The Impact of Climate Change on Rice Yields: Heterogeneity and Uncertainty

Kamilya Tazhibayeva
Robert Townsend

December 2012

Abstract

We specify a three-stage production function for rice cultivation which incorporates the sequential nature of both production shocks, including weather, and input choices based on sequentially updated information sets of history of realized shocks and observed changes in crop growth. The production function is CES across stages, thus taking into account substantial complementarities between different phases of the biophysical crop growth process, in contrast to the substitute nature of commonly used Cobb-Douglas specification. This framework is particularly well-suited for evaluating the effect of climate change on crop cultivation practices and yields. We apply the model to 11-year panel of rice farmers in Thailand. The panel structure of our data allows us to analyze both the cross sectional effects of weather shocks and climate change on yields, and the effects on the mean levels and shape of yield distribution for individual farmers. We find substantial heterogeneity among farmers in the effect of both weather shocks and climate change on yields. We consider two alternative climate change scenarios for Southeast Asia, one with mild increases of temperature and rainfall throughout the year and the other with more extreme temperature increases and less rainfall during months of rice cultivation. While from the farmer's perspective uncertainty of yields decreases with more extreme climate change, cross sectional heterogeneity in uncertainty increases. Our focus is on detailed partial equilibrium analysis of the effects of climate change on yields at the crop-plot level, accurate understanding of which is essential for global general equilibrium modeling of environmental changes. We integrate our economic model of rice production with soil science crop growth modeling, weather simulators, and global climate change models.

Kamilya Tazhibayeva: kamilya@mit.edu
Robert Townsend: rtownsen@mit.edu
We thank Andrew Foster, Michael Greenstone, Samuel Kortum, and the participants of the IIES conference on Climate and the Economy (2012) and Seminar in Environmental Economics and Policy at Harvard for helpful comments and suggestions; John Felkner for collaboration with DSSAT simulations and generation of climate scenarios; Jim Jones, Cheryl Porter, Gerrit Hoogenboom, and Paul Wilkens for comments and help with the DSSAT and climate modeling; and Neil Best for technical assistance. We acknowledge support from National Institute of Child Health and Human Development, the National Science Foundation, and the John Templeton Foundation.

1 Introduction

We quantify the impact of weather variation and climate change on the production of rain-fed rice. Rice is a major crop of Thailand both for domestic consumption and export[1], as it is for other countries in Southeast Asia. Crop failure and commodity shortages can lead to disaster, even national calamities. However our story turns out a bit differently. It is a tale of substantial heterogeneity in impact of weather variation for the given contemporary climate, as well as for climate shift scenarios, and changes in uncertainty that farmers face.

We analyze two alternative climate change scenarios, one mild and another more severe. Then, taking advantage of detailed panel data, we are able to distinguish two dimensions of the effect of climate change on rice yields. One is the extent of heterogeneity in yield distributions across households, both in means and in variation, or uncertainty, that is present under a given climate, and how this cross sectional heterogeneity in means and variation is affected by climate change. The other is how climate change affects yield uncertainty from the household's perspective. We find that extreme climate change increases heterogeneity in levels and uncertainty of yields across households but decreases uncertainty about yields for a given household.

Our detailed agricultural data are gathered in the field in the Northeast of Thailand, in a semi-arid tropics zone, as part of Townsend Thai monthly survey, and form a panel for 1999-2009 covering 11 crop cycles. The data we use cover 268 households and 2,887 crop plots in two provinces, Sisaket and Burirum. The data contain the details on inputs use including fertilizer, labor, and land planted; harvests and hence yields; crop operations by name, which allows a multistage approach to modeling crop production; and a suite of environmental data including initial plot soil measurements, monthly rainfall within village and temperature at nearby meteorological station.

We integrate soil science crop modeling, weather simulators, and global climate change models with this multistage crop production model. Farmers are imagined to maximize expected profits, as if they were risk neutral[2]. Yields in the final stage are a function of rainfall, temperature and inputs during that stage. The initial condition for the final stage is the condition of crop-plot[3] at the beginning of the final stage. In turn that plot condition is a function of production activities in the previous stage, both inputs use and realizations of shocks. We use a combination of a biophysical

[1]Thailand is the world's largest rice exporter, and rice is one of Thailand's top ten exports. Thailand's share of world's rice export averaged 30 percent for 1980-2006 (FAOSTAT, http://faostat.fao.org, exports measured in tons).

