The Economic 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

Rainfall index insurance can enable farm households to manage production risk, but demand in developing countries remains low at market prices, in part because the insurance trigger may not correlate well with individual farm losses. Area–yield crop insurance, which links payouts to average yield in a geographic zone, attempts to increase demand by more accurately targeting insurance payouts to production shortfalls. However, shifting from an exogenous weather-based to an endogenous yield-based index introduces concerns of asymmetric information, which can lead to market failures that constrain supply from providers. These features are inversely related: larger insurance zones inhibit index manipulation, but average yield is less informative about any individual plot. We quantify this tradeoff for maize in Ghana using a spatial yield model calibrated to match observed production. Insurers must demarcate zones of no more than 5,000 farmers for area–yield insurance to outperform weather insurance. The framework presented in this paper allows assessment of the relationship between index performance and asymmetric information in 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 Colin Carter, Richard Gallenstein, Jens Hilscher, Andrew Hobbs, Rich Sexton, 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

Risk remains a salient barrier to agricultural investment and rural development. Crop insurance can insulate farm households from production risk, but directly insuring individual on-farm yield may be hampered by asymmetric information (Gunnsteinsson, 2020). To limit market failures due to adverse selection and moral hazard, insurers in developing economies often base payouts on readily observable exogenous factors such as low rainfall.

Weather-based index insurance has proven to promote investment and prevent decapitalization in subsidized field trials (see Cole and Xiong, 2017, for a review), yet demand at market prices remains low in developing countries (e.g. Cole et al., 2017). One prominent factor diminishing its appeal is the presence of basis risk, which arises in cases of mismatch between insurance payouts and individual farm losses. Downside basis risk, when insurance fails to trigger despite revenue shortfalls, is especially costly to those near subsistence for whom unrecovered premia constitute a substantial burden in times of loss (Clarke, 2016).

Area–yield index insurance, in which payouts are based on average yield in a geographic zone, can raise demand by more comprehensively encompassing on-farm crop loss. Field trials show promise on very small zones (e.g. Casaburi and Willis, 2018; Stoeffler et al., 2021), but linking payments to an endogenously determined outcome reintroduces asymmetric information that can constrain suppliers' willingness to issue such policies. To sustain an area–yield insurance market, insured zones must be sufficiently large that providers are protected from coordinated responses by policyholders within the zone.

In this paper, we assess whether area–yield index insurance can lower basis risk for policyholders while still mitigating asymmetric information for providers. Our analysis complements work by Stigler and Lobell (2024) and Gallenstein and Dougherty (2024), and by Tsiboe et al. (2023) in the U.S. context, that quantifies the basis risk and associated insurance value to policyholders of switching from exogenous weather-based to endogenous yield- and price-based indices within a fixed insurance pool. We introduce a framework to weigh such improvements in index performance, which can raise demand from policyholders, against supply-side concerns of asymmetric information and index manipulation in small insurance zones that may hinder the financial viability of products offered by insurance providers.

The underlying insight is that as an index zone grows, and therefore the capacity to sustain internal coordination shrinks, basis risk increases. We identify the largest possible area–yield index zone that improves basis risk over rainfall insurance by calibrating a spatial model using data on maize in Ghana. Results indicate area–yield index insurance can only outperform rainfall index insurance, leading to greater consumer demand, if insurers operate zones of no more than 8 kilotonnes (kt), encompassing roughly 5,000 farmers on average. We encourage this style of economic analysis when designing crop insurance contracts.

2 Theory

Plot-level productivity can be described in relation 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 index realization that determines payouts, β scales the index to output, and ϵ_{it} is uninsured productivity variation.

