}$ are below their long term target $\tau_i \bar{\bar{y}}_{\cdot c \cdot}$, where τ_i is the trigger level chosen by farmer i. Miranda considers a simplified payout scheme,[6] where the indemnity is the difference between target county yields and actual county yields whenever actual county yields are below, i.e. $I^c_{ct} = \max(\tau_i \bar{\bar{y}}_{\cdot c \cdot} - \bar{y}_{\cdot ct}, 0)$. Note that for ease of exposition the indemnity is expressed in yields units, not scaled to dollars units, which are here unnecessary given that our focus is on yield, not revenue insurance. Turning to the farm-based insurance, we consider later on the same indemnity scheme, simply replacing county yields by individual field yields: $I^F_{ict} = \max(\tau_i \bar{y}_{ic\cdot} - y_{ict}, 0)$, where $\bar{y}_{ic\cdot}$ is the field-level mean. Miranda's model is based on β_{ic}, the coefficient of a regression over time of individual yields y_{ict} against county yields $\bar{y}_{\cdot ct}$:

farm-based insurance, all with different subsidy rates. Likewise, *yield exclusion* options allowing to exclude a particularly bad year from the farm-level premiums increase the attractivity of farm-level products.

[5] As an example, $\bar{y}_{\cdot ct} \equiv 1/n_{i^c} \sum_{i \in c} y_{ict}$ denotes the county mean over time.

[6] The actual indemnity scheme divides the difference by the trigger τ_i, and contains also a *protection factor*, which allows to scale up or down the indemnity payment. The RMA does to our knowledge not provide data on insurance take-up by protection factor level, so we simply set it to 100%, to ease comparison with farm-level insurance. See Skees et al. (1997) for details.

Figure 2: Demand for area versus farm-based insurance



Source: Own computation from Risk Management Agency's Summary of Business

$$y_{ict} = \alpha_{ic} + \beta_{ic}\bar{y}_{\cdot ct} + \epsilon_{ict} \tag{1}$$

Intuitively, β_{ic} indicates how well a farmer's yield is correlated to the county yield. The term ϵ_{ict} represents idiosyncratic farmer-specific shocks that cannot be insured by a county-level insurance scheme. Miranda analyses the benefits of area-based insurance using a mean-variance framework. When premiums are fair, the absolute difference in mean-variance utility between area-based and no insurance amounts to the difference in variance. Miranda shows that the variance reduction for field i, Δ_{ic}, is a function of the farmer's own β_{ic} and a county-level *critical beta* value $\tilde{\beta}_c$: $\Delta_{ic} = \sigma^2_{IC}[\beta_{ic}/\tilde{\beta}_c - 1]$, where σ^2_{IC} is the variance of the indemnity. The risk is reduced for all farmers above the critical beta, i.e. $\beta_{ic} > \tilde{\beta}_c$. Further theoretical refinements of Miranda's model focusing on the design of the optimal indemnity were made by Mahul (1999), Vercammen (2000) and Bourgeon and Chambers (2003).

The use of a mean-variance utility function is, however, somehow controversial. Jensen et al. (2016) argue in particular that the assumption of symmetry in preference between positive and negative shocks is not very relevant for the context of crop insurance, targeted at reducing negative shocks. Expected utility (EU) offers a theoretically-grounded alternative, and captures the asymmetry in preferences through the concavity of the risk aversion function. Unfortunately using a EU function does not lead to simple analytical expressions anymore. We can, however, use a second-order Taylor approximation and still obtain analytical results. The absolute difference in utility Δu_{ic} becomes now: $\Delta\,\mathrm{E}[u_{ic}] \approx -1/2u''(\mu_{ic})\Delta_{ic}$, where μ_{ic} is the expected value of the field-level mean $\bar{y}_{ic\cdot}$ and Δ_{ic} Miranda's variance reduction factor. Importantly, Miranda's result that only farmers with $\beta_{ic} > \tilde{\beta}_c$ will benefit from index insurance still holds.

Given the difficulty of obtaining analytical results in the general case, we will proceed below to a simple empirical evaluation of utility of the various insurance plans based on (simulated) yield data. Instead of expressing our comparison in utility units, we use certainty equivalents (CE), which are expressed in yield metrics. The certainty equivalent is the non-random value whose utility is the same as the expected utility from a random *lottery*, where the lottery here is simply the set of observed yields. That is, CE is the value such that $u(CE) = E[U(y)]$ holds. A higher CE is equivalent to a higher utility, and hence we simply compare index insurance versus no insurance based on their ratio CE^I/CE^{no}. A ratio >1 implies a higher utility of index insurance, $U^I > U^{no}$.

As discussed in the beginning, basis risk is often considered the main issue with index insurance. Basis risk is often defined as the risk that the farmer experiences a loss, while the index does not leads to a payout. The reverse situation of the farmer experiencing no loss yet receiving an indemnity is also possible, but usually not taken into account, as the emphasis is on the ability of an insurance scheme to reduce negative events, not to amplify positive ones.[7] Following Elabed et al. (2013) we can consider the *False Negative Probability FNP*:

$$FNP(\theta_c, \theta_i) \equiv P(\bar{y}_{\cdot ct} > \theta_c | y_{ict} < \theta_i)$$

Here θ_c is a county loss threshold, and θ_i is a farmer-specific subjective loss threshold. This measure is unfortunately unsatisfying for multiple reasons. First of all, it requires to define specific loss thresholds θ_c and θ_i, which is mostly arbitrary given that yields are a continuous variable. Second, this is only a probability between 0 and 1, and hence is not indicative of the amount of loss experienced. An insurance missing a particularly catastrophic event yet delivering payouts for all other small loss events would be deemed to offer a low basis risk despite not serving when it is the most needed.[8] We adopt here another definition of basis risk, related to the county yield regression (1). We look at the variance of residuals $\sigma^2_{\epsilon_{ic}}$ normalized by the field-specific variance, which is equivalent to $\sigma^2_{\epsilon_{ic}} / \sigma^2_{ic} = 1 - R^2_{ic}$. This represents the amount of idiosyncratic risk that can not be insured by the index. A value of 0 indicates perfect correlation with the index, while a value of 1 indicates that the variables are fully uncorrelated. In the latter case, $\beta_{ic} \to 0$ so that the reduction in variance Δ_{ic} is negative, indicating that the area insurance provides less utility than without insurance.[9]

The discussion so far focused on measuring the benefits of index insurance versus no insurance. The next question is how index insurance compares to farm-based insurance, which is a more stringent and also more informative test. Noting that no-insurance is equivalent to a farm-based insurance with a 0% trigger level, we seek to strengthen our comparison, comparing now the utility of index insurance to increasing coverage levels for the farm-based product. We name such measure the *farm-equivalent risk coverage*, which we define as the highest level of farm-based insurance for which index insurance is at least as good or better. The higher this number is, the more protection index insurance gives in terms of an ideal farm-based scheme. Formally, our measure is defined as:

$$\tau^* \equiv \max_{\tau \in \{0.2, \ldots, 0.9\}} \tau \quad \text{such that} \quad U^{area}_{90\%} > U^{farm}_{\tau\%}$$

We set the value of 90% for the area insurance as this is the value most selected by farmers, and search over a large set candidates values of $\{0.2, \ldots, 0.85, 0.95\}$ which includes all values offered by the RMA (from 0.5 to 0.85). For an index-insurance with a trigger of 90%, this measure will typically lie in the interval $[0\%, 85\%]$. We expect indeed that at an equal trigger level, a farm-based insurance at 90% will be preferred to an area-based insurance at 90%, given that the farm plan will additionally cover the idiosyncratic risk on top of the systematic covered by the index insurance. But this is not necessarily the case, and we observe

[7] It should be noted however that insurance windfalls have also an indirect negative impact by increasing premiums.
[8] See Clarke (see 2016); Barré et al. (see 2016) for an in-depth discussion of metrics for index insurance.
[9] This is true for a mean-variance utility function, as well as for any utility function up to a second-order approximation.

several cases where the area insurance at 90% is referred to a farm-based product at 90%, or even at 95%. To understand this counter-intuitive situation, think of a field perfectly correlated to the county area except for the only period when there is a large shock: say the field experiences a shock of say 84% compared to its mean, while the county average has a shock of 70% (we will assume there are enough periods so that the mean of the county is close to the field mean despite the one-time discrepancy). A farm-based product with coverage of 85%, 90% or 95% will provide an indemnity of 1%, 6% or 11% respectively. On the other side, the indemnity of the area product will be close to 20%. In this case the farmer will clearly prefer the area-based product at 90% to a farm product at 90%, and might in fact even prefer it compared to a farm product at 95%! Figure A.1 in the appendix shows an example of a field in our sample which has a farm-equivalent risk coverage of 90%. This is explained by a very high indemnity from the area product, at 110% of own yields, for the lowest yield realization (at 70%).

A limitation of our *farm-equivalent risk coverage* measure is that it is undefined in two cases. For one, if index insurance is not even as good as no insurance, then it is clear that it won't be better than any level of farm-based insurance. In this case, we attribute a τ value of 0%. The second limitation arises from the fact that we can only observe the utility of farm-level insurance for those trigger levels at which there is a yield fallout happening. If for a given field the minimum yield observed is say at 70% of the average yield, then farm plans covering 50% to 65% will not provide any protection, and hence will give the same utility as the situation without insurance. If however index insurance is inferior to the minimum observed relative yield, we only know that it lies in an interval $[0, y_{\min}/\bar{y}[$. These two limitations raise issues for the aggregation of our *farm-equivalent risk coverage* at the county level. To address these two issues, we consider rank statistics such as the median and the proportion of fields within a county which have $\tau_i > 0.85$, as well as $\tau_i > 0.50$.[10] These numbers correspond to the highest and lowest levels of farm protection available. The 50% level is also called *catastrophic* protection, so serves as a good benchmark for the minimum protection level index insurance should provide. The 85% level on the other side corresponds to the best available farm-level protection, so any field with a farm-equivalent risk coverage at 85% or above would strictly prefer area-based insurance over far-based insurance.

