The Market Potential for Area–Yield Crop Insurance: An Application to Maize in Ghana*

Ashish Shenoy[1C] and Mira Korb[1]

[1]University of California, Davis

[C]Corresponding author.
Email: shenoy@ucdavis.edu
Mail: 1 Shields Ave, Dept of Ag & Res Econ, Davis, CA 95818

Abstract

Area–yield crop insurance offers an appealing alternative to weather-based insurance because indemnities more closely target production shortfalls while still mitigating information asymmetry. These features are inversely related: larger insurance zones limit index manipulation, but average yield is less informative about any given plot. We present a framework to quantify this tradeoff and apply it to maize in Ghana using a spatial yield model calibrated to match observed production. We find area–yield insurance would require zones of no more than 7,000 farmers for the index to outperform weather insurance. Collusion would be difficult to sustain at this scale, confirming the market viability of area–yield insurance. The framework presented in this paper readily adapts to assess market potential for new crop insurance products.

Keywords: Agricultural insurance, area–yield insurance, basis risk, maize, Ghana

JEL Codes: G22, O13, Q14

*This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID, the University of California, or the United States Government.
We thank Richard Gallenstein, Andrew Hobbs, and Matthieu Stigler for helpful advice and feedback.
Authors declare we have no conflicts of interest. No institution had the right to review results before publication.

1 Introduction

Agricultural production risk remains a salient barrier to rural development. Crop insurance can insulate farm households from risk, but directly insuring individual on-farm yield invites adverse selection and moral hazard (Gunnsteinsson, 2020). To prevent market unraveling, insurers base indemnities on environmental factors such as rainfall.

Weather-based insurance has been shown to promote investment and prevent decapitalization in field trials (see Cole and Xiong, 2017, for a review), yet demand at market prices remains low (e.g. Cole et al., 2017). One prominent factor diminishing its value is the presence of basis risk, whereby insurance fails to trigger for non-weather-related loss. Mismatch between indemnities and production lowers insurance value, especially for those near subsistence for whom unrecovered premia impose a substantial burden (Clarke, 2016).

Newer products attempt to mitigate basis risk by indexing indemnities to average yield in an index zone. In principle, area–yield more comprehensively encompasses sources of crop loss—e.g. pests or fire—and therefore better reflects individual production. Field trials have shown promise, but have employed small index zones at or below the level of a single cooperative (e.g. Casaburi and Willis, 2018; Stoeffler et al., 2021). Narrowly tailored indices reintroduce concerns of asymmetric information as clients can coordinate to manipulate the index. For a viable area–yield insurance, index zones must be sufficiently large to withstand strategic manipulation.

In this paper, we present a framework to assess the market potential for area–yield index insurance to improve on weather insurance. Our analysis describes how as an index zone grows, and therefore the scope for market manipulation shrinks, so does the basis risk of an area–yield insurance contract. We quantify this tradeoff using data on maize in Ghana, and then identify the index zone size that makes area–yield insurance competitive with weather insurance in terms of basis risk.

The underlying insight of this paper is that not all insurance indices are equally manipulable. Notably, weather insurance can be tied to arbitrarily small areas without admitting information asymmetry. Our empirical exercise complements the decomposition by Stigler and Lobell (2023) of design and zonal risk within a fixed zone by letting zone size vary with contract type.

2 Theory

Agricultural insurance indemnifies negative productivity shocks. Plot-level productivity can be described with respect to an insurance contract by insured and uninsured components. Formally, let yield Y_{it} on plot i in year t be

$$Y_{it} = \gamma_i + \beta T_{it} + \epsilon_{it} \tag{1}$$

where γ_i is average (anticipated) yield, T_{it} is the insurance index realization, β is a scaling factor relating the index to output, and ϵ_{it} is the idiosyncratic uninsured variation.

A contract's insurance value, and therefore market demand, increases with the correspondence between the index and plot-level yield. The remaining uninsured variation constitutes basis risk, quantified for an index as the ratio of uninsured to total variance averaged across plots:

$$BR = \frac{1}{N} \sum_i \frac{\text{Var}_t\, \epsilon_{it}}{\text{Var}_t\, Y_{it}} \tag{2}$$

We define basis quality as one minus this value.

