Risk, Agricultural Production, and Weather Index Insurance in Village South Asia

Jeffrey D. Michler[1], Frederi G. Viens[2], and Gerald E. Shively[3]

[1] *Department of Agricultural and Consumer Economics, University of Illinois, Urbana, IL*
[2] *Department of Statistics, Purdue University, West Lafayette, IN*
[3] *Department of Agricultural Economics, Purdue University, West Lafayette, IN*

Abstract

We investigate the sources of variance in crop output and measure their relative importance in the context of weather index insurance for smallholder farmers. We use parcel-level panel data from South Asia and a multilevel modeling approach to isolate the different sources of variance. We then measure how large a role weather plays in explaining variance in yields. Using Bayesian methods, we draw the underlying distribution of the random error term responsible for weather uncertainty, which is highly skewed and non-normal. We find that variance in weather accounts for a small but important fraction of total variance in crop output. We also derive pricing and payout schedules for actuarially fair weather index insurance. Our results shed light on the low uptake rates of index insurance in South Asia and provide direction for designing index insurance with less basis risk for farmers.

1 Introduction

Agricultural production is complex and risky and weather is just one of several potential causes of yield variability. Other determinants of variability include the quantity and quality of inputs, the soil characteristics of farmed parcels, pest pressure, the inherent or learned abilities of farmers, their situations, and the policy environment in which they operate (Hardaker et al., 1997). In this paper we use parcel-level panel data from South Asia to measure the different sources of variability in crop output and assess their relative importance. Determining just how large a role weather plays in explaining variability in yields is vital in designing and implementing weather index insurance, an increasingly popular micro-level intervention in developing countries. Recent randomized control trials in South Asia have found limited farmer uptake of index insurance. Using a multilevel modeling approach and Bayesian estimation we find that a small but important fraction of parcel-level yield variability can be attributed to weather. This may partly explain the low uptake of weather index insurance by farmers.

The goal of weather index insurance is to assist farmers in managing covariate risk. In developing country agriculture, covariate risk is particularly difficult to informally insure against since, by its very nature, it affects neighboring households and limits the effectiveness of traditional risk sharing mechanisms. Several early studies looked at the role of weather risk on agricultural production in South Asia. Townsend (1994), in his study of risk and insurance, found that household consumption moves with village-level consumption and is not much affected by idiosyncratic shocks. What idiosyncratic risk exists is covered by households via credit markets, gifts, and asset sales. Townsend (1994) concludes that households are able to insure against non-covariate risk but that covariate risk remains a problem. Along similar lines, Rosenzweig and Binswanger (1993) focus on the role of weather risk on production and asset portfolios. They find that village-level rainfall variables explains only a small proportion of household-level profit variability. But, similar to Townsend (1994), they find households insure against non-covariate risk. Using the same ICRISAT data from villages in India, Townsend (1994) and Rosenzweig and Binswanger (1993) come to the same conclusion, namely that weather risk plays only a small role in determining income variability. However, households are much less successful in insuring against covariate weather risk than they are in insuring against non-covariate risk. This conclusion motivates recent attempts to design and implement weather-based index insurance for poor farmers.

The apparent need for weather index insurance has been met by surprisingly little uptake of the product, at least by farmers in India (Giné et al., 2008, 2012). There are many potential barriers to uptake of insurance, including trust in the insurance company, household liquidity constraints, and lack of financial literacy of household decision makers (Cole et al., 2013). In addition to these consumer-centric explanations for low uptake, flaws in the product itself may be a cause. Uninsured exposure to basis risk is a case in point. There are two potential sources of basis risk. One is the loss caused by an event not measured by the index. Low temperatures that retard crop growth or high winds that cause crop damage when the insurance index is based on rainfall. Another potential source of basis risk is low correlation between the index and covered losses. This can occur, for example, when measurement stations are spatially distant from insured crops. A less common example is potentially insured crops that are sufficiently resilient to the phenomena which forms the basis for the insurance, such as when a drought-tolerant crop variety is covered by rainfall-based insurance. These demand side considerations ought to be important when designing weather index insurance to ensure barriers to household uptake are minimized.[1]

The most common type of weather index insurance in South Asia is rainfall insurance (Barnett and Mahul, 2007; Giné et al., 2007; Akter et al., 2009). However, surprisingly little empirical research has been conducted to confirm the underlying correlation between rainfall variability and yield variability. Giné et al. (2012) claim "around 90% of variation in Indian crop production levels is due to rainfall volatility." They cite Parchure (2002) as a source but Parchure's actual assertion is that of crops lost for weather reasons, 90% of the losses are due to rainfall variance. Barnett and Mahul (2007) claim "rainfall variability accounts for more than 50% of the variability in crop yields" but fail to cite a source or provide data. Even less data exist on the correlation between yield variance and other measures of covariate risk, such as temperatures, wind speed, or the occurrence of natural disasters. Finally, to our knowledge, no data exist on covariate risk's share in explaining variance in crop yield. Previous studies which have attempted such measurements were constrained by data, econometric techniques, and computing power so that they were unable to net out measurement error and fully describe the determinants of output variability (Rosenzweig and Binswanger, 1993). We attempt to fill this research gap.

We examine the different sources of yield variance using a multilevel/hierarchical regression approach. This allows us to control for inputs at the parcel-level and also isolate the

[1] While household characteristics such as financial literacy are difficult to address in designing index insurance, basis risk is not. The design of the insurance contract should attempt to minimize basis risk by verifying that (1) the index measure is the primary driver of yield variability and (2) the index measure is highly correlated with output loss. Chantarat et al. (2013) develop a regression based approach to minimize basis risk for asset insurance in Kenya.

amount of yield variance due to parcel-level effects, household-level effects, and seasonal effects. Using Bayesian methods, we draw the underlying distribution of the random error term corresponding to different sources of risk. Bayesian methods are particularly useful because the underlying distribution for seasonal variability is highly skewed and non-normal. Overall, we find very low correlation between weather and yield variability in the data. And considering all sources of yield variance, variance due to weather is found to play a small role.

However, our results should not be interpreted as arguing that smallholder farmers do not need weather insurance. Rather, what they need is better insurance. The the skewness in the posterior distribution for weather variability highlights the need for access to affordable risk-management tools. In other words, despite the low explanatory power of seasonal variations in yield variability, our results still suggest that extreme weather events, while rare, are nevertheless potentially costly. We provide a brief quantitative assessment of index insurance contracts currently offered for sale in village South Asia by the insurance company ICICI Lombard. We conclude that these contracts are overpriced. This underscores that farmers may rationally prefer to focus their risk management choices on minimizing other sources of risk instead of purchasing weather index insurance, as currently marketed.

To tie our results as closely as possible to the issue of weather risk in agricultural production, we use the ICRISAT household survey data from the same Indian villages studied by Townsend (1994) and Rosenzweig and Binswanger (1993). However, we use an expanded panel dataset covering both India and Bangladesh. We use a fairly short panel of eight cropping seasons in our analysis. This short horizon has both advantages and disadvantages. The primary benefit is that over such a short time frame parcel and household characteristics are unlikely to have changed. By controlling for parcel and household effects, along with parcel-level inputs, we can measure the variance in crop output resulting from covariate risk. The primary liability of our dataset is that the two and a half years of production that we observe does not contain any large-scale weather related natural disasters, such as catastrophic drought or flooding. This likely means that the degree of correlation we measure between weather and yield variability is an underestimate of the true correlation. Nevertheless, our analysis sheds new empirical light on the potential value of weather insurance during typical production periods.