We define the performance of an insurance index as the correlation between realization of the index variable T_{it} and on-farm yield Y_{it}. Insurance value to policyholders, and therefore market demand, increases with index performance. The remaining uninsured variation constitutes basis risk, formally quantified as the ratio of uninsured to total production variance:

$$BR = \frac{\mathrm{Var}_t\, \epsilon_{it}}{\mathrm{Var}_t\, Y_{it}} \tag{2}$$

The maximally performing index perfectly reflects on-farm yield as $\beta T_{it} = Y_{it} - \gamma_i$ with zero residual variance in ϵ_{it}. This would enable insurance to directly cover individual on-farm losses, but may generate market failures if asymmetric information allows farmers to adjust their yield or coverage in response to the contract. Such strategic behavior is mitigated in the U.S. market, where individual indemnity insurance is prevalent, through precise input monitoring and history-dependence in long-term contracts (Mieno et al., 2018, see). However, in developing economies with lower institutional capacity and weaker contracting environments, insurance payouts are more commonly indexed to external outcomes.

Traditional index insurance contractually defines T_{it} using exogenous productivity-related factors such as rainfall or temperature. Such contracts avoid information asymmetry because, conditional on climate, weather is a publicly observable random shock outside farmers' control. In principle, insurance could be indexed to precise plot-level weather conditions with the appropriate measurement technology. However, even this level of granularity leaves substantial uninsured risk from non-weather-related loss.

Area–yield insurance, which defines T_{it} as average productivity within a geographic zone, offers an attractive alternative to improve index performance by better reflecting plot-level outcomes. Expanding the index zone to include many plots mitigates concerns of information asymmetry as individuals have less influence over the index outcome, but does so at the cost of increasing basis risk as the zone average becomes less informative about each plot in the zone.

In this study, we quantify how the index performance of area–yield insurance degrades with index zone size. We then identify how small a zone an insurer must demarcate to improve over weather insurance. We analyze maize in Ghana, and our methods readily extend to other crops, regions, and types of indices.

3 Data

The ideal data to estimate both average production as well as year-to-year variability within an insurance zone would be a plot-level panel. Unfortunately, long time series and high granularity are rare at large scale in developing countries.[1] We instead benchmark local production to the Global Agro-Ecological Zones (GAEZ) database, which combines time-invariant soil, terrain, and climate conditions to apportion national production across geographic units (FAO and IIASA, 2023). This cross-sectional apportionment is treated as anticipated productivity, defined by γ_i in (1). We then use data on annual output and area harvested reported by the Ghana Ministry of Food and Agriculture (MOFA) from 2006–2011 for the country's then 138 districts to compute yearly deviations. Full details are given in Appendix A.

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

3.1 Basis Risk

The index performance of area–yield insurance is the correspondence between plot and index zone productivity shocks. To quantify this correspondence, we decompose district-level production variability into shocks at the 9km×9km tract level delineated in GAEZ data by modeling the data-generating process for tract-level yield as a spatially autocorrelated random draw. Covariance parameters in the model are calibrated to match measured district-level deviations reported by MOFA from the GAEZ benchmark, with details provided in Appendix B. We calculate area–yield index performance and basis risk using the correlation between tract-level productivity shocks and index zone aggregate shocks implied by the calibrated model.