With weather insurance, T_{it} is some weather realization. This contract inherently avoids information asymmetry because, adjusting for local climate, seasonal weather is a random shock outside farmers' control. In principle, it could be indexed precisely to plot-level conditions with the appropriate measurement technology. However, even this level of geographic specificity leaves substantial uninsured risk in place.

Area–yield insurance offers an attractive alternative to improve basis quality. With area–yield contracts, $T_{it} = \bar{Y}_t$ within the index zone of plot i. At the extreme, perfect insurance sets $T_{it} = Y_{it}$ with zero residual variance in ϵ_{it}, but such a contract is infeasible due to information asymmetry. Expanding the index zone mitigates this concern as individuals have less influence over the index, but does so at the cost of basis risk as the zone average becomes less informative about any individual plot within the zone.

In this study, we quantify how the basis quality of area–yield insurance degrades with index zone size. We then compare area–yield insurance to weather insurance to identify how small a zone (and associated market size) an insurer must demarcate for area–yield crop insurance to improve basis quality. We analyze maize in Ghana, and our methods readily extend to other crops and regions.

3 Data

The ideal data to estimate (1) would be a long plot-level panel. Such granularity is rare over large areas in developing countries.[1] We instead use data from the Ghana Ministry of Food and Agriculture (MOFA), which reports annual output and area harvested from 2006–2011 for each of the country's then 138 districts.

We isolate yearly unanticipated variation in district productivity using the Global Agro-Ecological Zones (GAEZ) database (FAO and IIASA, 2023). The database combines detailed soil, terrain, and climate data to apportion national production across geographic units. We treat this apportionment as a district's typical productivity, and deviations reported by MOFA are defined to be unanticipated shocks. Full details are given in Appendix A.

3.1 Basis Risk

The basis quality of area–yield insurance is the correspondence between plot-level productivity and index zone yield. Absent plot-level yield data, we quantify this relationship at the level of 9km$\times$9km tracts as defined in GAEZ data by modeling tract-specific productivity as a jointly normal process with local covariance. Covariance parameters are calibrated to match district-level spatial variation, and we report the precise relationship between local productivity and insurance zone average implied by the calibrated data generating process. Calibration details are provided in Appendix B.

3.2 Market Size

The scope for market manipulation depends on market size within an insurance zone. We relate market volume to zone area using the GAEZ's local apportionment of national harvest quantity. We consider this apportionment to be each tract's typical output, and calculate a zone's total production in a typical year. For each zone size, we define market volume as the average of this value across all possible zones of that size.

[1] Advances in remote sensing offer future promise, but correspondence with ground truth remains low (e.g. Jin et al., 2017).

4 Results

Model calibration indicates correlation over the range of three GAEZ grid cells. Beyond 27km, common components of maize yield shocks are indistinguishable from background noise. We explore the implications of this spatial correlation for two types of area–yield contracts in Figure 1.