By using a multilevel approach to measuring the roles of idiosyncratic and covariate risk in agricultural production in village South Asia we contribute to three separate streams of literature. The first is the literature focused on production risk and insurance in the developing world. This literature is large and growing. Our aim is to connect to two strands of this literature. The first (Townsend, 1994; Rosenzweig and Binswanger, 1993; Rosenzweig and Wolpin, 1993; Giné and Yang, 2009) has focused on the theoretical and empirical relationship between production, risk, and insurance. The second (Barnett and Mahul, 2007; Giné et al., 2007; Skees et al., 2007; Giné et al., 2008; Akter et al., 2009; Gaurav et al., 2011; Clarke et al., 2012; Cole et al., 2013; Clarke et al., 2012) focuses on interventions to help insure production against risk.

Our second contribution is to the nascent economics literature that applies multilevel models to household data. Although we introduce no new methods, per se, in the paper, we do contribute a methodological innovation by expanding the use, understanding, and adaptability of multilevel modeling, Bayesian inference, and Gibbs sampling, all in the context of household data.

The third body of literature to which we contribute is the empirical estimation of production functions. Until recently it has been computationally difficult to estimate production functions with multiple nested levels of data. While panel data random effects methods are equivalent to a multilevel model with a single level, few alternatives exist for models with a high number of nested levels. Production data are often collected at the farm or factory level, which suggests at minimum a two level model. Adding additional levels to account for and distinguish among temporal and/or spatial variation can provide more efficient estimation of production. We contribute to this strand of literature by comparing production functions parameter estimates obtained by standard ordinary least squares and multilevel regression.

2 Data

To conduct our empirical analysis, we use household data from villages in India and Bangladesh. These data were collected as part of the Village Level Studies/Village Dynamics Study of South Asia (VDSA, 2013). The data set combines high and low frequency household data from 12 Bangladeshi villages and 30 Indian villages. The Indian villages include the three studied by Townsend (1994) and the ten studied by Rosenzweig and Binswanger (1993). However, while much of the previous work has relied on the low frequency data, we utilize a newly available high frequency data set. These data include monthly household observations on input purchases, labor expenditure for on-farm activities, and crop yields. The added value of high frequency data is that, with multiple crops on multiple parcels for multiple seasons in a single year, we can glean much more detailed and accurate farm production data than afforded by the low frequency data.

We utilize monthly parcel-level data aggregated to the seasonal level. We focus solely on rice production, which accounts for 67 percent of all parcel-level observations. This leaves us with 9,462 observations from 4,407 unique parcels from 910 distinct households across eight seasons. We exploit this nested data structure in our empirical analysis.

While Bangladesh accounts for the majority of observations, the number of parcels and households are fairly evenly divided: of 4,407 unique parcels, 2,626 are in Bangladesh and 1,781 are in India. Among the 910 households in our dataset, 399 come from Bangladesh and 511 come from India. Despite accounting for only 30 percent of villages in the survey, our focus on rice production means that 75 percent of total observations are from Bangladesh. Compared to those in India, household in Bangladesh grow rice in more seasons.

Statistics describing these data are presented in Table 1. Average yield of rice is 3,874 kilograms per hectare. On average, yields in Bangladesh (4,122 kg/ha) are higher than those in India (3150 kg/ha). Bangladesh's higher yields are generally due to more intensive use of inputs. Bangladeshi households use more fertilizer, irrigation, mechanized machinery, and pesticide than their Indian counterparts. Indian households use more labor, on average, than Bangladeshi households. On average, parcels are much smaller in Bangladesh (0.098 ha) than in India (0.259 ha).

3 Empirical Strategy

3.1 Ordinary Least Squares

We begin by estimating a simple linear regression for yield. Let y_{it} denote the log of yield for parcel i at time t. We estimate

$$y_{it} = X_{it}\beta + \epsilon_{it} \tag{1}$$

where X_{it} is a matrix of data (including a constant), β is a vector of coefficients to be estimated, and ϵ_{it} is an error term which we assume is i.i.d. $N(0, \sigma_\epsilon^2)$. While this specification allows us to estimate the impact of parcel-level inputs on yield at each point in time, it does not control for variation in parcel yields across time. To gain a sense of how seasonal changes in weather impact yields and the marginal value of inputs, we estimate a random effects model,

$$y_{it} = X_{it}\beta + s_t + \epsilon_{it} \tag{2}$$

where s_t is i.i.d. $N(\bar{s}, \sigma_s^2)$. This random effects or error component model takes into account that variation in yield is determined by parcel and time specific inputs X_{it},[2] a time-specific term s_t, and the idiosyncratic error term, ϵ_{it}. This specification allows us to focus on time specific means instead of an overall mean (the constant term), which makes sense since yields may differ across seasons.

We note several drawbacks associated with this linear estimation of the production function. First is that we are unable to estimate the role of weather in determining variations in yields. Equation (2) produces estimates of how each season affects yields. But this information does not allow for comparison of the effects of weather on yield relative to other inputs, but only the role of weather in determining each season's yield. A second drawback is that OLS limits our ability to control for additional clustered effects, such a parcel or household-level effects. While changes in season clearly impact the effectiveness of parcel-level inputs, equally relevant effects may exist at the parcel or household-level. Some households may be more efficient in their application of labor compared to others, while some parcels may be of better quality, resulting in less need for fertilizer. And even if it were computationally feasible to estimate time-specific, household-specific, and parcel-specific parameters using OLS, such grouped data would violate the assumption of independence for all data (Corrado and Fingleton, 2012). A final drawback, and perhaps the most important, is that OLS assumes the time-specific term is normally distributed. While in many applications this assumption is appropriate, we should not expect weather events to be normally distributed. If time-specific effects have a highly skewed distribution, OLS will produce biased parameter estimates.

3.2 The Multilevel Model

A multilevel or hierarchical modeling approach addresses the first two drawbacks associated with the standard linear approach to estimation. First, multilevel models offer a natural way to assess the role of changes in weather on variation in yields by explicitly modeling the variance and not just the mean of the data. Such models allow us to estimate the share in total variance attributed to variability in each group. In our case, a multilevel approach also allows us to disaggregate total variance into its sources, so as to measure the relative contribution of seasonality and weather risk in production. Second, a multilevel approach

[2] In this specification the data matrix X_{it} no longer includes a constant term.

allows us to incorporate random intercepts for each grouping of the data without adding to the computational burden and without violating independence assumptions.

