# 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**

Rainfall insurance can enable farm households to manage production risk, but demand remains low at market prices. Area–yield crop insurance, which links payouts to average yield in a geographic zone, attempts to increase demand by more accurately targeting production shortfalls. However, shifting from an exogenous weather-based to an endogenous yield-based insurance index introduces concerns of asymmetric information, which can constrain supply from providers. These features are inversely related: larger insurance zones prohibit 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 be willing to 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 market viability 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 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

Production risk remains a salient barrier to agricultural investment and rural development. Crop insurance can insulate farm households from risk, but directly insuring individual on-farm yield generates asymmetric information (Gunnsteinsson, 2020). To prevent market unraveling, insurers often base payouts on exogenous factors such as low rainfall.

Weather-based 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 (e.g. Cole et al., 2017). One prominent factor diminishing its appeal is the presence of basis risk, whereby insurance fails to trigger for non-weather-related loss. Mismatch between payouts and production shortfalls is especially costly to those near subsistence for whom unrecovered premia constitute a substantial burden in times of loss (Clarke, 2016).

Area–yield insurance, based on average yield across plots in a geographic zone, can raise demand by more comprehensively encompassing 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 endogenous outcome reintroduces asymmetric information that can constrain suppliers' willingness to issue policies. To sustain an area–yield insurance market, insured zones must be sufficiently large that providers are protected from strategic coordination by policyholders within the zone.

In this paper, we assess whether area–yield 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) that quantifies the 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 demand-side quality improvements to insurance design against supply-side concerns of asymmetric information and index manipulation in small insurance zones.

The underlying insight is that as an index zone grows, and therefore the scope for manipulation 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 be competitive if insurers are willing to operate zones of no more than 8kt, encompassing roughly 5,000 farmers on average. We encourage this style of viability analysis when designing crop insurance contracts.

## 2 Theory

Plot-level productivity can be described relative 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 $\gamma_i$ is average (anticipated) yield, $T_{it}$ is the index realization that determines payouts, $\beta$ scales the index to output, and $\epsilon_{it}$ is uninsured variation.

Insurance value to policyholders, and therefore market demand, increases with the correspondence between the index and realized yield. The remaining uninsured variation constitutes basis risk, quantified as the ratio of uninsured to total production variance:

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

For exposition, let basis quality be one minus this value.

Traditional index insurance defines $T_{it}$ using exogenous outcomes such as rainfall. Such contracts avoid information asymmetry because, conditional on climate, weather is a random shock outside farmers' control. In principle, insurance could be indexed to precise plot-level conditions with the appropriate measurement technology. However, even this level of specificity 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 lower basis risk by better reflecting plot-level outcomes. At the extreme, perfect insurance sets $T_{it} = Y_{it}$ with zero residual variance in $\epsilon_{it}$, but is infeasible due to information asymmetry.

Expanding the index zone to include multiple plots 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 each plot.

In this study, we quantify how the basis quality 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 indices.

## 3 Data

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

We isolate unanticipated productivity variation using the Global Agro-Ecological Zones (GAEZ) database (FAO and IIASA, 2023). The database combines time-invariant soil, terrain, and climate conditions to apportion national production across geographic units. We treat this apportionment as anticipated productivity, and yearly deviations reported by MOFA as 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 and index zone productivity shocks. We quantify this relationship for 9km×9km tracts as delineated in GAEZ data by modeling tract-level yield as a spatially autoregressive process. Covariance parameters are calibrated to match measured variation in MOFA data, and we report the precise relationship between local productivity and zone average implied by the calibrated model. 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 production volume to zone area using the GAEZ's local apportionment of national harvest. For each index zone size,

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

we define volume as the average of this value across all possible zones of that size.

# 4 Results

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 for area–yield insurance in Figure 1.