To measure the direct benefit of index insurance both versus no insurance and versus farm-insurance, we simply specify a-priori the same utility function for all fields, and evaluate our measures based on the yields with or without insurance. Following previous literature (Wang et al., 1998; Deng et al., 2008; Flatnes et al., 2018) we use a constant relative risk aversion (CRRA) iso-elastic utility function, with a parameter of 1.5. Fair premiums and indemnities are computed ex-post from the data. By following this procedure, we make two fundamental assumptions. Firstly, we are assuming that yields are the same whether or not the farmer takes insurance. This means that we are ruling out possible moral hazard. Second, we are computing ex-post fair premiums assuming the farmer takes the insurance every period, ruling out adverse selection. While this makes us depart from real-world characteristics in an important way, this allows us to focus on our main topic of interest, the utility of index insurance.

A defining characteristic of production in the Corn Belt is the practice of rotation between corn and soybeans (see Hennessy, 2006; Seifert et al., 2017). Given the large dataset we have, we observe almost every possible sequences of corn and soybean (and other crops), from always corn, always soy, always rotating to

[10] Using rank statistics will take care of the issue of aggregating over zero values. It will also partially address the problem of undefined values, although there is still a small percentage of fields with relative minimum above 85% (or above 55%) which would be miscounted. The total percentage of undefined values is however relatively small, between 2% and 5%.

any other intermediate combinations. This raises a problem for the computation of fair premiums for the area-based product. Our fair premiums are computed using county yields for the whole period. This means that the premiums will be fair for fields planting always corn (or always soy) over the whole period. But for other fields, the premium might be exceptionally favorable (say field is planted to corn only in drought year 2012 and receive huge indemnity) or very unfavorable (field is planted to corn every year but 2012). This brings important randomness in our data, making our comparisons blurred. To avoid this, we decide to simulate yields, providing us with a sample of corn and soy yields every year for each field. This has three further advantages. First of all, this allows us to extend the time length of our sample, which we simulate using NASS means from 1990 to 2018. Second, having more observations for each field increases the probability of observing lower minimum values for each field, attenuating the problem of undefinedness of our farm-equivalent measure, which is not defined if the observed relative minimum is too high. Finally, simulating data can be seen as a measurement error correction, where we adjust our sample to match official county means. Yield is simulated based on the field-to-county regression (1): for each field we estimate $\hat{\alpha}_{ic}$, $\hat{\beta}_{ic}$ and $\hat{\sigma}^2_{ic}$ based on the raw data. We then plug-in detrended NASS county means $\tilde{y}^{NASS}_{ct}$ from 1990 to 2018, and simulate residuals from a normal distribution $\mathcal{N}(0, \hat{\sigma}^2_{ic})$, that is $\hat{y}_{ict} \sim \mathcal{N}(\hat{\alpha}_{ic} + \hat{\beta}_{ic}\tilde{y}^{NASS}_{ct}, \hat{\sigma}^2_{ic})$. To avoid simulating outlying observations, we actually simulate using a truncated normal distribution, setting generous lower bounds of 10 [bu/acres] for both crops, and upper bounds of 100 [bu/acres] for soy and 350 [bu/acres] for corn. As a robustness check, we investigate the impact of using raw data instead in Section 4.1.

3 Data

The yield data comes from the Scalable Satellite-based Yield Mapper (SCYM) model initially developed by Lobell et al. (2015) and further refined in Jin et al. (2017). The SCYM method predict yields based on a satellite-derived vegetation index.[11] The model follows an innovative approach using an agronomic crop growth model to create a training set. In brief, the agronomic model is used to simulate multiple realizations of *pseudo* yields and *pseudo* vegetation values. These simulated pseudo values are used to train a regression between vegetation index and yields. The estimated parameters are then used to predict yield based this time on the satellite-observed vegetation index. The advantage of this method is that it does not require ground data for the training stage. When ground truth data is available, it can be used as true out-of-sample validation. When validated against ground truth data for more than twenty thousand corn fields, Deines et al. (2019a) find that the overall correlation for corn is 0.67 at the field level. Accuracy for soybeans is typically lower, between 0.5 and 0.6. Earlier versions of this dataset have been used in various studies, Lobell and Azzari (2017) look at increasing field heterogeneity over time, Seifert et al. (2018) study the effect of cover crops, Deines et al. (2019b) study the effect of conservation tillage, Stigler (2019) estimates the effect of rotation and Stigler (2018) measures the supply response to prices.

We extend here this dataset in several ways. Firstly, we use an improved version of the model based on recent developments by Deines et al. (2019a) for corn and Dado et al. (2019) for soybeans. Second, we extend the sample over time. Previous versions of the dataset were predicting yields for those pixels

[11] The model uses the so-called Green Chlorophyll Vegetation index (GCVI) which is similar in spirit to the widely known normalized difference index, NDVI.

designated as corn and soybeans by the Cropland Data Layer (CDL) from Boryan et al. (2011). The CDL crop map covers our nine states of interest in the Corn Belt, yet starts at different periods depending on the state, with starting dates ranging from 2000 to 2008. To have a consistent sample, we fill-in the missing years with the Corn-Soy Data Layer (CSDL) crop map of Wang et al. (2020). The CSDL crop map uses random forests to predict corn and soybean from 2000 onwards for the nine states we consider here. To measure the accuracy of ourresulting dataset, we aggregate the SCYM yield predictions at the county level, and compare those against NASS county means. We find a correlation between the our predicted values and official statistics of 0.81 for corn and 0.77 for soybeans.

The SCYM yield dataset is at the pixel level, which is not very relevant to our analysis here. We use a dataset of field boundaries, the Common Land Unit (CLU). This dataset was last available for the year 2008, so field shapes might have changed in the meanwhile. To address this, we select only fields for which there is a good classification agreement throughout the years. That is, we compute the frequency of the majority vote from the crop pixel classification map (CDL or CSDL), and retain only fields that have on average 85% of classification agreement. Put differently, we require that for a given field, the CDL consistently predicts either corn or soybean for 85% of the pixels throughout the years. The pixel count is done by only considering pixels within a negative 30m buffer, avoiding contamination by mixed pixels lying on the border of the field. Averaging of the SCYM yields within the field boundary is made using the same subset of interior pixels.

In addition to filtering field boundaries based on classification agreement, we also restrict the sample to consider field-crop pairs that have at least eight years of observations of corn or soybeans. This is to guarantee statistical accuracy when we estimate the field-to-county regressions. This implies that we might observe a field for only one crop or for both. Applying these two filters, we are left with 1,838,199 million fields in the 9 states we consider. Among these 1.8 millions fields, for 54% of those we observe data on both crops (i.e. we have more than ≥ 8 observations for each of corn and soybeans), while for 28% we observe only corn and 18% only for soy. The resulting dataset gives us 2,826,681 million field-crop pairs. The sample runs from 2000 to 2019, and the total size of our sample is $\sim$ 30 million field-crop-year observations.

4 Results

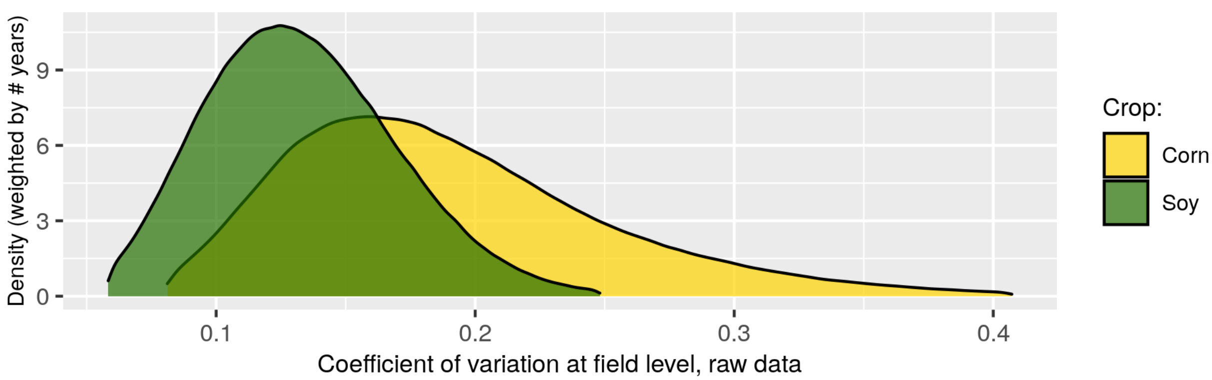
We start here by discussing some descriptive statistics of the temporal variance and the basis risk. Figure 3a shows the field-level temporal variance and 3b shows its idiosyncratic part. The temporal variance is expressed as coefficient of variation for ease of comparison across crops. The basis risk measure is the $1 - R^2$ from the field-to-county regression. While corn appears more variable than soybean, its variability is better captured by co-movement with county averages, so that the idiosyncratic variability is much lower than for soybeans. Interestingly, there is a large variation of our two measures for each crop.