3.2 Market Size

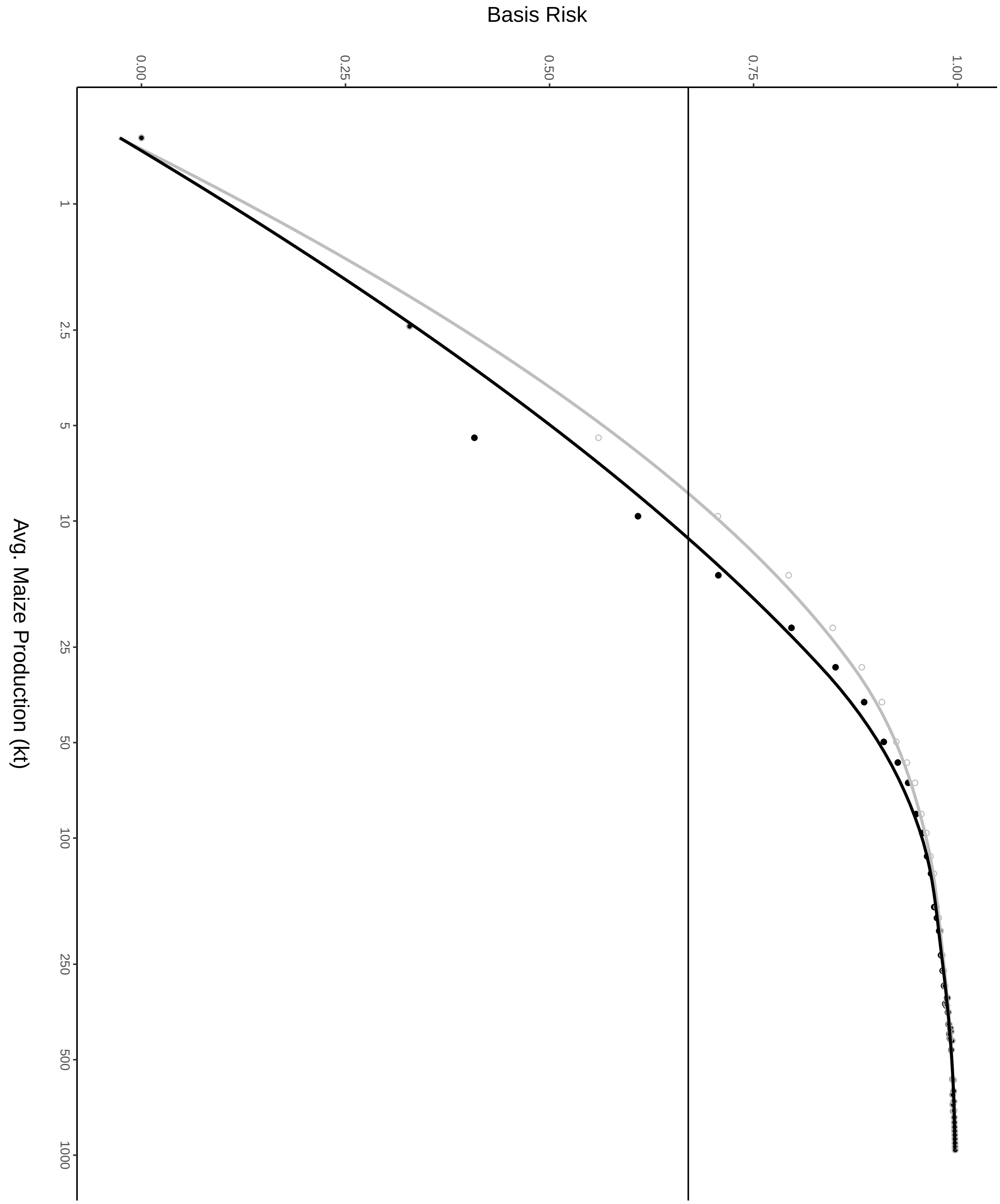
The scope for index manipulation and other forms of moral hazard depend on the size of the pool used to compute the index. We calculate production volume in an area–yield index zone using the GAEZ's measure of anticipated yield. For each index zone size, we define volume as the average of this value across all possible zones of that size.

4 Results

Model calibration indicates spatial correlation in productivity over the range of three GAEZ tracts. Beyond 27km, common components of maize yield shocks are indistinguishable from background noise. We report the implications for area–yield insurance in Figure 1.

The lighter curve represents basis risk averaged across all tracts in a fixed zone, reflecting how zones are traditionally demarcated. Basis risk is lowest at the center and increases toward the edges, where adjacent tracts with similar productivity shocks may fall outside the index zone. The darker curve illustrates the potential to improve index performance by designating tract-specific index zones centered around each insured tract. Such precision is becoming increasingly accessible as remote sensing enables measurement at finer spatial resolutions.