Gelman and Hill (2007) provide an introduction to multilevel analysis. We start with a simple two level model in which parcels are grouped within seasons. Let y_{it} denote the log of yield for individual parcel i coming from group t. We estimate

$$y_{it} = X_{it}\beta + \alpha_t + \epsilon_{it} \tag{3}$$

where X_{it} (with a constant), β, and ϵ_{it} are as previously defined and α_t is modeled by a group level regression

$$\alpha_t = \mu_t + \nu_t. \tag{4}$$

Here the group level intercepts are functions of a group mean, μ_t, and an error term, ν_t which is assumed to be i.i.d. $N(0, \sigma_\nu^2)$. This specification allows us to estimate the individual as well as the group level error term and calculate each level's share of the total variance. This calculation, called the intraclass correlation coefficient (ρ), is similar to the proportion of explained variance in OLS regression, and is calculated as:

$$\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\epsilon^2}. \tag{5}$$

As a matter of practicality there is little difference between equation (2) and equation (3). However, the value of a multilevel model becomes obvious as we seek to add additional levels.[3] In our parcel-level panel data analysis we can view each observation, n, as coming from a parcel group i; each parcel group as being nested within a household, h; and each household

[3]The model represented by equations (6a)-(6d) can be summarized as a single model with a matrix β, an intercept term μ_t, and a special noise structure which is a sum of independent noise terms with a nested dependence on indexes. Considered as a single noise term depending on ν_{iht}, this noise structure would have a covariance matrix which is the superposition of triangular structures. We find representation of the model by multiple equations to be more intuitive.

as observed within a time group, t. We can write the multilevel model as:

$$\textbf{Level 1}: y_{niht} = X_{niht}\beta + \alpha_{iht} + \epsilon_{niht} \tag{6a}$$

$$\textbf{Level 2}: \alpha_{iht} = \alpha_{ht} + \nu_{iht} \tag{6b}$$

$$\textbf{Level 3}: \alpha_{ht} = \alpha_{t} + \nu_{ht} \tag{6c}$$

$$\textbf{Level 4}: \alpha_{t} = \mu_{t} + \nu_{t}. \tag{6d}$$

Level 1 of the model estimates the log of yield as a function of inputs applied by household j to that specific parcel i at the given time t. Each n observation comes from a parcel cluster i which we assign a unique intercept, α_{iht}. This parcel-level intercept allows the relationship between inputs and yield to vary across parcels depending on parcel-level characteristics. While many parcel-level characteristics, such as soil color, may be observable, many other parcel-level characteristics are difficult to measure or costly to observe and thus remain unobserved to the econometrician. Such characteristics include soil micro-nutrients, grade, and aeration or composition. By including a unique intercept term for each parcel we can control for these parcel-level characteristics.

Level 2 of the model groups parcels within households. Here parcel intercepts, α_{iht}, are a function of household characteristics, α_{ht}, and a random disturbance term, ν_{iht}. The household-level intercept allows variation in parcel-level production to be dependent on household characteristics. In most production regressions there is an attempt to control for unobserved household ability through proxy variables such as age or education. Our multilevel approach allows us to control for any unobserved household-level characteristic by assigning each household a unique intercept term without the need to rely on proxies.

Level 3 of the model groups households within time. Here household-level intercepts, α_{ht}, are a function of time, α_{t}, and a random disturbance term, ν_{ht}. The time-level intercept allows variation in household-level efficiency to depend on seasonality. While household ability is often viewed as time invariant, in fact it can be viewed as time dependent. Households may increase in ability through education or experience. Additionally, household ability may be diminished or enhanced by changes in weather. Households with experience in dealing with drought conditions may find their ability diminished by flooding or cyclones. *A priori*, there is no reason to assume that seasonality or changes in weather have a monotonic effect on household characteristics. By allowing household-level intercepts to vary across time, we are relaxing the assumption that household characteristics are either time invariant or vary monotonically.

Level 4 of the model defines time-level intercepts as a function of a season mean, μ_t, and an error term, ν_t, as previously discussed. At all levels of the model the random disturbance terms, ν are assumed to be i.i.d. with mean zero and variance σ^2_ν.

We can verify if the normality assumption is reasonable by drawing profile zeta plots for each parameter, including the disturbance terms. Profile zeta plots are a way to visualize the sensitivity of the model fit to changes in values of particular parameters. While these plots are not the same as drawing the underlying distribution of an estimator, they represent a similar idea and can be interpreted as representing the underlying distribution. First, we estimate the model in equations (6a)-(6d). Then we hold a single parameter fixed and vary the other parameters, assessing the fit of each new iteration compared to the globally optimal fit. Our comparison statistic is the likelihood ratio test. We then apply a signed square root transformation to the LR statistic and plot the absolute value of the resulting function, $|\zeta|$, in comparison to the estimated parameter values. The zeta profile plots resulting from the model fit to equations (6a-d) are presented in Figure 1. Parameters with underlying normal distributions will have straight line zeta profile plots. When this is the case these parameters provide a good approximation to the normal distribution and standard confidence intervals can be used for inference. When zeta profile plot lines are not straight, the normal distribution is a poor approximation of the underlying distribution. For parameter estimates on data, a non-normal distribution simply requires an adjustment of the relevant test statistics. But, if the underlying distributions of our disturbance parameters are non-normal, we are faced with a violation of the normality assumptions of the error term. In Figure 1, zeta profile plots for the parcel and household disturbance terms, along with the idiosyncratic error term, represent good approximations of normality. However, the zeta profile plot for the disturbance term at the time level, ν_t, is highly skewed. Thus, we cannot assume that ν_t is distributed $N(0, \sigma^2_{\nu_t})$. Since we do not, *a priori*, know the density of ν_t, maximum likelihood methods cannot be used to estimate our model.

3.3 Bayesian Estimation of the Multilevel Model

While multilevel models address the first two drawbacks of OLS estimation, they still rely on the potentially unsupported assumption of normality of the error term at each level. We address this drawback to classical estimation of multilevel models by adopting a Bayesian framework and using Markov Chain Monte Carlo (MCMC) methods to obtain posterior estimates. For estimation of posterior distributions we use the Gibbs sampler, which iteratively constructs a sequence of samples from the univariate random values of each response variable, and of each model's parameters, conditional on all other parameters and variables. This method allows us to compute all features of the marginal and joint distributions because the marginal samples are iteratively fed back into the conditional posterior densities of all other parameters and variables for each sampling. This allows us to calculate unbiased point estimates and confidence intervals for all variables without recourse to normality assumptions, and indeed, these empirical data-driven estimations of some of the posterior densities will show non-normal characteristics.

Hanmaker and Klugkist (2011) provide an introduction to Bayesian estimation of multilevel models while Cameron and Trivedi (2005) provide a more detailed, and more technical, presentation. In this section, we first define the Bayesian estimator using the simple example of a two level hierarchical model. We then provide parameter definitions for our multilevel model first defined in equations (6a)-(6d).

For a general two level model, let y_{km} be a unique observation $k = 1, ..., K$ from the group $m = 1, ..., M$. Within each group, m, the data are distributed according to a particular distribution G with parameter γ such that $y_{km} \sim G(\gamma_m)$. We then assume that parameter γ_m comes from a distribution L with parameter λ such that $\gamma_m \sim L(\lambda)$. Finally, we assign a distribution to the hyperparameter λ so that $\lambda \sim Q(a, b)$. Given this setup, we can obtain the joint posterior distribution of all unknown parameters using Bayes' theorem:

$$p(\lambda, \gamma|y) \propto p(y|\gamma, \lambda)p(\gamma|\lambda)p(\lambda) \tag{7}$$

where $y = (y_{11}, ..., y_{1K}, ..., y_{M1}, ..., y_{MK})$ is all the data, $\gamma = (\gamma_1, ..., \gamma_M)$ is the group level parameter, and λ is the population parameter. The density of the data is obtained through

$$p(y|\gamma, \lambda) = \prod_{m=1}^{M} \prod_{k=1}^{K} p(y_{km}|\gamma_m, \lambda). \tag{8}$$

This is an independence assumption, and the individual density, $p(y_{km}|\gamma_m, \lambda)$, is assumed to be known. If this is the case, the prior for the group level effect is

$$p(\gamma|\lambda) = \prod_{m=1}^{M} p(\gamma_m|\lambda). \tag{9}$$

From the joint posterior distribution we can derive the conditional posterior distributions.