In our simulated sample, average corn yields over fields and years is 162 [bu/acre], and 51.7 for soy. Given the risk-aversion function we specified, this amounts to a certainty equivalent (CE) of 155 [bu/acre] for corn and 51 [bu/acre] for soy. The difference between the average and CE, called the *risk premium*, indicates the cost of risk. Here, risk induces a welfare loss in yield metric of 7 [bu/acre] for corn and 0.6 for soy, amounting to 4.3% and 1.2% in percentage.

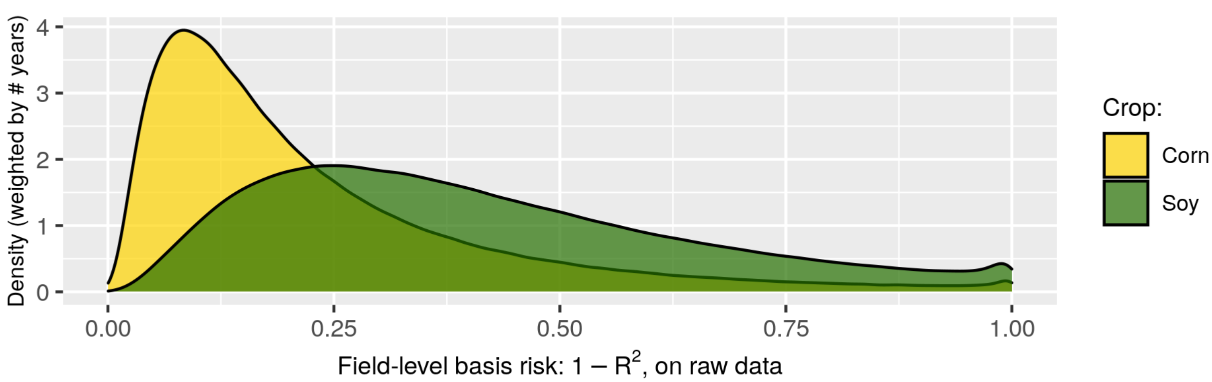
We turn now to our two main measures of the utility of index insurance, the utility of index insurance

Figure 3: Field level total and idiosyncratic variability (raw data)

(a) Coefficient of variation

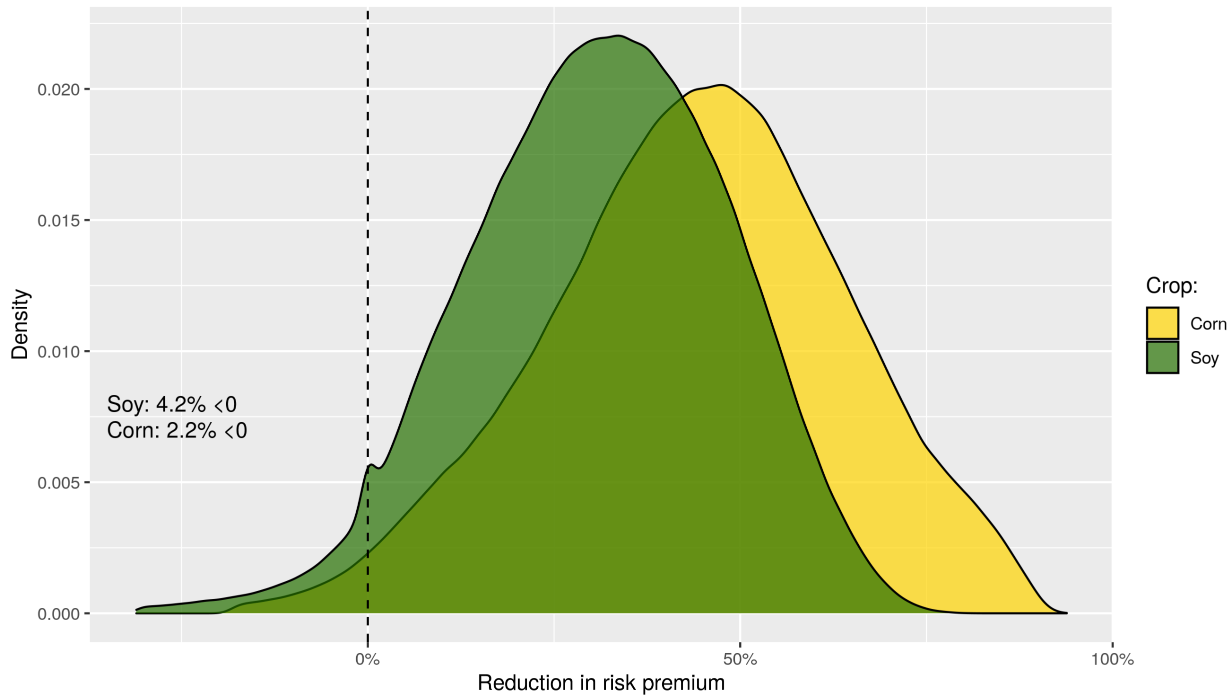


(b) Basis risk $1 - R^2$



Source: Own computation from SCYM. We kept only fields that have at least 8 observations for each crop. The density is further computed by weighting fields by the number of years available.

Figure 4: Comparing index insurance versus none: reduction in risk premium

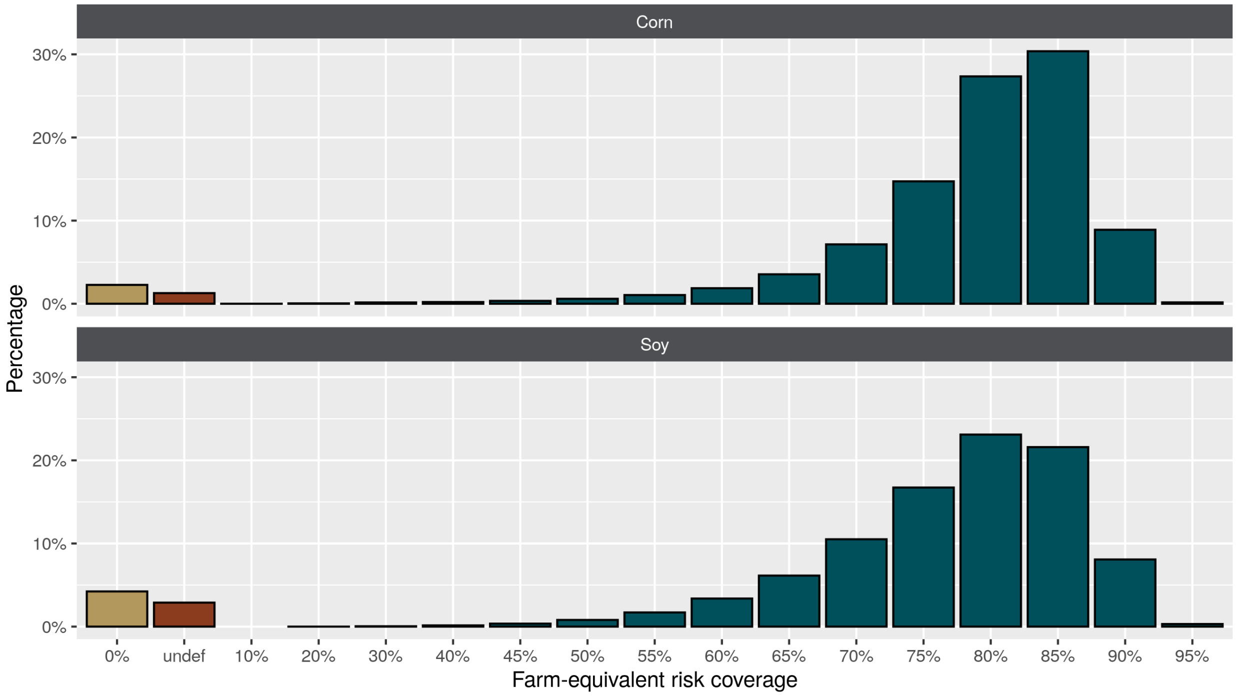


versus no insurance, and the *farm-equivalent risk coverage*. Figure 4 shows the density of the risk premium reduction. This indicates how the risk premium of no insurance is reduced with index insurance, a higher reduction meaning higher utility from index insurance. We see clearly that index insurance provides more risk reduction for corn compared to soybeans. The average reduction for corn is 43%, while that for soybeans is about 30%. The proportion of fields for which the reduction is negative, i.e. utility from index insurance is actually lower than without insurance, is also lower for corn, at 2.2% against 4.2% for soybeans. The better results for corn seems to be in line with the findings from Figure 3b, which showed that the basis risk was lower for corn.

Moving to the more stringent comparison between index insurance and farm insurance, Figure 5 shows our measure of *farm-equivalent risk coverage*. The category 0% indicates fields for which index insurance is not even as good as no insurance. These are the same percentage as the ones in Figure 4. The category *undef* corresponds to those fields for which our measure is undefined due to the fact that the utility of index insurance is lower than the smallest observed farm-equivalent category, yet higher than no insurance. Focusing on the subset of well-defined values, it is apparent again that corn provides a higher protection than soybeans. Looking at our measure of fields with at least an equivalent coverage of 85%, this number is relatively high, at 40% and 30% percent respectively. This is an important result, as it suggests that index insurance performs quite well relatively to the best available farm-level insurance level. Another interesting level to compare to is the 50% trigger, which is also the so-called *catastrophic* level offered at very low cost for farm-level insurance. There is now a large amount of fields for which index insurance is at least as good as this 50% level, 95% for corn and 92% for soybeans.[12] Note that the 90% and 95% triggers are not offered

[12]This number is possibly under-estimated due to the *undef* category, which corresponds to fields whose farm-equivalent coverage

Figure 5: Comparing index insurance versus farm insurance



by the RMA at the farm level, and we would expect that index insurance at 90% does not perform better than farm-level insurance at the same 90% trigger or even higher 95%. Computing the *farm-equivalent risk coverage* for the 90% trigger is however informative on the randomness of our simulation: had we simulated data for a longer period than 30 years, fields with index insurance better than the 90% farm-level trigger would certainly shrink rapidly.