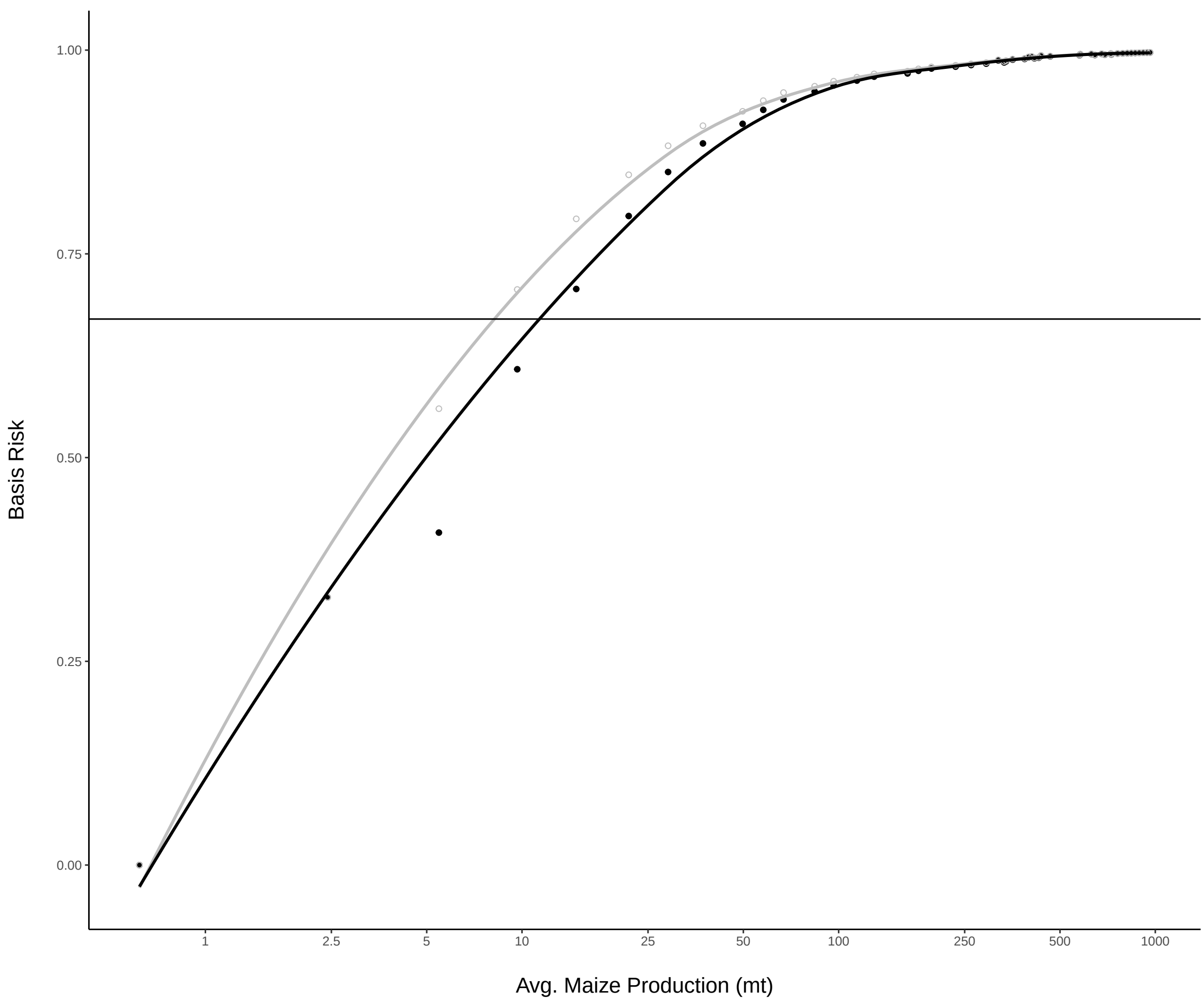
Basis quality is greatest at the center of an index zone and worse toward the edges, where nearby tracts with similar productivity lie outside the measured area. The lighter curve represents average basis risk across all tracts in a fixed zone, reflecting how zones are traditionally demarcated. The darker curve illustrates the potential to improve the contract by designating tract-specific index zones centered around the insured tract. Such precision is becoming increasingly accessible as remote sensing enables measurement at finer spatial resolutions.

Expanding insurance zone area initially increases basis risk by more in the fixed-zone contract than the tract-specific contract because it adds more peripheral tracts where the index performs poorly. The gap is most pronounced in the 0.5–25kt range, and subsequently closes as index zones grow too large to be informative anywhere. By 50kt, corresponding to roughly 80km×80km zones, the signal value of an area–yield index is almost completely degraded.

For comparison, the horizontal line in Figure 1 represents the basis risk in weather insurance. This benchmark is estimated from nation-level maize data in West Africa (Lobell and Burke, 2008) and plot-level maize data from Kenya (Stigler and Lobell, 2023). Both studies report a consistent correlation of roughly 0.33 between rainfall and maize production, corresponding to basis risk of 0.67. Basis quality does not vary with production volume because there is already minimal concern about asymmetric information in tract-level weather.

Insurers must be willing to create area–yield index zones producing 8kt or less—representing 34km×34km or smaller areas—for area–yield insurance to match the basis risk of weather insurance. This production volume corresponds to roughly 5,000 maize–producing households per zone (from Ghana Statistical Services, 2020). Allowing tract-specific zones relaxes this constraint to 11.3kt—40km×40km zones or 7,000 households. Collusion to manipulate an insurance index would be difficult to sustain at this scale, so we conclude there is sufficient scope for area–yield insurance to improve the basis quality of agricultural

Figure 1: Basis Risk versus Market Size in Area–Yield Insurance



Notes: Vertical axis measures basis risk defined by (2); horizontal axis denotes production volume in kilotonnes (kt). Grey circles and fitted curve represent average basis quality across all tracts in insurance zone. Black dots and fitted curve represent basis quality in central tract. Horizontal line shows basis risk of weather insurance.

insurance without unraveling due to asymmetric information.