[2]Household production separates from consumption and labor supply decisions when markets are complete. There is some evidence for this in the Townsend Thai Project monthly data. For details, see Alem and Townsend (2007). Levels of consumption smoothing by households in these data provide evidence of extensive social networks that enable consumption smoothing and thus approximate Arrow-Debreu institutions.

[3]We refer as "crop-plot" to plot- and year-specific observations, or crops cultivated on a given physical plot in a given year.

crop production model, called DSSAT, and previous labor effort to measure as best we can the condition of the plants, i.e. due to human and physical interaction. Households are forward looking and make expectations of weather and prices into the future based on current information. Finally, the timing of planting is incorporated through its effect on the timing of stages and therefore on the weather realization for a given plot. While weather is mostly uniform within a village, or even across nearby villages, its impact is heterogeneous as farmers plant at different time.

We model crop cultivation as having a general constant elasticity of substitution (CES) specification between production stages, rather than assuming a special case such as unitary or linear elasticity. The biophysical nature of crop development suggests limited substitutability between early and later stages of plant growth. For example, lack of rainfall when seedlings are planted in the beginning of crop cultivation will result in wilted seedlings, and abundant rainfall in subsequent months will not compensate for it. Similarly, one cannot substitute labor during planting stage with labor during harvesting stage; indeed, these two labor inputs are more likely to be complements. Our general CES approach does not impose any *a priori* restrictions on the level of substitutability or complementarity of production processes in different stages. Instead, we allow the data to determine these inter-stage elasticities.

This more realistic approach has important implications for the analysis of the effects climate change has on yields. Limited substitutability of production activities across stages means lack of flexibility in adjusting timing of inputs use once production has started. For example, once seedlings are transplanted, labor has to be spent on taking care of the crop, irrespective of anticipated future positive or negative weather shocks. Thus input choices are driven more by the demands of the current state of the crop, prior to realization of all weather shocks, and less by expectations of future weather shocks. This lack of flexibility in input timing once production is in process induces variability in yields for several reasons. First, it makes choice of planting date much more significant, as that is the only timing decision where the farmer has certain flexibility. As choice of planting time is made before weather shocks are realized, it is based on farmer's weather expectations, and introduces variability in yields due to imperfect foresight (and also potentially due to heterogeneity in expectation formation). Heterogeneity in planting date, in turn, introduces heterogeneity in stage-specific weather shocks despite the aggregate nature of weather, since crops on plots that were planted on different dates will be at different stages of growth at the same calendar time, and so the same aggregate weather realization will affect them differently.

In contrast, Cobb-Douglas specification, with unitary elasticity of substitution across production stages, allows farmers substantial flexibility in timing of inputs application even after planting has started. Consequently, under Cobb-Douglas specification input choices are driven by expectations of future weather shocks and current weather realizations, with current and future inputs being substituted for one another depending on the relative effect of current versus expected impact of weather realizations. That is, under Cobb-Douglas the important factor in farmer's decisions is farmer's weather expectations once planting has started, while under general CES specification that important factor is farmer's weather expectations before planting has started and his choice of planting date based is on these pre-planting expectations. As a result, Cobb-Douglas specification does not capture heterogeneity in planting time, and consequently it does not capture significant sources of heterogeneity in yields.[4]

We model production shocks, which are mainly weather realizations, as stage-specific and correlated across stages, employing a general specification that does not restrict the magnitudes of these correlations in any way. Our baseline model displays some key features. The data in fact display zero substitutability between soil quality and planting activities, and between growing stage and harvesting. There is more substitutability between planting stage and intermediate growing stage. This result underlines the importance of structural CES approach to modeling crop cultivation (and again the drawbacks of using linear or Cobb-Douglas specification for studying the effects of climate change on yields). Another implication of this result is that absence of substitutability between soil quality and planting activities underlines the importance of variation in soil quality in determining yields. In other words, initial, innate soil quality is a significant source of yield heterogeneity. Our estimates also indicate that the effect of weather, and rainfall in particular, is most pronounced during intermediate growth stage. We find that both DSSAT and previous stage labor are significant measures of intermediate outputs. These results both underline the importance of properly accounting for nonlinear interactions of weather and soil with crop development, as biophysical models like DSSAT do, and at the same time demonstrate that using such simulation models without accounting for human input is not sufficient. To our knowledge, we are among the first to integrate biophysical soil models with structural production function.[5]

We study though the lens of the model as it stands the impact of variation in weather, for the given contemporary climate. We also simulate the impact of weather generated for two alternative climates, the low and high emissions IPCC SRES climate change scenarios. For this study, we have chosen to use climate change predictions produced for the 4th Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC), released in 2007 (Cruz, Harasawa,

[4]We thank Andrew Foster for pointing out this advantage of CES specification.