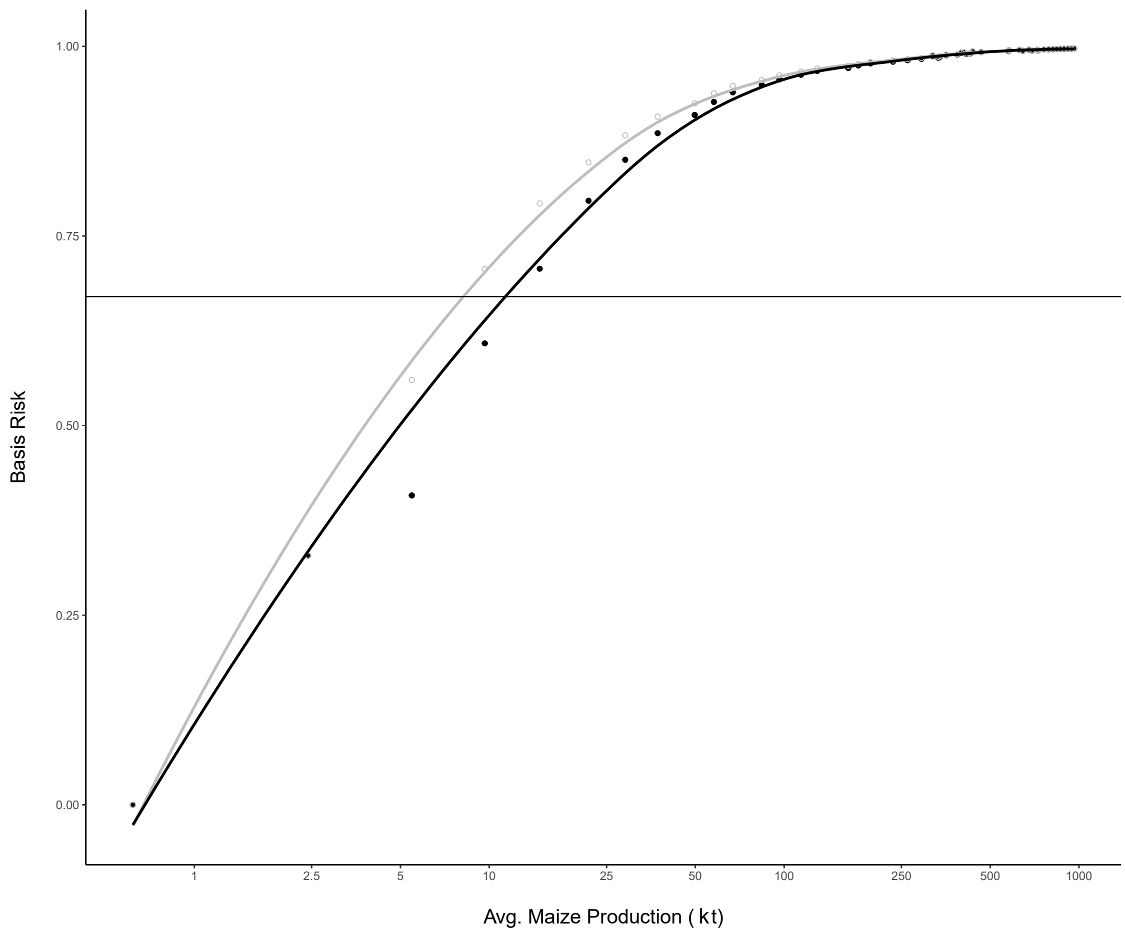
The lighter curve represents basis risk across all tracts in a fixed zone, reflecting how zones are traditionally demarcated. Basis risk is lowest at the center and increases toward the edges. The darker curve illustrates the potential to improve the contract by designating tract-specific zones centered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

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 weather 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 around 0.33, indicating basis risk of 0.67. Basis quality does not vary with volume because the exogenous index minimizes asymmetric information at any scale.

Insurers must be willing to create 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 volume corresponds to roughly 5,000 maize–producing households per insurance zone (from Ghana Statistical Services, 2020). Allowing tract-specific indices relaxes this constraint to 11.3kt—40km×40km zones containing 7,000 farm households. Collusion to manipulate an index would be difficult to sustain at these scales, so we conclude there is scope for area–yield insurance to improve basis quality without unraveling.

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 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.

# 5 Discussion

We introduce a general framework to characterize supply-side constraints in agricultural insurance that aggregates endogenous outcomes such as area–yield, and precisely calibrate it for Ghanaian maize. Our analysis addresses the tension between improving basis quality through geographic compactness and limiting asymmetric information by expanding market size when selecting an insurance zone. Determining an acceptable level of information asymmetry remains at the discretion of insurance providers.