We can now apply our simple example of the Bayesian estimator to the multilevel model given in equations (6a)-(6d). We redefine the model in probability terms:

$$\textbf{Level 1}: \; y_{niht} \sim N\left(X_{niht}\beta + \alpha_{iht}, \sigma^2\right) \tag{10a}$$
$$\textbf{Level 2}: \; \alpha_{iht} \sim N\left(\alpha_{ht}, \tau_i^2\right) \tag{10b}$$
$$\textbf{Level 3}: \; \alpha_{ht} \sim N\left(\alpha_t, \tau_h^2\right) \tag{10c}$$
$$\textbf{Level 4}: \; \alpha_t \sim N(\mu_t, \tau_t^2). \tag{10d}$$

We next define the hyperpriors, or the prior distributions of the hyperparameters, of the model:

$$\begin{aligned} \textbf{Hyperpriors}: \quad & \mu_t \sim N\left(a, b^2\right) \\ & \tau_t^2 \sim IG\left(d_t, g_t\right) \\ & \tau_h^2 \sim IG\left(d_h, g_t\right) \\ & \tau_i^2 \sim IG\left(d_i, g_i\right) \\ & \sigma^2 \sim IG\left(q, r\right). \end{aligned} \tag{11}$$

We select values for the distributions to ensure that priors are uninformative. We then use a burn-in period of 50,000 iterations and an additional 50,000 iterations to ensure convergence.

4 Econometric Results

We present the results from a large complement of regressions in Table 2. Models 1 and 2 are classical OLS regressions. Models 3-7a are multilevel models. All models are estimated using the log of yield as the dependent variable and logged value of inputs and independent variables. Hence, point estimates can be read directly as elasticities.[4]

We estimate the OLS regressions with complete pooling (Model 1) and with time-varying random effects (Model 2). Model 3 is multilevel model with covariates and data clustered at the season-level and is included for comparison with Model 2. Results from the first three models point to a fairly robust set of basic production patterns in the data. These include (i) positive and significant production relationships between yield and measured inputs (labor,

[4] Also included, but not reported in Table 2 is an indicator variable to control for differences in production between countries.

fertilizer, mechanization and pesticides); (ii) modest increasing returns to labor use, and overall increasing returns to scale; and (iii) strong seasonal patterns in yields. To facilitate the construction of our measures of variance decomposition and intraclass correlation coefficients (ICC), we also include results from four additional multilevel models. Model 4a contains no covariates and uses data clustered at the season-level; Model 5a contains no covariates and uses data clustered at the season and parcel-level; Model 6a contains no covariates and uses data clustered at the season, parcel, and household-level; and Model 7a contains covariates and uses data clustered at the season, parcel, and household-level.

Table 3 reports the residuals (panel A), ICCs (panel B) and variance shares (panel C) for Models 4a-7a. These statistics establish the key findings that inform our insights into the potential role of weather index insurance. We focus attention on panel C of Table 3, which compactly summarizes the model results expressed in the upper panels of the table. Reading across the columns of the table provides insights into the additional explanatory information that is provided by adding new hierarchical levels to the regression. Reading down the rows of the table allows us to assess the decomposition of variance and, by extension, the relative importance of each level in explaining overall variance in yields. So, for example, in a model with no covariates and using data clustered at the season-level only (Model 4a) we find that 12 percent of variance occurs across season, and 88 percent is left unexplained. Adding a household-level (Model 5a) re-attributes the variance such that 20 percent consists of between-household variance. Eight percent is between-season variance, and 72 percent remains unexplained. Adding a parcel-level to the model (Model 6a) results in 42 percent of variance being attributed to between-parcel differences, 15 percent being attributed to between-households differences, and 7 percent being attributed to between-season differences; 37 percent remains unexplained. Finally, Model 7a measures the share of yield variability accorded to each level while controlling for time-specific parcel-level inputs. In this, our most complete specification of the production function, 26 percent of variance comes from between-parcel differences, 25 percent comes from between-households differences, and 3 percent is attributed to between-season differences; 46 percent remains unexplained. These basic patterns highlight the relatively small importance of between-season yield variance compared with between-household or between-parcel yield variance.[5]

[5] Note that the share of variance that remains unexplained is not the appropriate measure of model fit. The reported residuals in Panel A of Table 3 lists the size of the residual or noise term in the regression, which is always decreasing as we add complexity to the model. Additionally, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) reported in Table 2 provides a measure of model fit. In both the AIC and BIC, lower values correspond to better model fit.

We present mean point estimates based on a parallel set of full Bayesian posterior density estimates for the four multilevel regressions in Table 4. Model 4b omits covariates and uses data clustered at the season-level; Model 5b contains no covariates and uses data clustered at the season and parcel-level and Model 6b contains no covariates and uses data clustered at the season, parcel, and household-level. Model 7b contains covariates and uses data clustered at the season, parcel, and household-level. The Bayesian estimation of the production function with a full set of covariates (Model 7b) generates point estimates that are virtually indistinguishable from those of the OLS and non-Bayesian multilevel regressions. As in the case of the fully specified non-Bayesian multilevel regression (Model 7a in Table 2), only the point estimate for irrigation is negative.

In general, the non-normality of the error term means that the classical and multilevel regression models are misspecified and we would expected point estimates from these models to be biased. However, in our application, only the seasonal noise term is not normally distributed. Furthermore, seasonal variance makes up only a small share of total variance. Thus, bias in the point estimates introduced through model misspecification is mitigated by the small impact our skewed noise term has in the total estimation procedure. We highlight that this is an artifact of the current application and data. If intra-seasonal variation accounted for more of the total variance we would expect a larger bias in our point estimates using classical regression techniques.

In parallel to Table 3, Table 5 reports the residuals (panel A), ICCs (panel B) and variance shares (panel C) for the Bayesian regressions reported in Table 4. These statistics confirm in general the findings from before and provide a strong test of the robustness of those results. We again focus attention on Panel C, where patterns are nearly identical to those reported in Table 3. In a model with no covariates and using data clustered at the season-level only (Model 4b) we find that 18 percent of variance occurs across seasons, and 82 percent is left unexplained. Adding a household-level (Model 5b) re-attributes the variance such that 19 percent consists of between-household variance, 13 percent is between-season variance, and 68 percent remains unexplained. Adding a parcel-level (Model 6b) results in 41 percent of variance being attributed to between-parcel differences, 14 percent being attributed to between-households differences and 10 percent being attributed to between season differences; 35 percent remains unexplained. Finally, although not strictly comparable to the previous models due to the inclusion of household and parcel-level covariates, Model 7b indicates that only 7 percent of the total variance can be explained by season, compared with 25 percent and 24 percent for parcel and household. Again we attribute a small role to between-season yield variance compared with between-household and between-parcel yield variance. A notable difference between the estimated residuals from the classical and Bayesian regressions are that the MLE estimation consistently underestimates the significant role seasonal effects play in yield variance. So, while the skewness of the seasonal residual does not have a pronounced effect on point estimates (due to the small size of the residual), estimation of that relative size of the seasonal residual consistently differs across estimation techniques.