5 Discussion

The analysis in this paper focuses on the tension between improving basis quality through geographic compactness and limiting asymmetric information by expanding market size when selecting an index zone. We present a general framework to evaluate this tradeoff and precisely characterize how basis risk grows with market size for Ghanaian maize. Complementary analysis quantifying how market size mitigates information asymmetry remains an open area for future research.

Area–yield insurance quality depends crucially on correlation across space. Lower spatial variance lowers basis risk within an index zone and allows for larger zones to achieve the same level of risk. We calibrate a tractable spatial model using district aggregate production. A drawback of this approach is that it limits analysis to simplified basis quality measures. Utility-based assessment (e.g. Conradt et al., 2015) is sensitive to distributional assumptions about the tails of data generating process, but would be possible with finer geographic resolution.

Results indicate insurers must demarcate zones of 8kt or less for area–yield insurance to be more attractive than weather insurance to maize farmers in Ghana. Given the available data, this comparison considers insurance at the level of 9km×9km tracts. Stigler and Lobell (2023) estimate residual plot-level yield variation at this resolution to be around 0.5, so smaller area–yield index zones would be needed to compete with more finely targeted weather insurance. On the other hand, many existing weather-based alternatives likely perform far worse because they aggregate much larger regional weather patterns (e.g. Awondo, 2019).

Alternate approaches to address basis risk focus on expanding the scope of named hazards to include other aggregate shocks such as pests or fire. If the correlation between these hazards and yield can be quantified, our framework readily accommodates comparisons with multiple–hazard insurance or other contract structures. We encourage this type of analysis to evaluate market potential when introducing new forms of agricultural insurance.

References

Awondo, Sebastain, "Efficiency of region-wide catastrophic weather risk pools: Implications for African Risk Capacity insurance program," *Journal of Development Economics*, 2019, *136*, 111–118.

Casaburi, Lorenzo and Jack Willis, "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya," *American Economic Review*, 2018, *108* (12), 3778–3813.

Clarke, Daniel, "A Theory of Rational Demand for Index Insurance," *American Economic Journal: Microeconomics*, 2016, *8* (1), 283–306.

Cole, Shawn and Wentao Xiong, "Agricultural Insurance and Economic Development," *Annual Review of Economics*, 2017, *9* (1), 235–262.

_, Xavier Giné, and James Vickery, "How Does Risk Management Influence Production Decisions? Evidence from a Field Experiment," *Review of Financial Studies*, 2017, *30* (6), 1935–1970.

Conradt, Sarah, Robert Finger, and Raushan Bokusheva, "Tailored to the extremes: Quantile regression for index-based insurance contract design," *Agricultural Economics*, 2015, *46* (4), 537–547.

FAO and IIASA, "Global Agro-Ecological Zones v4," 2023. Accessed: 12/01/2023.

Ghana Statistical Services, "2017/18 National Census of Agriculture," National Report, Republic of Ghana 2020.

Gunnsteinsson, Snaebjorn, "Experimental identification of asymmetric information: Evidence on crop insurance in the Philippines," *Journal of Development Economics*, 2020, *144*, 102414.

Jin, Zhenong, George Azzari, Marshall Burke, Stephen Aston, and David Lobell, "Mapping Smallholder Yield Heterogeneity at Multiple Scales in Eastern Africa," *Remote Sensing*, 2017, *9* (9), 931.

Lobell, David and Marshall Burke, "Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation," *Environmental Research Letters*, 2008, *3* (3), 034007.

Stigler, Matthieu and David Lobell, "Optimal index insurance and basis risk decomposition: an application to Kenya," *American Journal of Agricultural Economics*, 2023, *forthcoming*, 1–24.

Stoeffler, Quentin, Michael Carter, Catherine Guirkinger, and Wouter Gelade, "The Spillover Impact of Index Insurance on Agricultural Investment by Cotton Farmers in Burkina Faso," *World Bank Economic Review*, 2021, *36* (1), 114–140.

Supplementary Appendix for

"The Market Potential for Area–Yield Crop Insurance"

For Online Publication Only

A District-Level Yield Shock Calculation Details

We define district-level yield shocks in a given year to be the deviation of actual yield reported by the Ghana Ministry of Food and Agriculture (MOFA) from anticipated yield implied by the Global Agro-Ecological Zones (GAEZ) database. To compute the latter, we rescale the 2010 GAEZ apportionment of national production by aggregate production in a given year and adjust for changes in area harvested. With a longer panel, anticipated productivity could also be measured as the within-district mean over time.