[5]Existing literature contains a number of works that emphasize the potential of cross-disciplinary approach for the analysis of crop cultivation and present empirical frameworks that integrate biophysical and econometric models. Examples are Antle and Capalbo (2001), Stoorvogel, Antle, Crissman, and Bowen (2004), Antle and Stoorvogel (2006) and Valdivia, Antle, and Stoorvogel (2012). However, the economic behavior of farmers in these papers is represented by generalized reduced-form empirical specifications of input demands, rather than a structural model. These papers vary in the extent of integration of econometric and biophysical models.

Lal, Wu, Anokhin, Punsalmaa, Honda, Jafari, Li, and Ninh, 2007). We use an "ensemble-mean"[6] output of multiple, internationally reputable coupled Atmospheric-Oceanic General Circulation Models (AOGCMs) to produce predicted changes for the Southeast Asia region for the time period 2040-2069, relative to the 1960-1990 baseline period[7]. AOGCMs are computationaly intensive numerical models driven by equations for atmospheric and oceanic processes, which are integrated forward sequentially (e.g., temperature, moisture, surface pressure).

Because of the uncertainty in future anthropogenic global emissions (which may differ dramatically due to economic development, policy decisions or technology changes), as well as to assess the range of likely possible climate changes and impacts, we simulated two alternative economic scenarios selected from a set of widely-used scenarios developed for the IPCC Third Assessment Report: the Special Report on Emissions (SRES), the highest emissions trajectory scenario A1F1 and the lowest emissions trajectory scenario B1 (Nakicenovic, Alcamo, Davis, de Vries, Fenhann, Gaffin, Gregory, Grubler, Jung, and Kram, 2000)[8], both for the 2040-2069 time period. We did not specifically model El Niño impacts, as our primary focus was on impacts and adaptations to longer-term "baseline" changes.

IPCC ensemble-mean results predict a net increase in yearly average temperature of between 1.32°C (lowest emissions scenario B1) and 2.01°C (highest emissions scenario A1F1) and an increase in annual precipitation of 2.25 percent (lowest emissions) and 1.00 percent (highest emissions) for the 2040-2069 period, relative to the baseline 1961-1990 period (Cruz, Harasawa, Lal, Wu, Anokhin, Punsalmaa, Honda, Jafari, Li, and Ninh, 2007).

Assessing the impact of these changes on future agricultural outputs and crop yields is complex, as yields are a result of interactions between temperature, precipitation effects, direct physiological effects of increased CO2, and effectiveness and availability of adaptations (Parry, Rosenzweig, Iglesias, Livermore, and Fischer, 2004). Consequently, predictions for Asia are mixed. Some studies find decreases in rain-fed crops in South and South-East Asia (Rosenzweig, Iglesias, Yang, Epstein, and Chivian, 2001). Others such as Cruz, Harasawa, Lal, Wu, Anokhin, Punsalmaa, Honda, Jafari, Li, and Ninh (2007), using the HadCM2 global climate model, indicate that crop yields could likely increase up to 20 percent in East and South-East Asia, while Parry, Rosenzweig, Iglesias, Livermore, and Fischer (2004) find both increases and decreases in yields for Thailand

[6]"Ensemble-mean" predictions are the mean output from multiple models, run together to avoid potential bias or flaws inherent in any particular climate change model, providing a superior delineation of the forced climate change signal from the natural background variability of the system (Giorgi and Mearns, 2002).

[7]The models are listed on the IPCC website.

[8]The SRES scenarios, as with all economic scenarios of emissions and their reliability, are a source of some controversy. For example, the SRES scenarios have been criticized for their use of Market Exchange Rates (MER) for international comparison, in lieu of PPP exchanges rates, which correct for differences in purchasing power. However, for this micro-study, we accept these scenarios as given.

depending on CO2 regimes.

We find that from a household's perspective, mean yields decrease with more severe climate change. While this decrease is statistically significant, its magnitude is low. There is no significant change in mean household yields for a milder climate change. Yes at the same time, heterogeneity in mean yields across households, in the cross section, increases for both alternative climate change scenarios. From the perspective of the farmer, uncertainty in the distribution of yields decreases in the more extreme climate.