Figure 2 reports distribution profiles based on Model 7b for the associated parcel-level, household-level, time-level, and unexplained variance. These histograms are drawn from the posterior distributions estimated using the Gibbs sampler. Posterior distributions for the household-level, parcel-level, and unexplained variance are all close approximations to the normal distribution. However, the posterior distribution of the time/season variance term is highly skewed. We can draw several conclusions from this graph. First, τ_t^2 is not normally distributed, violating the normality assumption in OLS and multilevel regression. Second, normal confidence interval calculation will be inaccurate. Third, the long right hand side tail in the distribution indicates that there is a small but positive probability of extreme between-season variation. While our mean estimate for the share of yield variance due to weather is small, this estimate does not account for the small but positive probability of large yield variance due to extreme weather events. Thus, our mean estimate of seasonal variance explaining less than 10 percent of total variance is most likely an underestimate.

5 Actuarial Values of Rainfall-Index Insurance Risk

The likely importance of extreme and infrequent weather events in yield variability, evidenced by the aforementioned skewness in the posterior distribution of τ_t^2, highlights the need for smallholder farmers to have access to affordable risk-management tools. In other words, despite the low explanatory power of seasonal variations in yield variability, our data still suggest that potential insurance purchasers need to consider weather risk, since extreme weather events, while rare, are nevertheless potentially costly over the long run. This motives us to ask whether tools for managing this type of risk are adequately accessible. We finish this paper with a brief quantitative assessment which suggests that rainfall index insurance, as currently marketed in village South Asia, may be overpriced.

Our parcel-level panel data are not adapted to measuring the severity of weather events. In order to estimate the actuarial cost of these events, we rely on rain-index insurance contracts which ICICI Lombard offers in the region and which form the basis for the analysis conducted by Cole et al. (2013). These contracts are district-dependent, and are the sum of three individual contracts in the three successive phases of the monsoon season. The start and finish dates defining each of the three phases are determined for each village for each year, as a function of the daily evolution of rainfall for the village's corresponding weather station.[6] In Phases I and II, corresponding to planting and heading, a drought risk exists, while in Phase III, corresponding to harvest, the risk is from excess rain.[7] In Phase I and II the contract pays zero Indian Rupee (Rs.) if the cumulative rainfall exceeds a certain "strike" value s. It then pays a number of Rs. which decreases linearly with the cumulative rainfall in that phase, and if this rainfall amount reaches down to or beyond a certain low "exit" number e, the contract pays 1000 Rs. The contract is structured so as to have a piecewise linear continuous payout. Consequently, the slope of the linear part of the contract function, between the rainfall values of s and e, is equal to the negative number $1000/(e-s)$.[8] In Phase III, the contract is structured in an opposite way: the payout equals zero for cumulative rainfall below the strike value s, it equals 1000 Rs. above the exist value e, and it grows linearly with positive slope $1000/(e-s)$ for cumulative rainfall between s and e. The linear portion of the contract structure means that its actuarial value will be a bona fide quantitative measure of the severity of drought and/or flood risk. This would not be the case if the contract paid only either zero or a fixed lump sum, since in that case, the actuarial value would only measure the probability of a payout, not drought/flood severity.

We utilize daily rainfall data from 16 of the VDSA villages selected to provide a wide geographic distribution of Indian States with tropical wet-and-dry and humid subtropical climates from Andhra Pradesh to Bihar for the years 2010 and 2011.[9] Our method of calculating the actuarial value of the payouts differs from that used in previous studies of the same contracts. Cole et al. (2013) provide a straightforward and representative explanation

[6] Specifically, Phase I begins on the day when accumulated rainfall since the traditional monsoon start date exceeds 50mm. In Eastern and Central India, monsoon start date is generally considered 1 June, so Phase I begins for a village when accumulated rainfall since 1 June exceeds 50mm. In Western India monsoon start date is figured as 1 July. If accumulated rainfall in the first month fails to exceed 50mm, then the first phase automatically begins on the first of the next month (1 July in E. and C. India and 1 August in W. India).

[7] Phase I and II are each 35 days in length. Phase II begins the day after the close of Phase I. Phase III begins the day after the close of Phase II. In contrast to the first two phases, Phase III lasts for 45 days: World Bank (2011).

[8] Note that, presumably because of a typographical error, this "payout slope" is reported incorrectly as being constant over all districts in Table 1 of Cole et al. (2013).

[9] The rainfall data is part of the VDSA collection process. It is ostensibly collected in every village on a daily basis, although several villages have numerous missing data.

of the standard method of calculation. They utilize 36 years of rainfall data from five villages in Andhra Pradesh and 38 years of rainfall data from three villages in Gujarat to calculate values for each of the three contract phases. We instead use two years of rainfall data to determine rainfall in each of the three phases in 16 villages. Because of the wide geographic distribution, this averaging encompasses a number of differing weather conditions, which can be considered as a proxy for following a specific village over a long period of time, and ensures that a sufficient number of extreme weather events are included, a particularly useful feature in this region where droughts are sporadic. Additionally, we believe that the shorter time horizon utilized in our calculation more closely approximates the decision horizon for smallholder farmers in South Asia. Given that the average age of the head of household in our data is 49 years, rainfall information from 38 years ago is not likely to figure into insurance purchase decisions.

Using this shorter time horizon combined with wider geographic diversity, we calculate an estimate of the actuarial value of the payouts for the 3-phase contract by averaging payouts over all 84 village-year-phase combinations. To illustrate the effect of how different strike and exit values effect the insurance contract's value, we repeat the calculation for three of the contracts reported in Cole et al. (2013), with low, medium, and high payout structures, and report on the payouts' actuarial values and payout probabilities. The results from this analysis are given in Table 6.

Under the most generous payout structure ("high payout"), the probability of payout is just under 12 percent. This represents, on average, little more than one payout in one phase over a three-year period. Judging by the examples of contracts reported in Cole et al. (2013), it is more likely that smallholders would be offered contracts in the medium payout range, where payout probability is just over 7 percent. Farmers who do not look ahead more than three years in planning their activities would presumably find little incentive to purchase insurance where, on average, this 7 percent figure means they would be as likely as not to see no payout within their planning horizon. The payout probability in the case of low-payout contract is even more extreme: 3.6 percent. We suspect that in the villages we consider, only the high payout contract would likely be of interest to farmers, and then only if they perceived its premium as close to fair. Table 6 reports those actuarially fair premiums.

Cole et al. (2013) also report expected payouts as percentages of true premiums paid to the ICICI Lombard insurance company for two villages in Andhra Pradesh in 2006. These numbers are equivalent to saying that the premium is computed by multiplying the actuarially fair premium by a loading factor, $1 + \lambda$, via the formula:

$$1 + \lambda = \frac{\text{paid price}}{\text{actuarially fair price}}. \tag{12}$$

Thus λ is a proportional transaction cost which, in principle, reflects the cost to the company for doing business. The two loading factors computed from reported premiums, $\lambda = 2.00$ and $\lambda = 1.34$, are very large if one compares with other weather insurance markets: a typical value of λ in the US and other countries is between 0.1 and 0.2. For instance, Giné and Yang (2009) use a value of $\lambda = 0.175$, while the Poverty Global Practice Group at the World Bank reports values in the US, Burkina Faso, and Senegal of $\lambda = 0.10$ (de Nicola, 2015). While it may not be possible for insurers to offer contracts with such low values of λ in India and Bangladesh, our actuarially fair value for the high-payout contract, 217.8 Rs., can be compared to the prices quoted by ICICI Lombard in Cole et al. (2013), which range between 260 and 340 Rs. A premium of 300 Rs. for the high-payout contract would thus represent a loading factor of $\lambda = 0.38$, a significantly higher value than those practiced in other regions, but still much lower than those inferred from the two villages reported in Cole et al. (2013). Such excessive loading factors appear to be a key element in preventing farmers in village South Asia from accessing insurance as a risk-management resource, and further explains the low uptake rate of rainfall-index insurance.