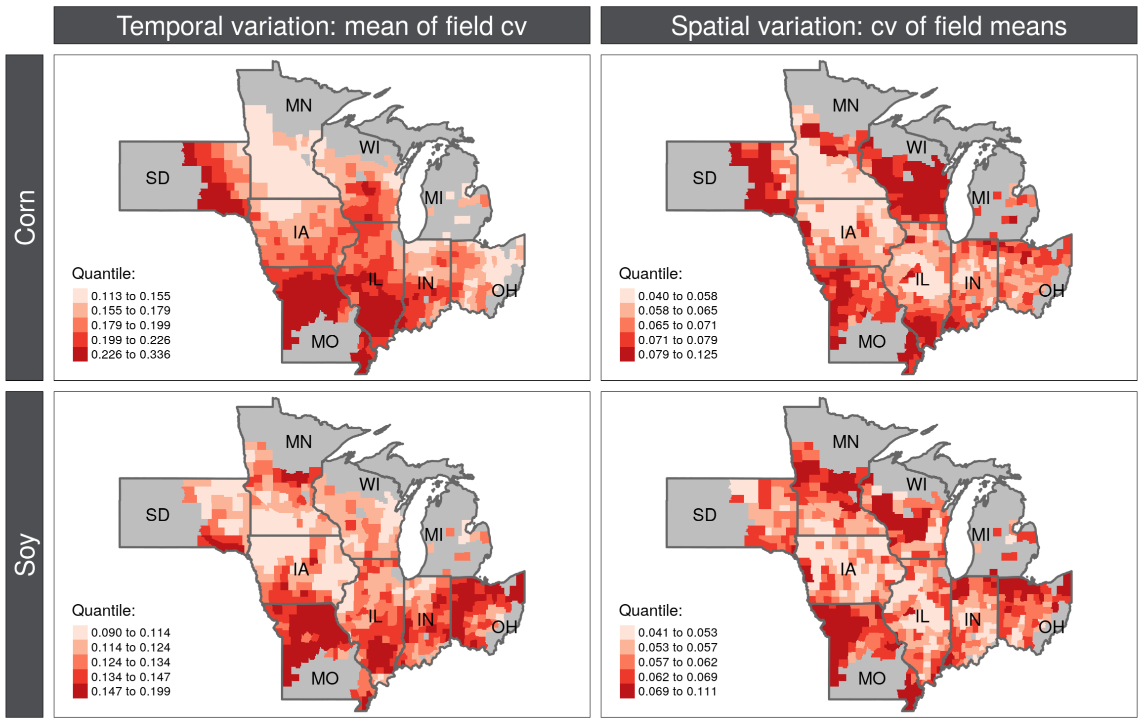
Concluding this first section, we find evidence in favor of index insurance, although the metric we consider play an important role. When doing a raw comparison versus no insurance at all, we find that a high number of fields, 98% of corn and 95% of soy fields, would benefit from index insurance. When comparing against the more stringent criterion of farm-equivalent risk coverage, we find again that for a relatively large number of fields, index insurance compares at least as well as the minimum catastrophic rate of 50% (95% and 72% for corn and soy). Finally, when subjected to the highest criterion comparing it to the maximum available farm-level trigger level of 85%, a smaller yet still large proportion of fields would still prefer index insurance, at 40% and 30%. Interestingly, this number is still much higher than the observed take-up of index insurance (see Figure 1).

We proceed now to a cross-county comparison of the suitability of index insurance, relating our measures of index insurance utility to characteristics of the 597 counties in our dataset. We start by showing the spatial and temporal variation of yields between counties. The county-average *temporal variability* is computed as the mean of every field's temporal variance.[13] The county average *spatial variability* is computed as

is identifiable only over an interval. Likely, some of these fields would have a value above 50%, yet are not counted as $< 50\%$ in this statistic.

[13]Note that this county average of field-level temporal variance $\overline{\sigma_i^2}$ is related yet distinct from the variance of the county average,

Figure 6: Corn: temporal and spatial variability

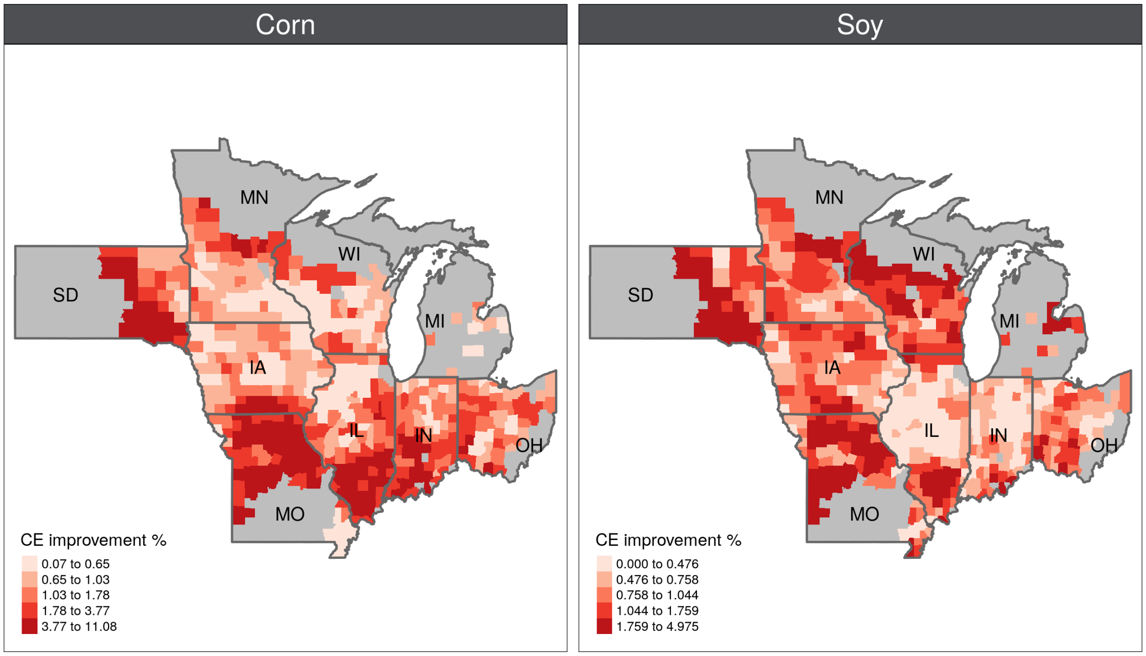


the variance of the field means. It basically indicates how different are fields within a county, and is related to the concept of local yield gap. Figure 6 shows these measures for each crop. The two measures show clear spatial patterns with a *core versus periphery* pattern, where variability is relatively low in the center of the Corn Belt, in particular in Iowa (IA), North of Illinois (IL) and South of Minnesota (MN). On the other side, bordering regions such as South Dakota (SD), Missouri, and South of Illinois (IL), Indiana (IN), Ohio, Michigan (MI) and Wisconsin (WI), have markedly higher variability. This spatial pattern is similar across crops or variables, with a correlation between variables of 0.43 for corn and 0.36 for soy, while for the same variability measure, the correlation between corn and soy is 0.5 for the temporal variation, and 0.55 for the spatial variation.

Turning now to the insurance utility measures based on the simulated data, Figure 7 compares index insurance versus none by showing the percentage difference in certainty equivalent (CE) of index insurance compared to no insurance. Interestingly, the benefit of index insurance is high in the *periphery* regions that have a high variability. It is indeed particularly high in Missouri (MO) and southern Illinois (IL), which are regions with both high temporal and spatial variability. In contrast, index insurance seems to be of more limited use in the *core* regions such as northern Iowa (IA) and southern Minnesota. Figure 8 compares on the other side the index insurance to the farm-based insurance, using our measure of *farm-equivalent risk coverage* at 85%. The conclusion is now reversed: counties in the *periphery* have a low farm-equivalent risk coverage, while those in the *core* show much higher benefits from index insurance. Most counties in Iowa (IA) have 40% or more of their fields which would strictly benefit from index insurance even compared to

$\mathrm{Var}(\bar{y}_{\cdot t}) = 1/N\overline{\sigma_i^2} + (N-1)/N\bar{\rho}$, where $\bar{\rho}$ is the average correlation among fields.

Figure 7: Utility comparison: area-plan versus no insurance



farm-based protection at 85%.

The previous results reveal an interesting reversal: counties where index insurance appears to be the most useful when assessed against no insurance turn out to be counties in which index insurance is the least beneficial in terms of farm-equivalent protection. This paradox stems from the fact that temporal and spatial variability happen to be positively correlated, yet these have opposite effects on the usefulness of index insurance. To make this point clearer, Figure 9 shows the value of index insurance (compared to no insurance or to farm insurance), projected in the spatial-temporal variability space. The index insurance utility metrics were interpolated, and red dots indicate the actual value of the near 600 counties in our sample. The first row shows the values for the utility of index insurance versus no insurance. Highest utility is found on the east side of the space, where temporal variability is highest. The gradient of utility along the temporal variability (x-axis) is so strong that it appears to obscure the effect of spatial variability (y axis), which seems to have almost no impact on the utility of index insurance. However, when looking at the utility of index insurance compared to farm insurance on the second row, results are reversed, much as we saw from maps 7 and 8. The highest utility of index insurance is now on the south-west part of the graph, where spatial and temporal variability are lowest. The east portion that was previously giving the highest utility when compared to no insurance gives now almost the lowest utility.