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



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

Over small areas, basis risk grows faster with size in the fixed-zone 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 narrows as zones grow too large to be informative. By 50kt, corresponding to 80km×80km zones, the signal value of area–yield is almost completely degraded.

For comparison, the horizontal line represents basis risk in weather index insurance. This benchmark is calibrated from analyses of national maize production in West Africa (Lobell and Burke, 2008) and plot-level maize production in Kenya (Stigler and Lobell, 2024). Both studies report correlation between rainfall and output of around 0.33, indicating basis risk of 0.67. Index performance does not vary with volume because the exogenous index already minimizes asymmetric information, so there is no benefit to coarsening the resolution of the index as the market grows.

For area–yield insurance to improve on the basis risk of weather insurance, index zones must be at most 34km×34km—representing production of 8kt or less. This volume corresponds to roughly 5,000 maize–producing households per insurance zone (from Ghana Statistical Services, 2020). Allowing tract-specific area–yield indices relaxes this constraint to 40km×40km zones—containing 7,000 farm households producing 11.3kt of maize. Internal collusion to manipulate an index would be difficult to sustain at these production scales, so we speculate there is scope for area–yield to improve performance in index insurance in this setting without triggering market failure.

5 Discussion

We introduce a general framework to characterize the tension in area–yield index insurance between basis risk, which can lower demand, and asymmetric information, which can constrain supply. This tension, inherent to any agricultural index insurance that aggregates endogenous outcomes, arises because index performance is improved through compactness in insurance pools, while mitigating asymmetric information requires expanding the size of the pool used to compute the index. We calibrate a production model to quantify the tradeoff between these two features for Ghanaian maize.

For area–yield insurance, this tradeoff depends crucially on spatial correlation in productivity because lower variance allows for larger insurance zones to achieve the same level of basis risk. We implement a tractable calibration approach to infer higher-resolution spatial variation from district-level production data. Our methodology limits analysis to correlational measures of index performance. More complex utility-based assessment (e.g. Conradt et al., 2015) is sensitive to distributional assumptions about the tails of data generating process, and would be possible to incorporate into our framework with richer spacio-temporal yield data.

The calibrated model indicates maize insurers in Ghana must designate insurance zones encompassing no more than 5,000–7,000 farmers for an area–yield index to outperform weather index insurance. This scale is likely too large to sustain internal coordination to manipulate yield realizations, but the capacity for collusion and extent of moral hazard are ultimately context-specific empirical questions beyond the scope of this paper to quantify. In the end, determining an acceptable level of exposure to information asymmetry remains at the discretion of insurance providers.

This paper analyzes area-yield insurance on 9km×9km tracts, matching the spatial resolution of rainfall insurance offered by the Ghana Agricultural Insurance Pool. Both weather and crop productivity are observable through remote sensing at this level of resolution, but Stigler and Lobell (2024) estimate residual plot-level yield variation of around 0.5 within tracts of this size. Therefore, smaller index zones would be needed to compete with more finely targeted plot-level weather insurance.[2] On the other hand, many existing weather-based products aggregate larger regional patterns (e.g. Awondo, 2019) and likely perform far worse, leaving substantial scope for improving index performance.

Alternate approaches to address basis risk with exogenous indices expand the scope of named hazards to include, e.g., pests or fire. Our framework readily accommodates comparisons with such contract structures. We encourage this type of analysis to evaluate potential supply-side market constraints when introducing new forms of agricultural insurance based on endogenous outcomes.

[2]It is worth noting that differences in the cost of measuring plot-level weather and yield outcomes add an additional dimension of complication to comparisons of market viability between different types of high-resolution insurance products.

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.

Gallenstein, Richard and John Dougherty, "Can revenue index insurance outperform yield index insurance?," *American Journal of Agricultural Economics*, 2024, *forthcoming*.