In addition to this pricing problem, a finer look at the payout probabilities per phase in our data indicate that in Phase III, there is essentially no chance of any payout, except for one payout event in the most generous payout structure. Because of the limited data, this is difficult to interpret, but could be an indication that the Phase-III portions of the contracts are not adapted to the farmers' needs.

6 Conclusion

Uptake of weather index insurance has been low in many low-income settings, including South Asia. Despite long standing evidence that households in rural South Asia are unable to fully insure covariate risk (Townsend, 1994; Rosenzweig and Binswanger, 1993), few studies have attempted to measure the role variance in weather plays in determining variability of crop output. In this paper we addressed this research gap using agricultural production data covering more than 4,000 parcels in India and Bangladesh. We examined the different causes of yield variance using a multilevel/hierarchical regression approach. This allowed us to control for inputs at the parcel-level and also isolate the amount of yield variability due to parcel-level effects, household-level effects, and seasonal effects. We adopt a Bayesian estimation approach to accommodate for the highly skewed distribution of the seasonal disturbance term. Overall, we find a low level of correlation between weather and yield variance, compared to other sources of variability. This suggests that farmers may rationally prefer to focus their risk management choices on minimizing other sources of risk instead of purchasing weather index insurance.

Although there are many potential impediments to insurance uptake, this research provides evidence for an obvious but until now overlooked explanation: insurance contracts are overpriced from the perspective of the farming household. Product design and ratemaking of index insurance in South Asia has generally taken a long view, utilizing rainfall data as far back as four decades to calculate actuarial rates (Parchure, 2002; Giné et al., 2007; Clarke et al., 2012). From the perspective of the insurance company, such a long time horizon is helpful to rely on the company's longevity to manage risk. However, risk diversification can also rely on geographic distribution. Moreover, given that the average age of the head of household in our data, rainfall information from when a household head was a child is not likely to figure into insurance purchase decisions. Taking a short-term perspective, we calculated actuarial value of payouts using rainfall data from recent seasons. We find that loading factors on these contracts are highly excessive and that this pricing problem provides little incentive for smallholder farmers to purchase insurance. These results suggest that micro-insurance markets hold promise for improving household risk management only to the extent that products are able to adapt to the planning horizons of potential purchasers.

References

Akter, S., R. Brouwer, S. Choudhury, and S. Aziz (2009). Is there a commercially viable market for crop insurance in rural Bangladesh? *Mitigation and Adaptation Strategies for Global Change 14*, 215–29.

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*(5), 1241–7.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Chantarat, S., A. G. Mude, C. B. Barrett, and M. R. Carter (2013). Designing index-based livestock insurance for managing asset risk in Northern Kenya. *Journal of Risk and Insurance 80*(1), 205–37.

Clarke, D. J., O. Mahul, K. N. Rao, and N. Verma (2012). Weather based crop insurance in India. Policy Research Working Paper, World Bank.

Clarke, D. J., O. Mahul, and N. Verma (2012). Index based crop insurance product design and ratemaking: The case of modified NAIS in India. Policy Research Working Paper, World Bank.

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*(1), 104–35.

Corrado, L. and B. Fingleton (2012). Where is the economics in spatial econometrics? *Journal of Regional Science 52*, 329–355.

de Nicola, F. (2015). Handling the weather: Insurance, savings, and credit in West Africa. Policy Research Working Paper, World Bank.

Gaurav, S., S. Cole, and J. Tobacman (2011). Marketing complex financial products in emerging markets: Evidence from rainfall insurance in India. *Journal of Marketing Research 48*, S150–62.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Giné, X., R. Townsend, and J. Vickery (2007). Statistical analysis of rainfall insurance payouts in Southern India. *American Journal of Agricultural Economics 89*(5), 1248–54.

Giné, X., R. Townsend, and J. Vickery (2008). Patterns of rainfall insurance participation in rural India. *World Bank Economic Review 22*(3), 539–66.

Giné, X., J. Vickery, L. Menand, and R. M. Townsend (2012). Microinsurance: A case study of the Indian rainfall index insurance market. In C. Ghate (Ed.), *Handbook of the Indian Economy*. New York: Oxford University Press.

Giné, X. and D. Yang (2009). Insurance, credit, and technology adoption: Field experiment evidence from Malawi. *Journal of Development Economics 89*, 1–11.

Hanmaker, E. L. and I. Klugkist (2011). Bayesian estimation of multilevel models. In J. Hox and J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis*, pp. 137–61. Hoboken: Taylor and Francis.

Hardaker, J. B., R. B. M. Huirne, and J. R. Anderson (1997). *Coping with Risk in Agriculture*. Wallingford: CAB International.

Parchure, R. (2002). Varsha bonds and options: Capital market solutions for crop insurance problems. *working paper*.

Rosenzweig, M. and H. Binswanger (1993). Wealth, weather risk, and the consumption and profitability of agricultural investments. *The Economic Journal 103*(416), 56–78.

Rosenzweig, M. R. and K. I. Wolpin (1993, April). Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India. *Journal of Political Economy 101*(2), 223–244.

Skees, J. R., J. Hartell, and A. G. Murphy (2007). Using index-based risk transfer products to facilitate micro lending in Peru and Vietnam. *American Journal of Agricultural Economics 89*(5), 1255–61.

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica 62*(3), 539–91.

VDSA (2013). Village Dynamics in South Asia (VDSA) database. generated by ICRISAT/IRRI/NCAP in partnership with national institutes in India and Bangladesh. (http://vdsa.icrisat.ac.in).

World Bank (2011). Weather index insurance for agriculture: Guidance for development practitioners. Agriculture and Rural Development Discussion Paper 50, World Bank Press.

Table 1: Descriptive Statistics

	Total	Bangladesh	India
yield (kg/ha)	3874	4122	3150
	(2145)	(1971)	(2450)
labor (hrs/ha)	998.8	975.7	1066
	(639.4)	(442.5)	(1013)
fertilizer (kg/ha)	314.2	341.4	234.8
	(252.1)	(250.0)	(241.3)
irrigation (liter/ha)	25.66	34.34	0.225
	(74.18)	(84.12)	(5.12)
mechanization (Rs/ha)	5631	5832	5039
	(4662)	(3886)	(6382)
pesticide (Rs/ha)	429.9	529.4	138.4
	(692.5)	(736.5)	(425.9)
parcel area (ha)	0.139	0.098	0.259
	(0.180)	(0.070)	(0.307)
rainfall (mm)	543.6	712.8	305.9
	(258.0)	(157.8)	(168.8)
number of observations	9462	7054	2408
number of parcels	4407	2626	1781
number of households	910	399	511

Table displays means of data with standard deviations in parenthesis. For inputs in value terms, we have converted Bangladeshi Taka (BDT) to Indian Rupee (Rs) using exchange rates at the start of the relevant monsoon season (1 BDT = 0.6744 Rs in 2010 and 0.5995 Rs in 2011). Rainfall is the sum of rainfall over the 70 days corresponding to phase 1 and phase 2 of rainfall index insurance contracts offered by ICICI Lombard.