This paper is organized into seven sections. Section 2 outlines the multistage model of rice cultivation. Section 3 discusses the structure of production shocks and the error term. Section 4 describes the data. Section 5 presents estimates of the multistage model. Section 6 outlines the integration of economic, crop growth, weather and climate models and presents our analysis of climate change impact on yields. Section 7 concludes the paper.

2 Modeling Rice Cultivation

Economic analysis of production traditionally assumes that production process occurs in one stage. All input choices are made at the start of production. Within the single production stage, all inputs are utilized simultaneously and timing of input usage does not affect realized output. Inputs are defined solely on the basis of their physical characteristics.

The single stage approach is ill-suited for analysis of agricultural crop production (Antle, 1983; Antle and Hatchett, 1986)[9]. Crop production is defined by the process of a crop's biological growth, which consists of distinct, chronologically sequential phases. Crop's need for and responsiveness to a given physical input varies across different growth phases. This makes the timing characteristic of inputs important in analysis of crop cultivation. Depending on the progress of crop growth, the farmer may want to adjust his use of inputs. As a result, input decisions are sequential in nature and are not all made at the start of production. The farmer responds to realized production shocks as captured in the state of the crop-plot, while forecasting future shocks and actions. The farmer can also use realized production shocks to update his information set and therefore his expectations of production shocks for upcoming stages. This can introduce a bias in estimation when production shocks influencing input choices are not seen by the econometrician and end up in the yield error term.

With crop cultivation, each sequential stage can be thought of as a separate production subprocess with its own production function. We map the growth phases of biological development of

[9]See also Just and Pope (2001) and Just and Pope (1978) for rigorous discussion of agricultural production functions.

the rice plant into economic production stages by matching the timing of production operations to the timing of plant development. First is the juvenile growth phase, during which germination takes place. It corresponds in the production process to planting of seeds and growing and transplanting of seedlings. The second is the intermediate phase, during which panicle initiation and heading occur. It corresponds to crop maintenance stage, which includes such operations as weeding and fertilizing. Third is the final phase, during which grains fill and mature. It corresponds to harvest collection and storage.

Using this mapping, we construct a three-stage rice production function. Within each stage, several operations can be performed simultaneously. Output from the previous stage is an initial condition for next stage production subprocess. Input decisions are made at the start of each stage, with updated expectations based on history at that point in time. This approach incorporates the sequential nature of crop production, where production shocks and input decisions from earlier stages affect crop-plot conditions and therefore input decisions at later stages. We assume that crop cultivation process is CES across stages and Cobb-Douglas within stages, with constant returns to scale in both instances. This approach enables us to describe crop cultivation process as a system of equations, with one equation specifying final output as a function of all inputs from all three production stages, and three sets of equations describing input demands for each of the three production stages.[10] We now derive this system of equations.

Let i index the three production stages and let vector $x_i = \left(x_{i1}\ x_{i2}\ ...\ x_{i,N_i}\right)'$ denote N_i inputs for stage i. Let y_i be the realized output of stage i, with y_0 describing initial conditions of production such as plot characteristics. Let ε_i be production shock realized during stage i, with ε_0 describing weather in pre-planting months. Then output in stage i is $f_i\left(y_{i-1}, x_i, \varepsilon_i\right) = y_i \exp\left(\varepsilon_i\right)$, for $i = 1, 2, 3$, where f_i is stage i - specific stochastic production process and y_i is stage i-specific CES production function[11]:

$$f_i\left(y_{i-1}, x_i, \varepsilon_i\right) = A_i \left(\theta_i \left(y_{i-1} \exp\left(\varepsilon_{i-1}\right)\right)^{\gamma_i} + \left(1-\theta_i\right)\left(B_i \prod_{n=1}^{N_i} x_{in}^{\alpha_{in}}\right)^{\gamma_i}\right)^{1/\gamma_i} \exp\left(\varepsilon_i\right), \qquad (1)$$

where $\gamma_i \leq 1$, and elasticity of substitution between production stages i and $i-1$ is equal to $1/\left(1-\gamma_i\right)$.[12] The order of events in each stage i is as follows. Input decisions x_i are made based on the history of production shocks and intermediate outputs realized in previous stages, and

[10]We later estimate the composite production function and input decision rules as a system of equations, with error terms correlating across equations.