For the calculation used in this paper, let $\tilde{Q}_i$ and $\tilde{A}_i$ represent tract-level output and area, respectively, reported in the GAEZ database. These values are imputed in the data for each tract, defined as a 9km×9km grid cell, around the year 2010 by taking averages of national output and area over the period 2009–2011. National values are apportioned among tracts according to local soil, terrain, and climate conditions. Importantly, this apportionment uses fixed tract characteristics without regard to time-varying features such as rainfall or pest damage in the imputation period. Therefore, we interpret the GAEZ area and output projections to reflect anticipated productivity (γ_i) independent of year-specific shocks.

To convert tract-specific anticipated 2010 productivity from GAEZ into district-level anticipated productivity in year t for comparison to MOFA data, we proceed in four steps. First, we aggregate across tracts in a district to compute the district-level anticipated yield in 2010.

$$\tilde{\gamma}_d = \frac{\tilde{Q}_d}{\tilde{A}_d} \equiv \frac{\sum_{i \in d} \tilde{Q}_i}{\sum_{i \in d} \tilde{A}_i} \tag{A.1}$$

Second, we compute the change in district-level output we would expect in year t if only area harvested deviated from the GAEZ estimate, with no difference in district-level productivity.

$$\tilde{Q}_{dt} = \tilde{\gamma}_d A_{dt} \tag{A.2}$$

Third, we calculate the ratio of observed national production in year t to what would be predicted from changes in area alone.

$$R_t = \frac{\sum_d Q_{dt}}{\sum_d \tilde{Q}_{dt}} \tag{A.3}$$

Consider this ratio to be a rescaling factor reflecting nation-wide technology or agricultural intensity. Finally, the year-specific anticipated productivity in a district is calculated as the GAEZ-defined anticipated productivity multiplied by that year's national rescaling factor.

$$\gamma_{dt} = \tilde{\gamma}_d R_t \tag{A.4}$$

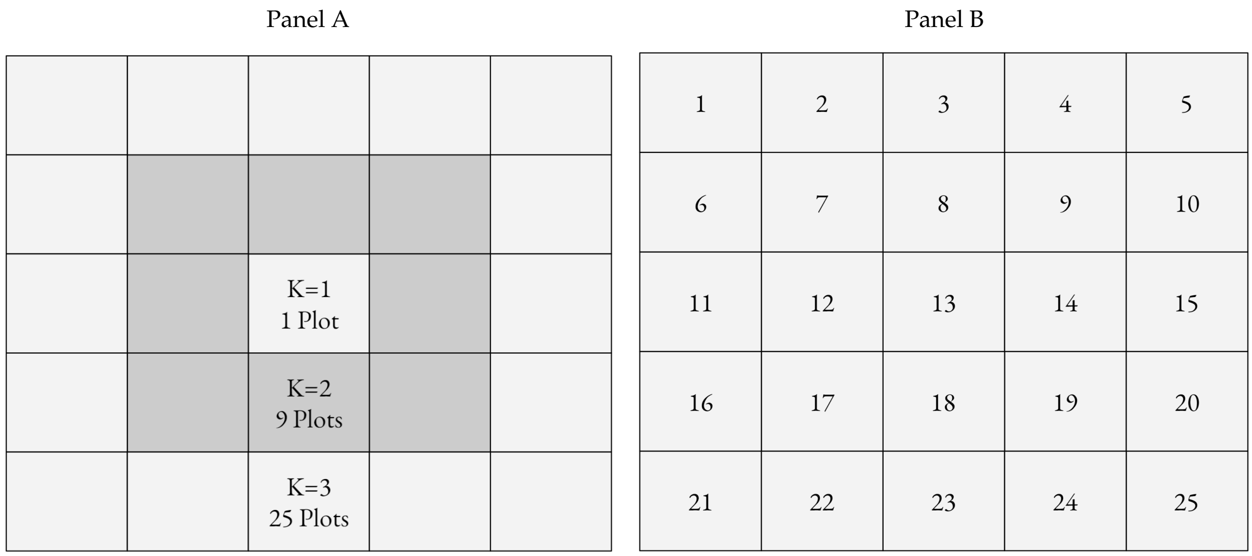
Differences between this anticipated production derived from fixed geographic characteristics and actual production reported by MOFA constitute the insurable yield shocks analyzed in this study.