Table 2: Results of Classical Estimation of Production Function

	OLS		*Multilevel Model*				
	(1)	(2)	(3a)	(4a)	(5a)	(6a)	(7a)
log labor	1.235***	1.248***	1.246***				1.438***
	(0.031)	(0.032)	(0.032)				(0.035)
log fertilizer	0.279***	0.278***	0.279***				0.348***
	(0.013)	(0.013)	(0.013)				(0.015)
log irrigation	−0.012	−0.016*	−0.014				−0.046***
	(0.008)	(0.009)	(0.009)				(0.010)
log mechanization	0.048***	0.050***	0.050***				0.053***
	(0.008)	(0.008)	(0.008)				(0.009)
log pesticides	0.014***	0.013**	0.013**				0.036***
	(0.005)	(0.005)	(0.005)				(0.006)
constant	1.635***		1.716***	3.707***	4.868***	4.827***	3.149***
	(0.218)		(0.229)	(0.497)	(0.131)	(0.136)	(0.119)
parcel						1.210	0.792
household					0.783	0.719	0.773
season			0.062	0.592	0.496	0.483	0.259
unexplained			1.384	1.605	1.483	1.135	1.044
Observations	9,462	9,462	9,462	9,462	9,462	9,462	9,462
R^2	0.341	0.970					
Log Likelihood			-16,527	-17,922	-17,668	-17,466	-15,980
Akaike Inf. Crit.			33,072	35,851	35,344	34,943	31,982
Bayesian Inf. Crit.			33,136	35,873	35,373	34,979	32,061

Note: All specifications include a statistically significant country indicator as a control. Columns (1) and (2) are classical OLS regressions without accounting for the multilevel structure of the data. Column (2) includes eight jointly significant season indicators as control variables. Columns (3a)-(7a) are multilevel models. Column (3a) is multilevel model with covariates and data clustered at the season-level. Columns (2), which includes season dummy variables, and (3a), which includes season-level clustering, are functionally equivalent to each other. Column (4a) contains no covariates and data clustered at the season-level. Column (5a) contains no covariates and data clustered at season and household-level. Column (6a) contains no covariates and data clustered at the season, household, and parcel-levels. Column (7a) contains covariates and data clustered at the season, household, and parcel-levels. Standard errors are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01)

Table 3: Residuals, ICCs, Variance Shares from Multilevel Model

	Multilevel Model			
	(4a)	(5a)	(6a)	(7a)
		Panel A *Estimated Residuals*		
parcel			1.210	0.792
household		0.783	0.719	0.773
season	0.592	0.496	0.483	0.059
residual	1.605	1.483	1.135	1.044
		Panel B *Intraclass Correlation Coefficients*		
parcel			0.418	0.271
household		0.200	0.566	0.528
season	0.120	0.281	0.632	0.530
		Panel C *Shares of Variance From Each Level*		
parcel			42%	27%
household		20%	15%	26%
season	12%	08%	07%	0.2%
unexplained	88%	72%	37%	47%

Note: Residuals in Panel A come from estimation of models in corresponding columns in Table 2. Residuals at each level correspond to σ_ν, or the explained portion of the total residual, while σ_ϵ is the unexplained portion of the total residual. Intraclass correlation coefficients in Panel B are calculated using the formula: $\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2+\sigma_\epsilon^2}$. Panel C decomposes the ICC into percent of variance accorded to each level.

Table 4: Results from Bayesian Estimation of Production Function

	Bayesian Estimates			
	(4b)	(5b)	(6b)	(7b)
log labor				1.442***
				(0.036)
log fertilizer				0.349***
				(0.015)
log irrigation				−0.046***
				(0.010)
log mechanization				0.054***
				(0.009)
log pesticide				0.036***
				(0.006)
α_{ht} (parcel)			0.963	−0.004
			(0.813)	(3.670)
α_t (household)		−0.073	−0.106	−2.854
		(0.367)	(0.987)	(3.010)
μ_t (season)	7.944	7.847	6.882	−0.690
	(0.322)	(0.445)	(1.582)	(4.574)
τ_i^2 (parcel)			1.212	0.787
			(0.029)	(0.034)
τ_h^2 (household)		0.793	0.719	0.777
		(0.037)	(0.042)	(0.034)
τ_t^2 (season)	0.769	0.646	0.632	0.081
	(0.321)	(0.260)	(0.285)	(0.040)
σ^2 (unexplained)	1.606	1.482	1.134	1.046
	(0.011)	(0.012)	(0.013)	(0.014)
Obs	9,462	9,462	9,462	9,462
DIC	35,823	35,559	32,284	30,191

Note: Results are Bayesian estimates of the multilevel models. Column (4b) contains no covariates and data clustered at the season-level. Column (5b) contains no covariates and data clustered at season and parcel-level. Column (6b) contains no covariates and data clustered at the season, parcel, and household-level. Column (7b) contains covariates and data clustered at the season, parcel, and household-level. Models were estimated using 4 chains, a burn-in of 25,000 iterations, and an additional 25,000 iterations for estimation. Standard errors are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01)

Table 5: Residuals, ICCs, Variance Shares from Bayesian Estimation

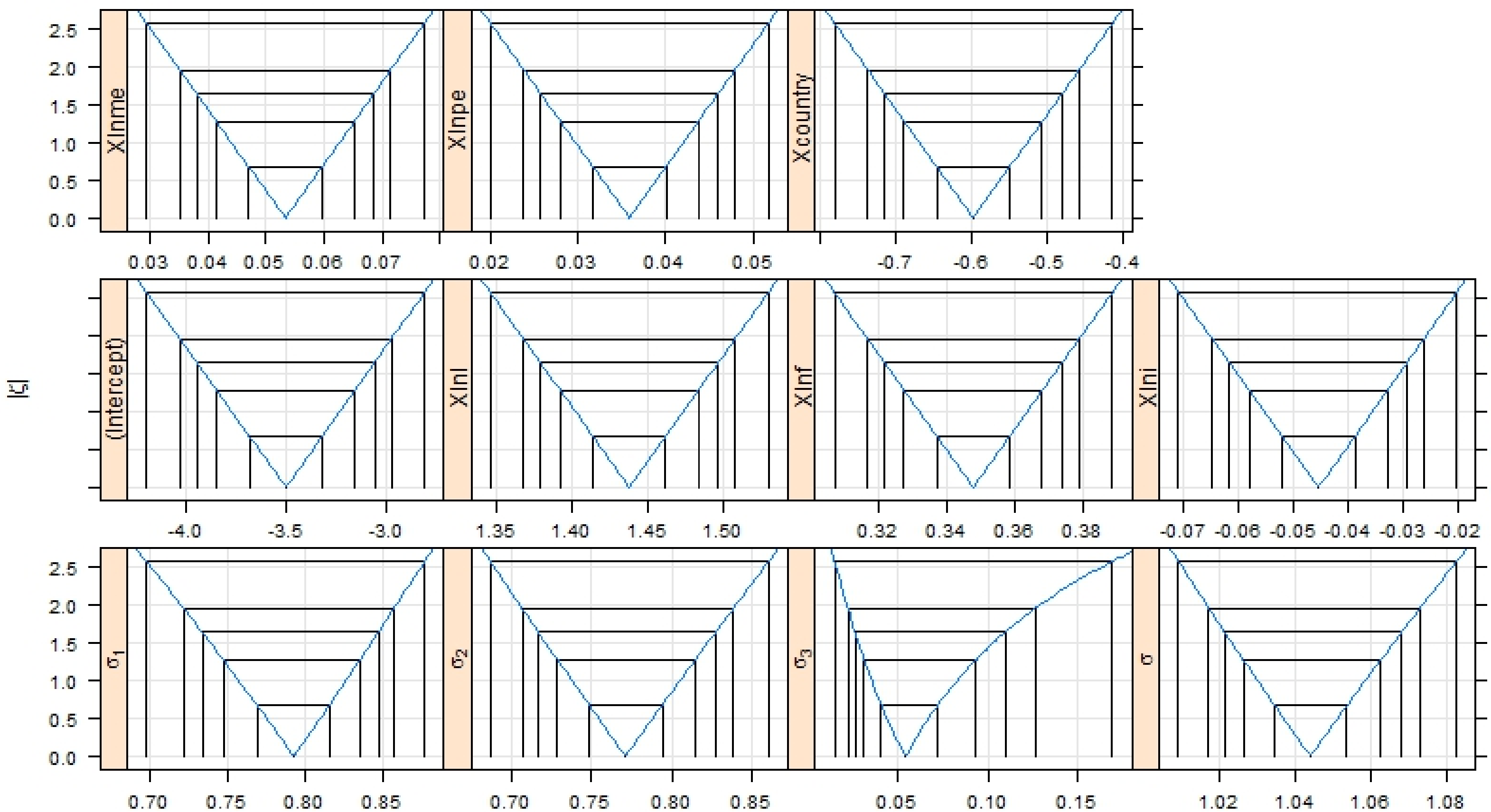
	Bayesian Estimates			
	(4b)	(5b)	(6b)	(7b)
		Panel A *Estimated Residuals*		
parcel			1.212	0.787
household		0.793	0.719	0.777
season	0.769	0.646	0.632	0.081
residual	1.606	1.482	1.134	1.046
		Panel B *Intraclass Correlation Coefficients*		
parcel			0.400	0.265
household		0.194	0.541	0.525
season	0.187	0.323	0.650	0.528
		Panel C *Shares of Variance From Each Level*		
parcel			40%	26%
household		19%	14%	26%
season	19%	13%	11%	0.3%
unexplained	81%	68%	35%	47%