[11]Values of inputs, outputs and production shocks are plot-specific. Plot indexing is omitted for simplicity of presentation.

[12]Special cases are Cobb-Douglas specification with $\gamma_i = 0$, linear with $\gamma_i = 1$, and Leontieff with $\gamma_i = -\infty$.

before stage i shocks are realized. Next, production takes place and inputs x_i are used at the same time as production shocks for the current stage, ε_i, are realized. At the end of the stage, output for the current stage, y_i, is observed. Substituting in recursively for intermediate outputs, we obtain a composite production function which describes final harvest as a function of initial plot conditions, and inputs and realized production shocks from all three stages: $f\left(y_0, \{x_i, \varepsilon_i\}_{i=1}^3\right) = y_3 \exp(\varepsilon_3)$, or

$$f\left(y_0, \{x_i, \varepsilon_i\}_{i=1}^3\right) = \tag{2}$$
$$\underbrace{\delta_3\left[\delta_2\left(\delta_1\left[y_0^{\gamma_1} + (\rho_1 z_1)^{\gamma_1}\right]^{\gamma_2/\gamma_1}\exp\left(\gamma_2\varepsilon_1\right) + (\rho_2 z_2)^{\gamma_2}\right)^{\gamma_3/\gamma_2}\exp\left(\gamma_3\varepsilon_2\right) + \left(\rho_3 z_3\right)^{\gamma_3}\right]^{1/\gamma_3}}_{y_3}\exp\left(\varepsilon_3\right),$$

where $\delta_i = A_i\theta_i^{1/\gamma_i}$, $\rho_i = A_i\left(1-\theta_i\right)^{1/\gamma_i}$, and $z_i = B_i\prod_{n=1}^{N_i} x_{in}^{\alpha_{in}}$.

At each stage, the farmer chooses inputs to maximize expected profits. Let p denote the price of final output and $w_i = \left(w_{i1}\ w_{i2}\ ...\ w_{iN_i}\right)$ denote a vector of stage i input prices. At each stage i, expectations are taken conditional on all information available to the farmer at that point, which includes all realized production shocks, intermediate outputs and factor prices from previous stages, and all current stage prices: $I_i = \left\{\left\{y_j\right\}_{j=0}^{i-1}, \left\{\varepsilon_j\right\}_{j=0}^{i-1}, \left\{w_j\right\}_{j=1}^{i}\right\}$. Stage 3 information set also includes final output price p. We assume that output and factor prices are independent from production shocks.

At the beginning of stage 3, the farmer chooses profit-maximizing levels of stage 3 inputs, x_3. At that point, only stage 3 production shock is not yet realized. Therefore, the farmer's information set at the beginning of stage 3, $I_3 = \left\{\left\{y_j\right\}_{j=0}^{2}, \left\{\varepsilon_j\right\}_{j=0}^{2}, \left\{w_j\right\}_{j=1}^{3}, p\right\}$, includes realization of stage 2 output and all production shocks that occurred in earlier stages. The farmer solves

$$\max_{\{x_{3n}\}_{n=1}^{N_3}} p\underbrace{A_3\left(\theta_3\left(y_2\exp\left(\varepsilon_2\right)\right)^{\gamma_3} + \left(1-\theta_3\right)\left(B_3\prod_{n=1}^{N_3} x_{3n}^{\alpha_{3n}}\right)^{\gamma_3}\right)^{1/\gamma_3}}_{y_3} E_3\left[\exp\left(\varepsilon_3\right)\right] - \sum_{n=1}^{N_3} w_{3n}x_{3n},$$

with expectation of stage 3 production shock, $E_3\left[\exp\left(\varepsilon_3\right)\right]$, conditional on information set I_3. The first order conditions are

$$pE_3\left[\exp\left(\varepsilon_3\right)\right]\frac{\partial y_3}{\partial x_{3n}} = w_{3n}\ \forall n \in \{1, ..., N_3\}$$

and have the standard interpretation that at the optimum level, input's marginal product is equal to its real price.