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.

Mieno, Taro, Cory G. Walters, and Lilyan E. Fulginiti, "Input Use under Crop Insurance: The Role of Actual Production History," *American Journal of Agricultural Economics*, 2018, *100* (5), 1469–1485.

Stigler, Matthieu and David Lobell, "Optimal index insurance and basis risk decomposition: an application to Kenya," *American Journal of Agricultural Economics*, 2024, *106*, 306–29.

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.

Tsiboe, Francis, Jesse Tack, and Jisang Yu, "Farm-level evaluation of area- and agroclimatic-based index insurance," *Journal of the Agricultural and Applied Economics Association*, 2023, *2* (4), 616–633.

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 would be measured as the within-district mean over time. However, because MOFA data are only consistently reported for six years, district historical averages would be computed with considerable error over this short span. Moreover, technological improvements and market development led yields to increase nationwide during this period, so inferring average productivity adds the additional complication of separating year-to-year fluctuations from the long-term trend in this already short time series. By contrast, the GAEZ apportionment utilizes detailed microdata on underlying (fixed) geographic characteristics, and therefore avoids these concerns.

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 Tract-Level 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.

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

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

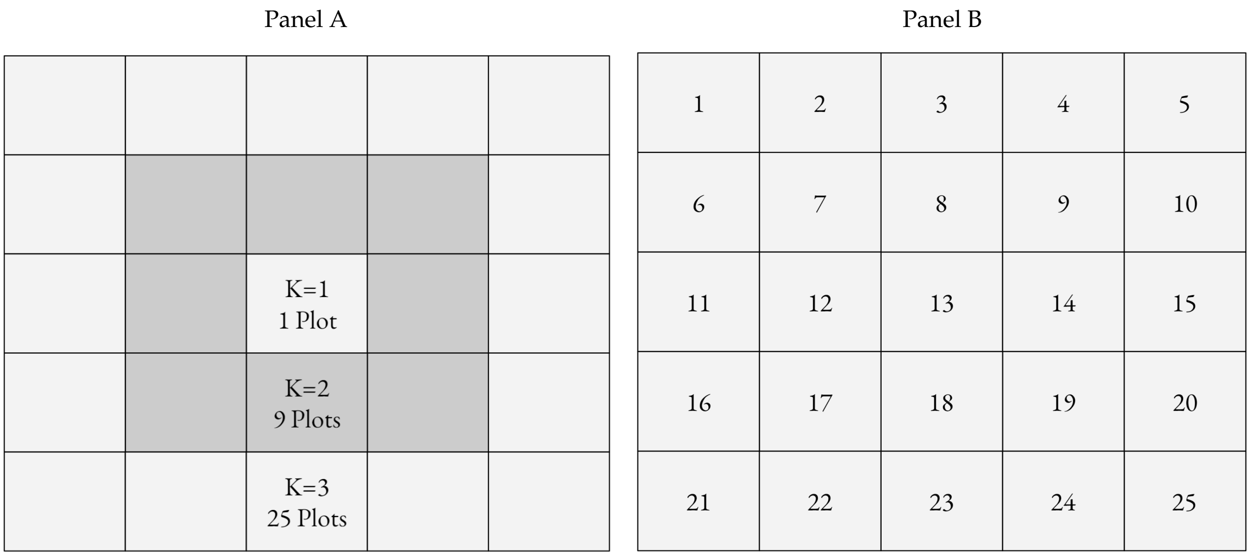
drawn i.i.d across tracts and years. Tract-level yield is a weighted combination of a tract's own characteristic and that of its neighbors. Formally, let

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

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

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).

simplicity) as

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

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

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

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

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)

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

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

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

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 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 $\vec{\boldsymbol{\mu}}_t = \{\mu_{1,t}, \ldots, \mu_{138,t}\}$ 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 $\vec{\boldsymbol{\mu}}_t$ 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

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}, 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.