Note that this calculation treats nation-wide productivity fluctuations as uninsurable variation embedded into γ_i. We believe this treatment to be sensible for two reasons. First, national fluctuations are likely caused by predictable factors such as regional climate patterns, technological developments, or macroeconomic conditions that influence access to farm inputs. It is less likely that movement in aggregate output comes from idiosyncratic shocks to tracts that are incidentally similar across the entire nation. Second, it would require substantial capital reserves for a domestic insurer to indemnify a simultaneous negative shock to the entire country. It is far more credible for an insurance company to diversify geographically within the nation and protect against locally idiosyncratic risk.

B Area–Yield Index Basis Risk Calculation Details

The relationship between individual tract productivity and average yield in an insurance zone depends crucially on the spatial correlation of productivity shocks. To quantify spatial correlation, we model the data-generating process for tract-level productivity as a joint normal distribution with correlation across nearby tracts. We then calibrate parameters to match the observed spatial variation in yield shocks across districts in MOFA data using maximum likelihood. Finally, we use the calibrated model to calculate the basis quality of insurance zones of arbitrary size.

Figure B.1: Aggregation of Characteristic Shocks into Tract Productivity



Notes: Panel A depicts balls of size 1, 2, and 3 around the central tract. Panel B numbers tracts in the grid for reference in equations (B.3)–(B.6).

B.1 Data Generating Process

We model tract-level productivity as a jointly normal process with correlation in nearby tracts that decays with distance. To operationalize this, let each tract receive a characteristic shock

$$\omega_{it} \sim (0, \sigma^2) \tag{B.1}$$

drawn i.i.d across tracts and years. Tract-level yield is a weighted combination of a tract's own characteristic and that of its neighbors. Formally, let

$$\begin{aligned} Y_{it} &= \gamma_i + \mu_{it} \\ \mu_{it} &= \frac{1}{2K-1} \sum_{j \in S_K(i)} \omega_{jt} \end{aligned} \tag{B.2}$$

where $S_K(i)$ represents all tracts in a K-sized ball around tract i. That is, $S_1(i)$ contain the tract i itself. $S_2(i)$ is tract i and the eight tracts directly adjacent to it, including those that share a corner. $S_3(i)$ adds the 16 tracts that directly encircle $S_2(i)$, and so on. Panel A of Figure B.1 balls of size 1, 2, and 3 around the central tract.

With this construction, tract-level productivity shocks μ_{it} have the same variance as the characteristic shocks ω_{it} because there are $(2K-1)^2$ tracts in a K-sized ball. However, there is spatial correlation in μ_{it} between tracts to the extent that they consist of overlapping characteristics. As an illustrative example, consider the area depicted by Panel B of Figure B.1. When $K = 2$, the productivity shocks on select tracts can be written (suppressing time subscripts for simplicity) as

$$\mu_7 = \frac{1}{3}(\omega_1 + \omega_2 + \omega_3 + \omega_6 + \omega_7 + \omega_8 + \omega_{11} + \omega_{12} + \omega_{13}) \tag{B.3}$$

$$\mu_8 = \frac{1}{3}(\omega_2 + \omega_3 + \omega_4 + \omega_7 + \omega_8 + \omega_9 + \omega_{12} + \omega_{13} + \omega_{14}) \tag{B.4}$$

$$\mu_9 = \frac{1}{3}(\omega_3 + \omega_4 + \omega_5 + \omega_8 + \omega_9 + \omega_{10} + \omega_{13} + \omega_{14} + \omega_{15}) \tag{B.5}$$

$$\mu_{19} = \frac{1}{3}(\omega_{13} + \omega_{14} + \omega_{15} + \omega_{18} + \omega_{19} + \omega_{20} + \omega_{23} + \omega_{24} + \omega_{25}) \tag{B.6}$$

The variance of each of these terms is σ^2. The covariance in productivity on adjacent tracts 7 and 8 is determined by the shared terms in (B.3) and (B.4)

$$\text{cov}(\mu_7, \mu_8) = \frac{1}{9}(\text{var}(\omega_2) + \text{var}(\omega_3) + \text{var}(\omega_7) + \text{var}(\omega_8) + \text{var}(\omega_{12}) + \text{var}(\omega_{13})) = \frac{2}{3}\sigma^2$$

For non-adjacent tracts 7 and 9, the covariance in productivity is determined by only three overlapping terms

$$\text{cov}(\mu_7, \mu_9) = \frac{1}{9}(\text{var}(\omega_3) + \text{var}(\omega_8) + \text{var}(\omega_{13})) = \frac{1}{3}\sigma^2$$

and even more distant tracts 7 and 19 share a single overlapping term so $\text{cov}(\mu_7, \mu_{19}) = \frac{1}{9}\text{var}(\omega_{13}) = \frac{1}{9}\sigma^2$.