From the first order conditions it follows that $\alpha_{3j}w_{3k}x_{3k} = \alpha_{3k}w_{3j}x_{3j}$ for all $j, k \in \{1, ..., N_3\}$. This lets us express all stage 3 inputs in terms of one stage 3 input, say, x_{31}, as $x_{3k} = \frac{\alpha_{3k}w_{31}}{\alpha_{31}w_{3k}}x_{31}$ for all $k \in \{1, ..., N_3\}$. Solving first for the optimal x_{31}, we can then solve for optimal stage 3 inputs levels x_{3k} $\forall k \in \{1, ..., N_3\}$:

$$x_{3k} = \left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3} \frac{\alpha_{3k}y_2 \exp(\varepsilon_2)}{B_3 w_{3k}\lambda_3}\left[(1-\theta_3)^{\frac{1}{\gamma_3 - 1}}\left(A_3 B_3 \lambda_3 p E_3\left[\exp(\varepsilon_3)\right]\right)^{\frac{\gamma_3}{\gamma_3-1}} - 1\right]^{-1/\gamma_3} \tag{3}$$

$\forall k \in \{1, ..., N_3\}$, where $\lambda_3 = \prod_{n=1}^{N_3}\left(\frac{\alpha_{3n}}{w_{3n}}\right)^{\alpha_{3n}}$. Using the approximation $\ln(x-1) \approx \ln x$, we can obtain the log-linear approximation

$$\begin{aligned}\ln x_{3k} \approx \ln \alpha_{3k} + C_3 + \ln y_2 + \varepsilon_2 - \frac{1}{1-\gamma_3}\ln\frac{w_{3k}}{p} + \\ + \frac{\gamma_3}{1-\gamma_3}\sum_{n=1,n\neq k}^{N_3}\alpha_{3n}\ln\frac{w_{3k}}{w_{3n}} + \frac{1}{1-\gamma_3}\ln E_3\left[\exp(\varepsilon_3)\right] \ \forall k \in \{1, ..., N_3\},\end{aligned} \tag{4}$$

where the common component of the constant term is $C_3 = \frac{1}{\gamma_3}\ln\theta_3 + \frac{1}{1-\gamma_3}\ln A_3(1-\theta_3) + \frac{\gamma_3}{1-\gamma_3}\left(\ln B_3 + \sum_{n=1}^{N_3}\alpha_{3n}\ln\alpha_{3n}\right)$. Input demand is increasing in previous intermediate output. Assuming that current inputs and previous intermediate output are complements rather than substitutes, so that $\gamma_3 < 0$, input demand is increasing in expected production shock and in relative prices of other stage 3 inputs, and decreasing in its own real price.

At the beginning of stage 2, the farmer chooses profit-maximizing levels of stage 2 inputs, x_2, given realized stage 1 output and taking into account his anticipated stage 3 inputs demands. At this point, the farmer's information set is $I_2 = \left\{\left\{y_j\right\}_{j=0}^{1}, \left\{\varepsilon_j\right\}_{j=0}^{1}, \left\{w_j\right\}_{j=1}^{2}\right\}$. Farmer solves

$$\max_{\{x_{2n}\}_{n=1}^{N_2}} E_2\left[p\right]E_2\left[y_3\exp(\varepsilon_3)\right] - \sum_{n=1}^{N_2}w_{2n}x_{2n} - E_2\left[\sum_{n=1}^{N_3}w_{3n}x_{3n}\right],$$

where expectations are conditional on information set I_2, y_3 is given by equation (1) and is a function of expected stage 3 inputs demands (3). Substituting equation (3) for stage 3 input demands $\{x_{3n}\}_{n=1}^{N_3}$, we can express stage 3 production costs, $\sum_{n=1}^{N_3}w_{3n}x_{3n}$, and deterministic stage 3 output,

y_3, in terms of deterministic component of stage 2 output, y_2:

$$E_2\left[\sum_{n=1}^{N_3} w_{3n}x_{3n}\right] = \left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3} \frac{y_2}{B_3} E_2\left[\frac{\exp\left(\varepsilon_2\right)}{\lambda_3 g_3^{1/\gamma_3}}\right]$$

$$\text{and } E_2\left[y_3\right] = A_3\theta_3^{1/\gamma_3} y_2 E_2\left[\exp\left(\varepsilon_2\right)\left(1+\frac{1}{g_3}\right)^{1/\gamma_3}\right],$$

where $g_3 = (1-\theta_3)^{\frac{1}{\gamma_3-1}}\left(A_3 B_3 \lambda_3 p E_3\left[\exp\left(\varepsilon_3\right)\right]\right)^{\frac{\gamma_3}{\gamma_3-1}} - 1$. The first order conditions for $x_{2k}, k = 1, ..., N_2$, are

$$A_3\theta_3^{1/\gamma_3}\frac{\partial y_2}{\partial x_{2k}} E_2\left[\left(\frac{1+g_3}{g_3}\right)^{1/\gamma_3}\exp\left(\varepsilon_2+\varepsilon_3\right)\right] = \frac{w_{2k}}{E_2\left[p\right]} + \frac{1}{B_3}\left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3}\frac{\partial y_2}{\partial x_{2k}} E_2\left[\frac{\exp\left(\varepsilon_2\right)}{\lambda_3 p g_3}\right].$$