The reversal we document here leads to a puzzling paradox: those places where risk is highest and hence where insurance is the most needed are also those where index insurance is the least useful. Said differently, index insurance leads to good farm-equivalent coverage only in those counties that have the lowest risk. Because of the positive correlation between temporal and spatial variability, when average individual risk increases, so does the spatial variability, deteriorating the benefits of index insurance. A

Figure 8: Farm-coverage equivalent of area insurance: percentage of fields within county with coverage > 85%

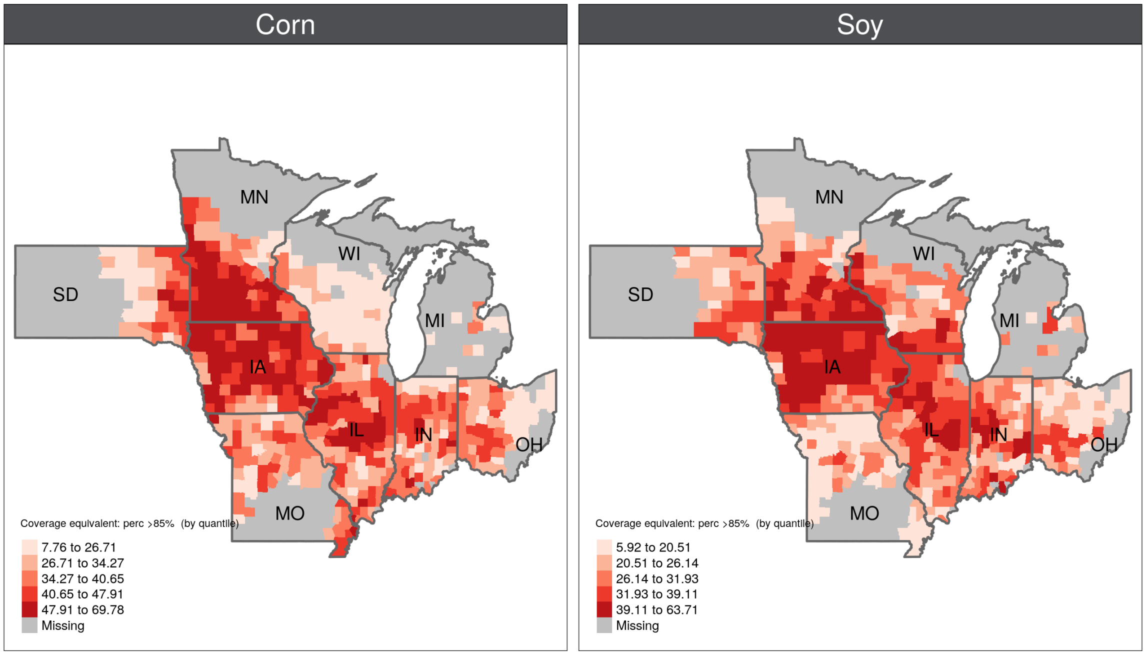
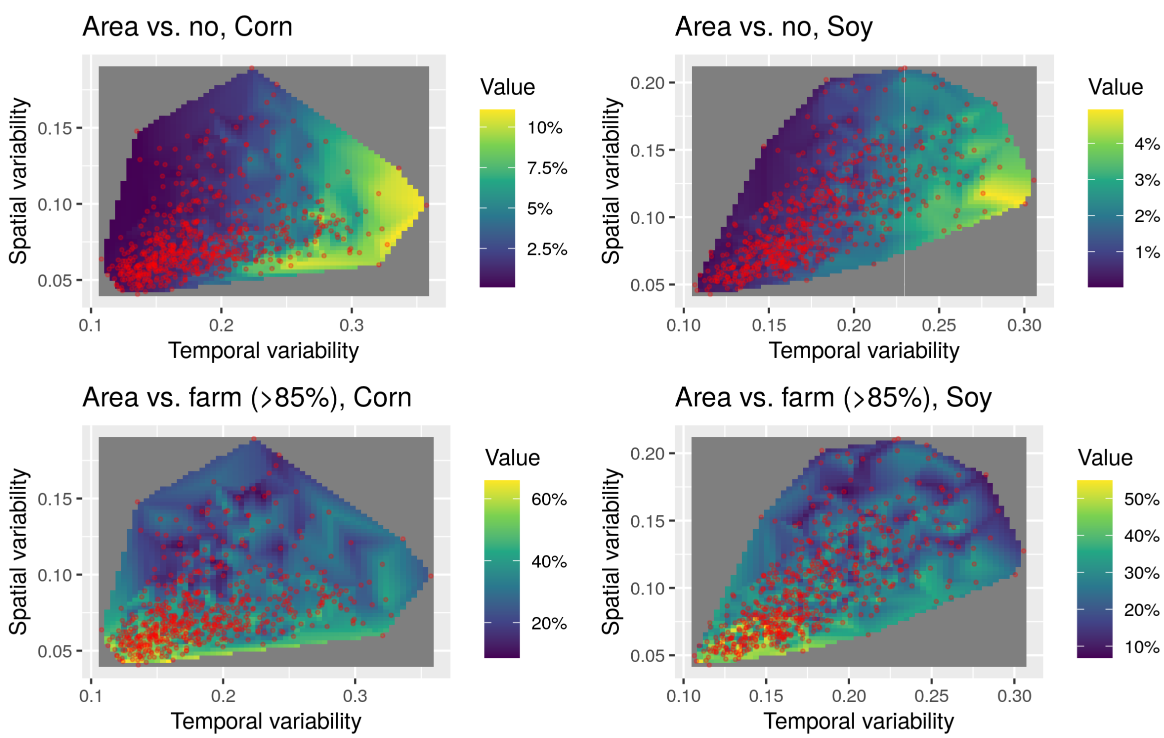


Figure 9: Utility of insurance according to county temporal and spatial variability



direct consequence of this is that selecting good zones for index insurance is a difficult task: for one, easy available statistics such as the temporal variance of the zone average[14] are potentially misleading, leading to choose zones where index insurance offers the lowest farm-equivalent coverage. What is clearly needed beyond the variance of county average is information on the spatial variability, which is much harder to obtain in practice.

4.1 Robustness checks: effect of using simulated data

Our analysis so far relied on the simulated data. The main reason was that using the raw data has a large amount of missing values due to the practice of crop rotation. For a given field, we typically only observe 50% of the corn and 50% of the soy yields over a given period. This implies that even if the premiums are fair at the county level (i.e. the premium corresponds to the average of the indemnities), thy might be very unfair of very unfavorable for a given field depending on the crop sequence. Imagine indeed that a field cultivated corn only in the drought year 2012, receiving a huge indemnity yet paying a small premium once. On the other hand, a field might cultivate corn every year buy 2012, paying the premium every year yet not receiving the large indemnity for 2012. Using the simulated data allowed us to avoid this randomness, by filling-in missing years. A second reason for using simulated data was that it would extend the sample over time, and usually also extend the empirical support of the yield distribution. This is particularly useful given that our measure of farm-equivalent risk coverage is not defined beyond the empirical minimum.[15] Extending the range of the data hence reduces the number of undefined cases.

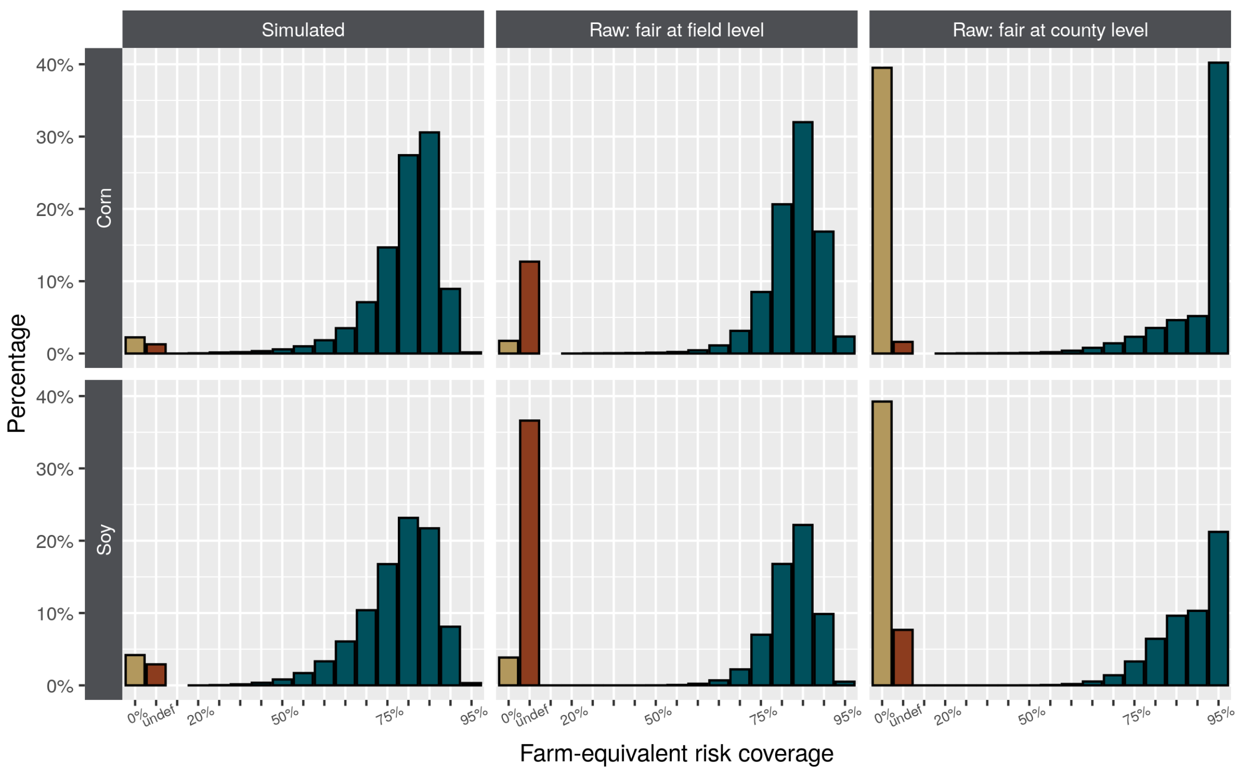
We investigate now the effect of using the raw data instead of the simulated one. We do this in two steps. In the first step, we isolate the effect of the possible *unfairness* at the field, and construct premiums that are fair at the field level. That is, we assume the insurer is computing a field-specific premium, taking into account only the years in which a field is planted to the specific crop. In a second step, we relax that assumption, and use premiums that are fair at the county level.