The extent of spatial correlation is captured by K—expanding the ball increases the overlap between adjacent tracts and introduces correlation between more distant tracts. Note that characteristic shocks ω_{it} have no physical interpretation. They do not, for example, represent spillovers from nearby rainfall or pests. The use of ω_{it} is merely a modeling technique to describe correlation in productivity shocks μ_{it} that decays with distance in a parsimonious way for calibration.

B.2 Calibration with Maximum Likelihood

The data-generating process can be summarized by the two parameters (σ, K) that describe the variance and spatial correlation, respectively, of productivity shocks across tracts. We next calibrate these parameters to match the observed distribution of district-level yield shocks inferred from MOFA production data.

To map the model to data, define district-level yield to be a weighted average of yield across all GAEZ tracts in the district, weighted by harvested area in the tract. The productivity shock in the district can then be written as a weighted average of productivity shocks across tracts (μ_{it}) in the district, which can in turn be written as a weighted average of

Table B.1: Parameter Estimates and Log Likelihoods

K	σ	log(Likelihood)
1	1.760	-606.7
2	0.821	-508.9
3	0.837	-580.5
4	1.093	-725.0
5	1.354	-807.0
6	1.632	-874.1
7	1.802	-879.9
8	1.932	-868.6
9	2.152	-898.7
10	2.312	-905.1

characteristic shocks (ω_{it}) on tracts in and adjacent to the district. That is,

$$\mu_{dt} = \frac{1}{A_d} \sum_{i \in d} A_i \mu_{it} = \sum_{i \in S_{1--K}(d)} C_i \omega_{it} \tag{B.7}$$

for some weights C_i defined by harvested area A_i and (B.2).

Each μ_{dt} is the sum of independent, normally distributed variables ω_{it}. Therefore, the vector of district-level yield shocks $\vec{\mu}_t = \{\mu_{1,t}, \ldots, \mu_{138,t}\}$ in a given year can be written as a multivariate normal random variable with a covariance matrix defined as a function of parameters (σ, K) by the overlapping ω_{it} components in districts' yield processes.

To calibrate the model, we search over the parameter space for values that maximize the joint likelihood of producing the six realizations of $\vec{\mu}_t$ observed in the 2006–2011 production data reported by MOFA. Optimization is implemented using maximum likelihood by fixing K, calculating the value of $\sigma|K$ that maximizes the likelihood of the observed yield shocks for a given K, and then searching over the range $K \in \{1, \ldots, 50\}$, spanning the breadth of the country. Likelihoods are presented for $K \in \{1, \ldots 10\}$ in Table B.1. We also allow the weight assigned to ω to decay with distance rather than be constant within the K-sized ball, but find the maximum likelihood falls with even very slight decay.

B.3 Computation of Basis Risk

Finally, we use the calibrated data generating process to compute the covariance between the shock to average yield in an insurance zone and tract-specific shocks within the zone. Each of these values can again be expressed as sums of characteristic shocks ω_{it}, and therefore follow a joint normal distribution with covariance determined by the degree of overlap between an individual tract's productivity components and those of the full insurance zone.

We report two measures of the basis quality for a zone of arbitrary size $N \times N$. First, we report the average basis risk across all tracts in the zone. This value will be smaller for tracts toward the center of the zone, whose productivity components overlap with more of the zone, and greater for tracts toward the edge. Basis quality computed in this manner corresponds to crop insurance with fixed, predefined insurance zones, following how area–yield is commonly implemented.

Second, we report the basis risk on only the most central tract(s), for which the area–yield index will be most informative. This measure represents an upper bound to what is achievable with area–yield insurance. Tract-specific index computation is becoming increasingly feasible with remote sensing technology that removes cost barriers to making multiple yield measurements.