That is, marginal cost of an intermediate stage 2 input, equal to the right hand side of the above first order condition, consists of two components, concurrent and anticipated future marginal cost. Concurrent marginal cost is the input's real price, $w_{2k}/E_2\left[p\right]$. Anticipated future marginal cost of an intermediate stage 2 input is its marginal effect on expected production costs of the future production stage 3. Levels of stage 2 inputs affect optimal usage of future production stage 3 inputs, and therefore expected stage 3 production costs, through their effect on the level of intermediate output y_2 which is the initial condition for stage 3 production. At the optimal level of stage 2 input demand, this composite marginal cost is equal to that input's marginal product, which is the left hand side of the above first order condition. Note how both stage 2 input's marginal product and future marginal cost depend on expected production shock not only for the current stage, but also for the subsequent stage 3.

Solving again first for the optimal x_{21}, we can then solve for optimal stage 2 input levels:

$$x_{2k} = \left(\frac{\theta_2}{1-\theta_2}\right)^{1/\gamma_2}\frac{\alpha_{2k} y_1 \exp\left(\varepsilon_1\right)}{B_2 w_{2k}\lambda_2}\left[(1-\theta_2)^{\frac{1}{\gamma_2-1}}\left(A_2 B_2 \lambda_2 E_2\left[P_3\exp\left(\varepsilon_2\right)\right]\right)^{\frac{\gamma_2}{\gamma_2-1}} - 1\right]^{-1/\gamma_2} \tag{5}$$

$\forall k \in \{1, ..., N_2\}$, where $\lambda_{2n} = \prod_{n=1}^{N_2}\left(\frac{\alpha_{2n}}{w_{2n}}\right)^{\alpha_{2n}}$, $P_3\left(\varepsilon_3, \{w_{3n}\}_{n=1}^{N_3}, p\right) = \theta_3^{1/\gamma_3} p R_3^{\frac{\gamma_3-1}{\gamma_3}}$, and

$$R_3\left(\varepsilon_3, \{w_{3n}\}_{n=1}^{N_3}, p\right) = \left(A_3\exp\left(\varepsilon_3\right)\right)^{\frac{\gamma_3}{\gamma_3-1}} - \left((1-\theta_3)\left(B_3\lambda_3 p\right)^{\gamma_3}\right)^{\frac{1}{1-\gamma_3}}.$$

Component R_3 captures the net indirect effect of a change in stage 2 input on stage 3 production process. This indirect effect comes from the direct positive effect of stage 2 input use on stage 2 output. On the one hand, higher stage 2 output, which is used as an input in stage 3, results in higher stage 3 output, other things being equal. On the other hand, recall from equation (3) that demands for stage 3 inputs increase in stage 2 output. As a result, stage 3 output increases, but so do stage 3 production costs. From equation (5), the marginal effect of expected stage 3 production shock on stage 2 input demands, $\partial x_{2k}/\partial E_2[\varepsilon_3]$, is positive, while the marginal effect of expected stage 3 real input prices on stage 2 input demands, $\partial x_{2k}/\partial E_2\left[w_{3k}/p\right]$, is negative. As in stage 3, stage 2 input demand is increasing in previous intermediate outputs, in expected production shocks, and in relative prices of other stage 2 inputs, and decreasing in its own real price.

Rewrite R_3 as

$$R_3\left(\varepsilon_3, \{w_{3n}\}_{n=1}^{N_3}, p\right) = \frac{(1-\theta_3)^{\frac{1}{\gamma_3-1}} (A_3 B_3 \lambda_3 p \exp(\varepsilon_3))^{\frac{\gamma_3}{\gamma_3-1}} - 1}{(1-\theta_3)^{\frac{1}{\gamma_3-1}} (B_3 \lambda_3 p)^{\frac{\gamma_3}{\gamma_3-1}}}. \tag{6}$$