Figure 10 shows the histogram of our *farm-equivalent risk coverage* metric, using the simulated data (hence corresponding to Figure 5 above), the raw data with premiums fair at the field level or at the county level. Looking first at the raw data with premiums fair at the field level (second column), we see that the main difference lies in the *undef* category, which corresponds to fields for which index insurance was higher than no insurance (0%), yet lower than the level of the lowest-observed farm insurance, and hence is not clearly defined. Abstracting from this category, results look qualitatively similar. Turning to the last column showing fields with county-fair premiums, we see a striking increase in fields that either have a 0% equivalent coverage, or 100%. This illustrates well the problem of premiums being "over-" and "under-fair" premiums depending on whether or not a field planted the crop in the bad year 2012. Indeed for corn, among those fields for whom we find a 0% farm-equivalent value, 92% of those did not plant to corn that year, while among those that have the maximum 95% farm-equivalent value, 95% planted corn in 2012. For soybeans, the influence of the year 2012 is less strong, which is partly explained by the fact that the impact of the 2012 drought on yields was lower compared to corn.

[14]Remember that our measure of temporal variability used here is derived from the average field-level variance, which is not equal to the variance of the average.

[15]Remember that if the minimum of a field is say 80% of its mean, the utility of a coverage at 75% is undefined.

Figure 10: Effect of using simulated versus raw data



4.2 Robustness check: simulating using longer sample sizes

In the main analysis we used simulated data over a period of 29 years. Compared to other studies, this is rather a small number: Ye et al. (2020) use for example 1000 simulations for each farm. Our choice for 29 years is the result of a trade-off between computational burden on one side and accuracy of the risk calculation on the other side. Noting that our sample has more than 1.8 million (M) fields and 2.8M field-crop pairs, using 29 years of data already results into a dataset of more than 80M rows.

We investigate here nevertheless the impact of increasing the sample size of the simulated yields series. Note that the initial simulated data was simulating using (detrended) NASS county aggregates from 1990 to 2018 plugged into the field-county regression (1). This approach cannot be used anymore for larger series, so we proceed instead to simulating NASS data itself. To do so, we estimate an AR(2) model on each individual series, and predict from this model. Innovations for the predictions are taken from the empirical distribution of the residuals of the fitted AR model. We opt for the empirical distribution for the innovations, as it allows to capture large negative shocks that would be difficult to model with standard parametric distributions.

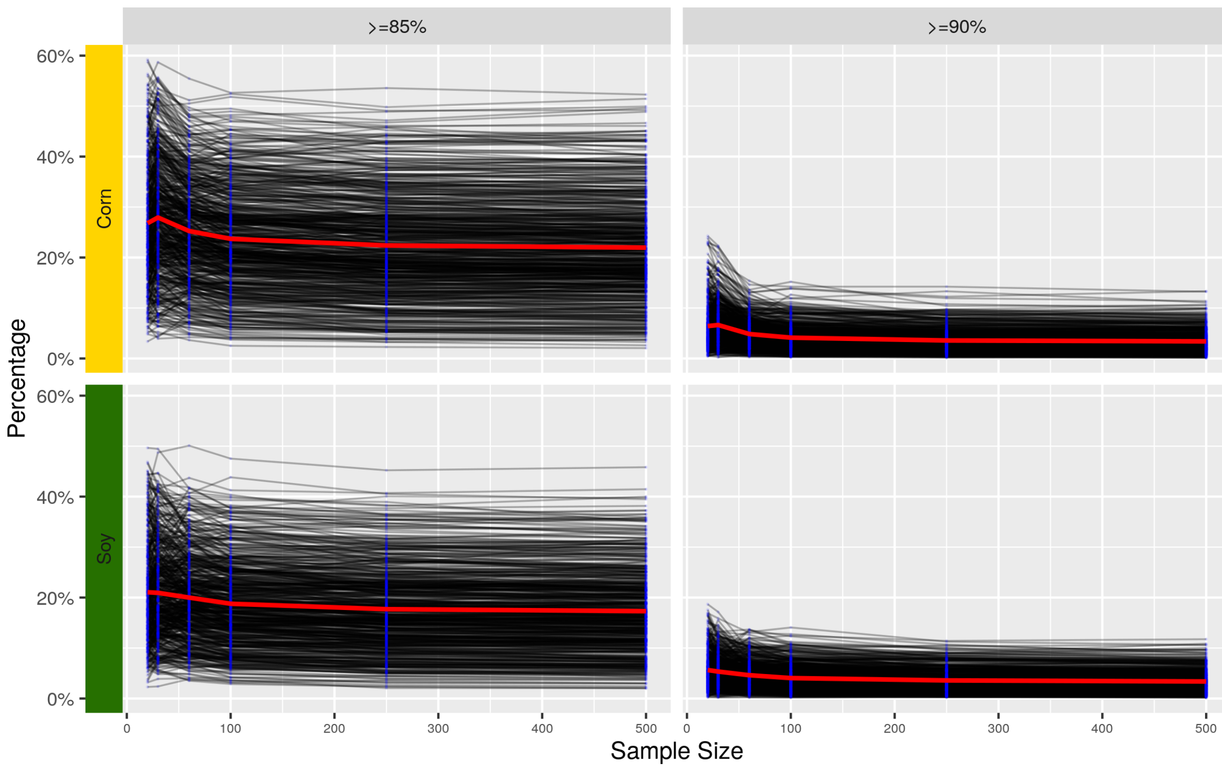
Figure 11 shows the county-level measure of the percentage of fields with index insurance at least as good as 85%,as well as 90%. The red line indicates the average over all the counties. Blue dots represent the sample sizes taken into consideration: 20, 30, 60, 100, 250 and 500. Results indicate that with larger sample sizes, our measure of the benefits of index insurance decreases, yet get stabilized relatively quickly for sample sizes of 100 or larger. This suggests that there is a small bias in our estimates on the order of 5%. The second panel shows the results for the percentage of fields for which index insurance is at least as good as a 90% farm-insurance. Remember that the coverage for the index insurance has been selected at 90% throughout this paper. Intuitively, we would expect that at equal coverage, farm-level insurance does better than area-based. Presence of fields for which area schemes are nevertheless better than farm-based was assumed so far to arise from simulation noise. Figure 11 partly confirms this intuition: the percentage decreases with bigger sample size. However, it does not goes to zero, indicating the presence of fields which always prefer area-based insurance.

4.3 Robustness check: taking subsidies into consideration

In the main analysis, we considered fair premiums. However, premiums are actually heavily subsidized by the Federal Government. Subsidies vary depending on the level of coverage and the type of scheme, see Table A.1. The subsidies start at around 60% for the lowest level of coverage, and decrease to 40-50% for the highest ones. Subsidies are higher for the area-based scheme, in particular for higher coverage levels. This suggests that taking subsidies into account should increase the attractiveness of area-based insurance compared to our benchmark.

To verify the impact of subsidies, we re-run the same analysis as above, this time applying the subsidies to the premiums. Without surprise, our two metrics of the utility of index insurance both increase with subsidies for almost all fields ($> 99\%$). As a consequence, the aggregate number of fields preferring index insurance to no insurance or to the highest farm-level insurance also increase.

Figure 11: Effect of using simulated versus raw data



4.4 Robustness check: using cumulative prospect theory

The analysis so far was based on the expected utility (EU) framework. But several authors have pointed out the widespread under-coverage observed in practice cannot be explained by expected utility (Babcock, 2015; Feng et al., 2020). According to EU theory, farmers should seek maximal coverage with fair premiums. Du et al. (2017) develop a framework to include subsidies in expected utility computations, yet find that this still does not explain the low coverage chosen in practice. Babcock (2015) uses instead cumulative prospect theory (CPT), finding that it captures better the observed behavior of the three farms he considers. Cumulative prospect theory (Tversky and Kahneman, 1992) allows to capture phenomena like loss aversion, probability weighting and reference dependence. Reference dependence refers to the existence of a *reference* point below which outcomes are considered losses, and above which values are considered gains. It is not obvious what this reference point should be for farmers' choice of crop insurance. We follow here Babcock (2015), and consider two possible reference points.[16] The first one includes the expected yield plus the premium. The second uses only the expected yield, considering that premiums are considered a sunk cost. For the choice of the value and decision weighting functions, we use exactly the same functions and parameters as in Babcock (2015), who used values directly derived from (Tversky and Kahneman, 1992). Like Babcock (2015), we use the empirical distribution and hence assign weights $1/T$ to each yield outcome.