Note: Residuals in Panel A come from estimation of models in corresponding columns in Table 4. Residuals at each level correspond to σ_ν, or the explained portion of the total residual, while σ_ϵ is the unexplained portion of the total residual. Intraclass correlation coefficients in Panel B are calculated using the formula: $\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2+\sigma_\epsilon^2}$. Panel C decomposes the ICC into percent of variance accorded to each level.

Table 6: Actuarial Values and Payout Probabilities

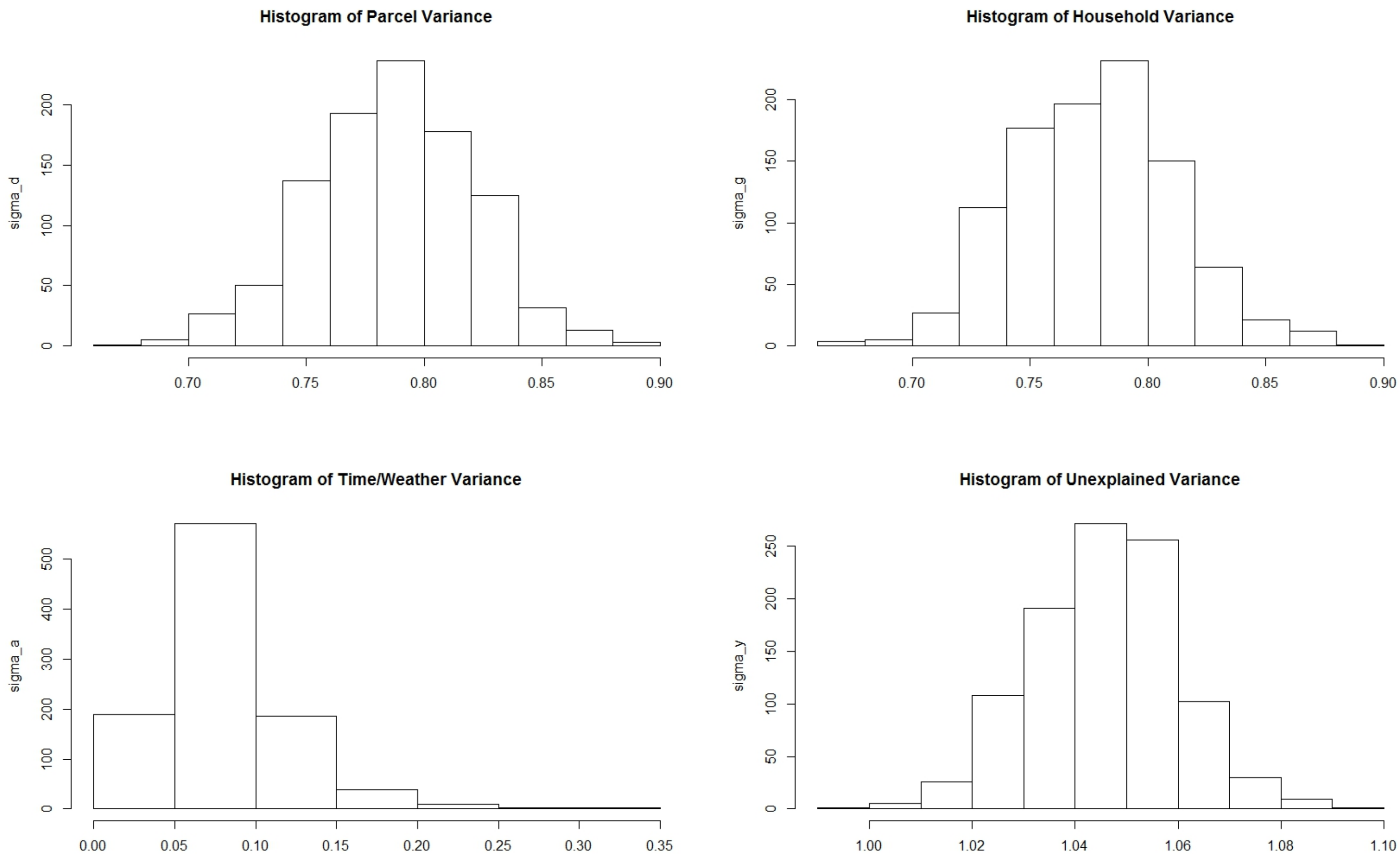
	Contract Structure			*Summary Statistics*	
	Strike	Exit	Max Payout	Actuarially Fair Premium	Probability of Payout
			Panel A *High-Payout Contract Structure*		
Phase I	70	10	1000		
Phase II	80	10	1000	217.8	0.119
Phase III	375	450	1000		
			Panel B *Medium-Payout Contract Structure*		
Phase I	50	5	1000		
Phase II	60	5	1000	128.2	0.071
Phase III	560	670	1000		
			Panel C *Low-Payout Contract Structure*		
Phase I	25	0	1000		
Phase II	25	0	1000	79.7	0.036
Phase III	500	580	1000		

Figure 1: Zeta Profile Plots for Multilevel Model



Note: Xlnme = log mech; Xlnpe = log pest; XCountry = country; Xlnl = log labor; Xlnfe = log fert; Xlni = log irrig, σ_1 = parcel-level disturbance term; σ_2 = household-level disturbance term; σ_3 = time-level disturbance term; σ = idiosyncratic error term.

Figure 2: Histograms of Level Variance from Bayesian Estimation



Note: Histograms drawn from posterior distributions of variance terms estimated from model (7b) reported in Table 4.