Area–yield insurance viability depends crucially on spatial correlation because lower variance allows for larger zones to achieve the same level of risk. We implement a tractable calibration approach using district-level production, which limits analysis to correlational basis quality measures. 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 with richer spacio-temporal yield data.

This paper analyzes area-yield insurance on 9km×9km tracts, matching the spatial resolution of rainfall insurance offered by the Ghana Agricultural Insurance Pool. Stigler and Lobell (2024) estimate residual plot-level yield variation at this scale of around 0.5, indicating smaller index zones would be needed to compete with more finely targeted weather insurance. On the other hand, existing weather-based products that aggregate larger regional patterns likely perform far worse (e.g. Awondo, 2019).

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 supply-side market potential when introducing new forms of agricultural insurance based on endogenous outcomes.

# 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.

**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.

# 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 ($\gamma_i$) 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.

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

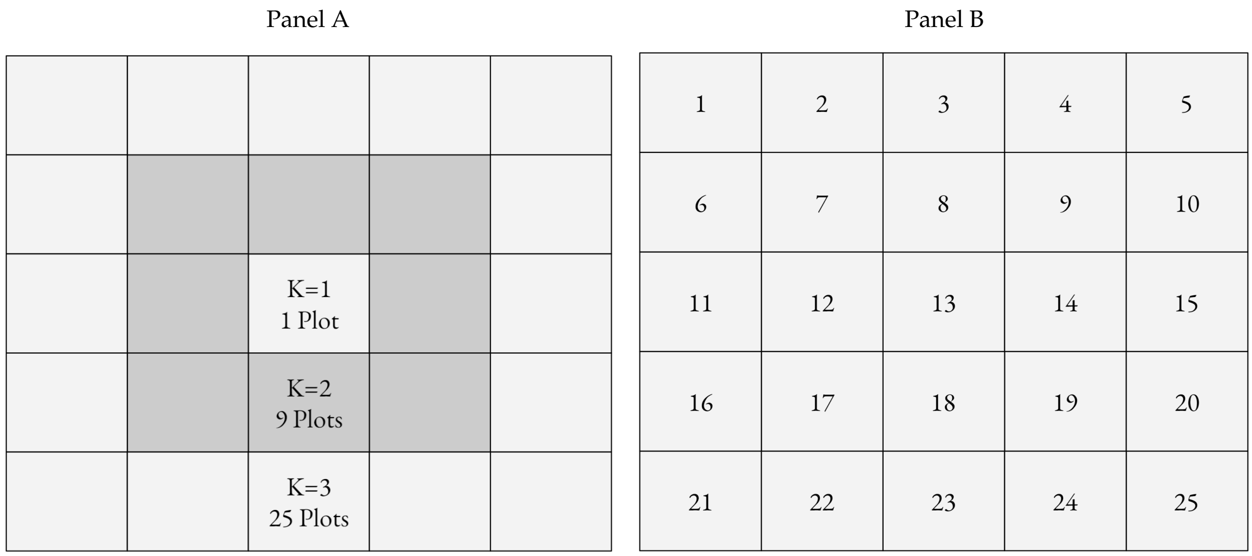
Differences between this anticipated production derived from fixed geographic characteristics and actual production reported by MOFA constitute the insurable yield shocks analyzed in this study.

Note that this calculation treats nation-wide productivity fluctuations as uninsurable variation embedded into $\gamma_i$. 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 $\mu_{it}$ have the same variance as the characteristic shocks $\omega_{it}$ because there are $(2K-1)^2$ tracts in a $K$-sized ball. However, there is spatial correlation in $\mu_{it}$ 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

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

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

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

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

The variance of each of these terms is $\sigma^2$. 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 $\omega_{it}$ have no physical interpretation. They do not, for example, represent spillovers from nearby rainfall or pests. The use of $\omega_{it}$ is merely a modeling technique to describe correlation in productivity shocks $\mu_{it}$ 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 $(\sigma, K)$ 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 ($\mu_{it}$) in the district, which can in turn be written as a weighted average of

Table B.1: Parameter Estimates and Log Likelihoods

| $K$ | $\sigma$ | 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 ($\omega_{it}$) 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 $\mu_{dt}$ is the sum of independent, normally distributed variables $\omega_{it}$. 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 $(\sigma, K)$ by the overlapping $\omega_{it}$ 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 $\omega$ 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 $\omega_{it}$, 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.