We re-run the analysis evaluating now our metrics of index insurance with the cumulative prospect theory functions. Figure 12 shows the distribution of the farm-equivalent coverage with the standard expected utility (CRRA), as well as the CPT with the two reference points. *R1* refers to the point including expected yields plus premium and *R2* to expected yields only. Using CPT induces a quite sizable reduction on the benefits of index insurance, in particular using R1. The percentage of fields which do not benefit at all from index insurance increases with R1. On the other side, the percentage of fields which would be indifferent or prefer index insurance to farm insurance decreases under both reference points. That number falls to a 30% to 40% under expected utility to a 10-15% under CPT. This is an interesting results, as it seems more realistic than the 40% predicted under expected utility. Nevertheless, that number is still quite larger than the take-up observed in practice, which is never more than 5%.

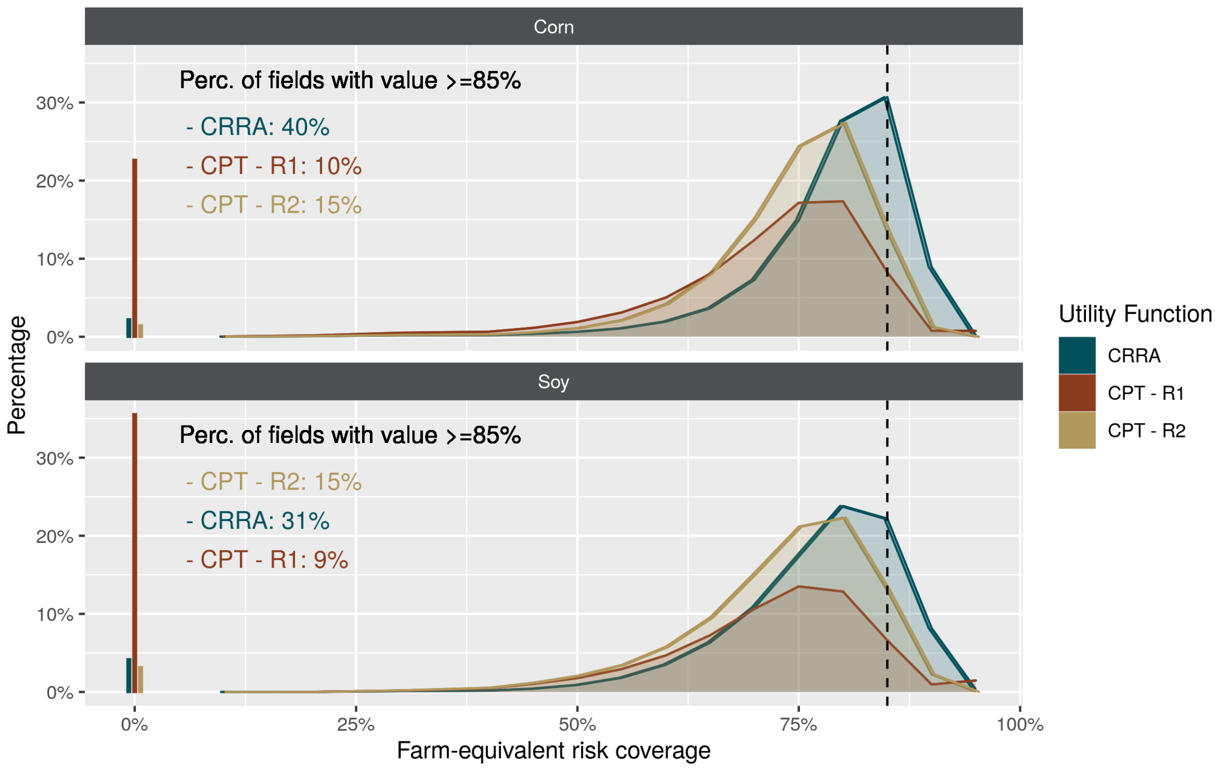
5 Conclusion

In this study, we investigate the suitability of crop insurance in the US using a unique dataset of nearly two million fields observed over 20 years through satellite remote sensing. We run a large-scale simulation seeking to replicate observed yields as closely as possible, yet abstracting from moral hazard or adverse selection issues. We develop several metrics of suitability of index insurance based on expected utility theory, comparing index insurance to no insurance but also to farm-level insurance. Thanks to the very large scale of our dataset spanning close to 600 index insurance *zones*, we are also able to investigate the characteristics of the counties which make insurance more beneficial.

Our current results bring a new positive light for index insurance. Our simulations show that absent adverse selection and moral hazard, index insurance brings a positive improvement for almost all fields.

[16] We do not consider his third reference point which is based on indemnities only. This is because with fair premiums, considering indemnities on their own amounts to choosing a lottery that increases risk yet has zero expected gain. No farmers would want such a lottery.

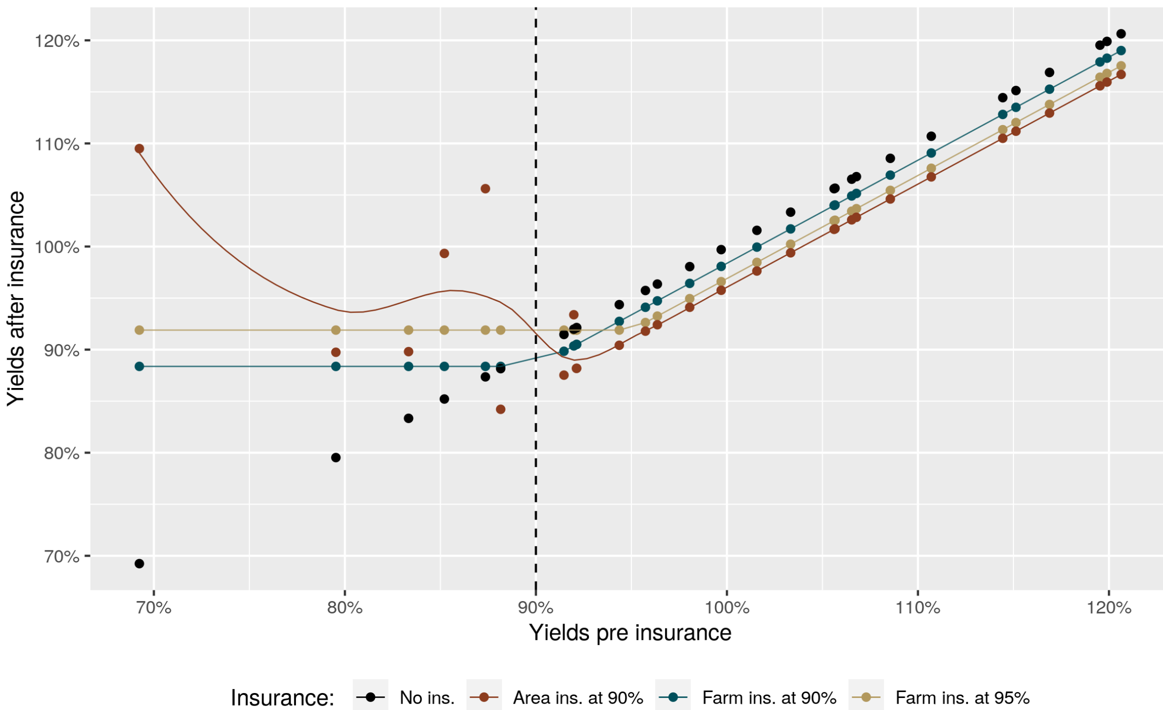
Figure 12: Effect of using cumulative prospect theory



When expressed in our new measure of farm-equivalent coverage, index insurance is at least as good as a 50% farm coverage for a majority of fields, indicating that it can serve the basic function of protecting against catastrophic events. Furthermore, when assessed against the highest-available level of 85%, 30% of the fields still benefit more from an index insurance at the 90% coverage level. Our results are robust to relaxing several assumptions of the model, although they tend altogether to slightly reduce the benefits observed. The largest changes stem from using cumulative prospect theory instead of expected utility. This is an interesting result that will deserve further analysis. Our second contribution is to uncover a small paradox in the spatial determinants of the suitability of index insurance. We show that the metric chosen to assess those benefits plays a crucial role, and similar metrics can lead to opposite results. When assessed against no insurance, index insurance seems to be the most beneficial in the counties in the outer Corn Belt, which have a higher temporal variability. On the other hand, when assessed against farm-based insurance, index insurance is now the most beneficial in the counties in the core of the Corn Belt, which have the lowest temporal variability. This result is explained by the fact that temporal and spatial variability tend to be correlated at the county level. While temporal variance increases the benefit of index insurance (as it does for any insurance), spatial variability reduces it. We believe that these new results have a relevance outside of the Corn Belt, and hopefully will help design better index insurance products in other settings with less data available.

This study could be extended in several ways. For one, we assumed away adverse selection and moral hazard, and relaxing each of these assumptions would be interesting on its own, although it will require in turn to make stronger assumptions for modeling each component. In this regard, it is important to emphasize that we constructed the premiums to be fair ex-post, which implies that no adverse selection is possible. Predicting ex-ante premiums, following the large literature based on Harri et al. (2011), would be a worthwhile extension, opening the door to models of adverse selection following Just et al. (1999). Our results showing the different predictions obtained from cumulative prospect theory are also very promising. Those could be extended to the question of the coverage of farm-level insurance, extending the work by Babcock (2015). Finally, multiple improvements could be on the methodological side when we seek to model to correlation between yields. This is definitely a high-dimension problem given that we have at most twenty time periods yet hundreds of even a few thousand of variables. While there exist several techniques to model covariance matrices in very large dimension, less has been done when there are missing values. We are currently working on integrating the sparsity shrinkage method by **?**, which would be used on the residuals covariance matrix from the initial regression. This would correspond to the approximate factor model with observed factor developed by Fan et al. (2011).

Figure A.1: Yield with and without insurance: illustration for a single field



A Appendix

A.1 Supplementary figures

Table A.1: Subsidy rate for farm- and area-based plans, yield protection

		Subsidy rate	
Coverage type	Coverage Level	Farm yield	Area yield
Catastrophic	50%	100%	
Additional	50%	67%	
	55%	64%	
	60%	64%	
	65%	59%	
	70%	59%	59%
	75%	55%	59%
	80%	48%	55%
	85%	38%	55%
	90%	-	51%

Source: RMA Insurance Handbook

References

Babcock, B. A. (2015): "Using Cumulative Prospect Theory to Explain Anomalous Crop Insurance Coverage Choice," *American Journal of Agricultural Economics*, 97, 1371–1384.

Barnett, B. J., J. R. Black, Y. Hu, and J. R. Skees (2005): "Is Area Yield Insurance Competitive with Farm Yield Insurance?" *Journal of Agricultural and Resource Economics*, 30, 1–17.

Barnett, B. J. and O. Mahul (2007): "Weather Index Insurance for Agriculture and Rural Areas in Lower-Income Countries," *American Journal of Agricultural Economics*, 89, 1241–1247.

Barré, T., Q. Stoeffler, and M. Carter (2016): "Assessing index insurance: conceptual approach and empiricalillustration from Burkina Faso," Tech. rep., University of California Davis.

Binswanger-Mkhize, H. P. (2012): "Is There Too Much Hype about Index-based Agricultural Insurance?" *The Journal of Development Studies*, 48, 187–200.

Boryan, C., Z. Yang, R. Mueller, and M. Craig (2011): "Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program," *Geocarto International*, 26, 341–358.

Boucher, S., M. Carter, and C. Guirkinger (2008): "Risk Rationing and Wealth Effects in Credit Markets: Implications for Agricultural Development," *American Journal of Agricultural Economics*, 90(2), 409–423.

Bourgeon, J.-M. and R. G. Chambers (2003): "Optimal Area-Yield Crop Insurance Reconsidered," *American Journal of Agricultural Economics*, 85, 590–604.

Carriker, G. L., J. R. Williams, G. A. Barnaby, and J. R. Black (1991): "Yield and Income Risk Reduction under Alternative Crop Insurance and Disaster Assistance Designs," *Western Journal of Agricultural Economics*, 16, 238–250.

Carter, M., A. de Janvry, E. Sadoulet, and A. Sarris (2017): "Index Insurance for Developing Country Agriculture: A Reassessment," *Annual Review of Resource Economics*, 9, 421–438.

Clarke, D. J. (2016): "A Theory of Rational Demand for Index Insurance," *American Economic Journal: Microeconomics*, 8, 283–306.

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery (2013): "Barriers to Household Risk Management: Evidence from India," *American Economic Journal: Applied Economics*, 5, 104–35.

Cole, S. A. and W. Xiong (2017): "Agricultural Insurance and Economic Development," *Annual Review of Economics*, 9, 235–262.

Dado, W., J. M. Deines, and D. B. Lobell (2019): "Improving satellite-based soybean yield mapping across irrigated and rain-fed conditions," in *American Geophysical Union, Fall Meeting*.

Deines, J. M., W. Dado, R. Patel, and D. B. Lobell (2019a): "Insights into Effective Satellite Crop Yield Estimation from an Extensive Ground Truth Dataset in the US Corn Belt," in *American Geophysical Union Fall Meeting*.

Deines, J. M., S. Wang, and D. B. Lobell (2019b): "Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt," *Environmental Research Letters*, 14, 124038.

Deng, X., B. J. Barnett, G. Hoogenboom, Y. Yu, and A. G. y. Garcia (2008): "Alternative Crop Insurance Indexes," *Journal of Agricultural and Applied Economics*, 40, 223–237.

DENG, X., B. J. BARNETT, AND D. V. VEDENOV (2007): "Is There a Viable Market for Area-Based Crop Insurance?" *American Journal of Agricultural Economics*, 89, 508–519.

DU, X., H. FENG, AND D. A. HENNESSY (2017): "Rationality of Choices in Subsidized Crop Insurance Markets," *American Journal of Agricultural Economics*, 99, 732–756.

ELABED, G., M. F. BELLEMARE, M. R. CARTER, AND C. GUIRKINGER (2013): "Managing basis risk with multiscale index insurance," *Agricultural Economics*, 44, 419–431.

FAN, J., Y. LIAO, AND M. MINCHEVA (2011): "High-Dimensional Covariance Matrix Estimation In Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356.

FENG, H., X. DU, AND D. A. HENNESSY (2020): "Depressed demand for crop insurance contracts, and a rationale based on third generation Prospect Theory," *Agricultural Economics*, 51, 59–73.

FLATNES, J. E., M. R. CARTER, AND R. MERCOVICH (2018): "Improving the Quality of Index Insurance with a Satellite-based Conditional Audit Contract," Tech. rep., Basis, Feed the Future Innovation Lab for Markets, Risk and Resilience, UC Davis.

HARRI, A., K. H. COBLE, A. P. KER, AND B. J. GOODWIN (2011): "Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating," *American Journal of Agricultural Economics*, 93, 707–717.

HENNESSY, D. A. (2006): "On Monoculture and the Structure of Crop Rotations," *American Journal of Agricultural Economics*, 88, 900.

JENSEN, N. D., C. B. BARRETT, AND A. G. MUDE (2016): "Index Insurance Quality and Basis Risk: Evidence from Northern Kenya," *American Journal of Agricultural Economics*, 98, 1450–1469.

JENSEN, N. D., A. G. MUDE, AND C. B. BARRETT (2018): "How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya," *Food Policy*, 74, 172 – 198.

JIN, Z., G. AZZARI, AND D. B. LOBELL (2017): "Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches," *Agricultural and Forest Meteorology*, 247, 207 – 220.

JUST, R. E., L. CALVIN, AND J. QUIGGIN (1999): "Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives," *American Journal of Agricultural Economics*, 81, 834–849.

KARLAN, D., R. OSEI, I. OSEI-AKOTO, AND C. UDRY (2014): " Agricultural Decisions after Relaxing Credit and Risk Constraints *," *The Quarterly Journal of Economics*, 129, 597–652.

LOBELL, D. B. AND G. AZZARI (2017): "Satellite detection of rising maize yield heterogeneity in the U.S. Midwest," *Environmental Research Letters*, 12, 014014.

LOBELL, D. B., D. THAU, C. SEIFERT, E. ENGLE, AND B. LITTLE (2015): "A scalable satellite-based crop yield mapper," *Remote Sensing of Environment*, 164, 324 – 333.

MAHUL, O. (1999): "Optimum Area Yield Crop Insurance," *American Journal of Agricultural Economics*, 81, 75–82.

MIRANDA, M. AND K. FARRIN (2012): "Index Insurance for Developing Countries," *Applied Economic Perspectives and Policy*, 34, 391–427.

MIRANDA, M. J. (1991): "Area-Yield Crop Insurance Reconsidered," *American Journal of Agricultural Economics*, 73, 233–242.

SCHNITKEY, G., J. COPPESS, N. PAULSON, AND C. ZULAUF (2015): "Perspectives on Commodity Program Choices Under the 2014 Farm Bill," *farmdoc daily*, 5:111.

SEIFERT, C. A., G. AZZARI, AND D. B. LOBELL (2018): "Satellite detection of cover crops and their effects on crop yield in the Midwestern United States," *Environmental Research Letters*, 13, 064033.

SEIFERT, C. A., M. J. ROBERTS, AND D. B. LOBELL. (2017): "Continuous Corn and Soybean Yield Penalties across Hundreds of Thousands of Fields," *Agronomy Journal*, 109, 541–548.

SKEES, J. R., J. R. BLACK, AND B. J. BARNETT (1997): "Designing and Rating an Area Yield Crop Insurance Contract," *American Journal of Agricultural Economics*, 79, 430–438.

SMITH, V. H., H. H. CHOUINARD, AND A. E. BAQUET (1994): "Almost Ideal Area Yield Crop Insurance Contracts," *Agricultural and Resource Economics Review*, 23, 1–9.

STIGLER, M. (2018): "Supply response at the field-level: disentangling area and yield effects," Tech. rep., UC Davis, ARE.

——— (2019): "Measuring rotation effects in the US Corn Belt," Tech. rep., Chapter 2 of dissertation, `https://github.com/MatthieuStigler/MatthieuStigler.github.io/raw/master/docs/rotation_effects_Stigler_standalone.pdf`.

TVERSKY, A. AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

VERCAMMEN, J. A. (2000): "Constrained Efficient Contracts for Area Yield Crop Insurance," *American Journal of Agricultural Economics*, 82, 856–864.

WANG, H. H., S. D. HANSON, R. J. MYERS, AND J. R. BLACK (1998): "The Effects of Crop Yield Insurance Designs on Farmer Participation and Welfare," *American Journal of Agricultural Economics*, 80, 806–820.

WANG, S., S. DI TOMMASO, J. M. DEINES, AND D. LOBELL (2020): "Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive," *Scientific Data*, 7, 307.

YE, T., W. HU, B. J. BARNETT, J. WANG, AND Y. GAO (2020): "Area Yield Index Insurance or Farm Yield Crop Insurance? Chinese Perspectives on Farmers' Welfare and Government Subsidy Effectiveness," *Journal of Agricultural Economics*, 71, 144–164.