For $\gamma_3 < 0$, it appears that the effect of expected increase in stage 3 output due to higher stage 2 input use is very pronounced, while the effect of higher stage 3 production costs is negligible. Note that $\partial R_3/\partial \gamma_3 < 0$, so the stronger is complementarity between production stages 2 and 3 (that is, the more negative γ_3 is), the more prevalent is the former effect over the latter. We approximate R_3 as $R_3 \approx (A_3 \exp(\varepsilon_3))^{\frac{\gamma_3}{\gamma_3-1}}$.[13] Applying the log approximation to equation (5), we can write stage 2 input demands as

$$\begin{aligned} \ln x_{2k} \approx {} & \ln \alpha_{2k} + C_2 + \ln y_1 + \varepsilon_1 - \frac{1}{1-\gamma_2} \ln \frac{w_{2k}}{E_2[p]} + \\ & + \frac{\gamma_2}{1-\gamma_2} \sum_{n=1, n\neq k}^{N_2} \alpha_{2n} \ln \frac{w_{2k}}{w_{2n}} + \frac{1}{1-\gamma_2} \ln E_2[\varepsilon_2 + \varepsilon_3] \end{aligned} \tag{7}$$

$\forall k \in \{1, ..., N_2\}$, where the common component of the constant term is $C_2 = \frac{1}{\gamma_3(1-\gamma_2)} \ln \theta_3 + \frac{1}{\gamma_2} \ln \theta_2 + \frac{1}{1-\gamma_2} \ln A_3 A_2 (1-\theta_2) + \frac{\gamma_2}{1-\gamma_2}\left(\ln B_2 + \sum_{n=1}^{N_2} \alpha_{2n} \ln \alpha_{2n}\right)$.

We solve the farmer's stage 1 maximization problem in a similar manner, and obtain the fol-

[13] This approximation eliminates dependence of stage 2 input demands on expected stage 3 factor prices.

lowing log approximation[14] of stage 1 input demands:

$$\ln x_{1k} \approx \ln \alpha_{1k} + C_1 + \ln y_0 - \frac{1}{1-\gamma_1} \ln \frac{w_{1k}}{E_1[p]} + \tag{8}$$

$$+ \frac{\gamma_1}{1-\gamma_1} \sum_{n=1,n\neq k}^{N_1} \alpha_{1n} \ln \frac{w_{1k}}{w_{1n}} + \frac{1}{1-\gamma_1} \ln E_1[\varepsilon_1 + \varepsilon_2 + \varepsilon_3] \ \forall k \in \{1, ..., N_1\},$$

where expectations are conditional on the information set $I_1 = \{y_0, \varepsilon_0, w_1\}$ and the common component of the constant term is $C_1 = \frac{1}{\gamma_3(1-\gamma_1)} \ln \theta_3 + \frac{1}{\gamma_2(1-\gamma_1)} \ln \theta_2 + \frac{1}{\gamma_1} \ln \theta_1 + \frac{1}{1-\gamma_1} \ln A_3 A_2 A_1 (1-\theta_1) + \frac{\gamma_1}{1-\gamma_1} \left(\ln B_1 + \sum_{n=1}^{N_1} \alpha_{1n} \ln \alpha_{1n} \right)$.

Rice cultivation process is described by a system of equations consisting of cumulative production function equation for final output (2) and equations approximating input demand for each of the three stages: (8), (7), and (4).

2.1 Comparison of CES and Cobb-Douglas Specifications

For comparison, let's consider Cobb-Douglas specification of stage production functions, so that production activities in different stages are substitutes with unit elasticity of substitution, rather than potentially complements as allowed for with our CES specification. Let stage production functions be

$$f_i(y_{i-1}, x_i, \varepsilon_i) = A_i y_{i-1}^{\beta_i} \prod_{n=1}^{N_i} x_{in}^{\alpha_{in}} \exp(\varepsilon_i).$$

Constant returns to scale imply that $\beta_i + \sum_{n=1}^{N_i} \alpha_{in} = 1$. Solving farmer's profit maximization problem at the beginning of stage 3, we obtain the following stage 3 input demands:

$$\ln x_{3k} = \ln \alpha_{3k} + C_3 + \ln y_2 + \varepsilon_2 - \frac{1}{\beta_3} \ln \frac{w_{3k}}{p} + \frac{1}{\beta_3} \sum_{n=1,n\neq k}^{N_3} \alpha_{3n} \ln \frac{w_{3k}}{w_{3n}} + \frac{1}{\beta_3} E_3[\varepsilon_3]$$

$\forall k \in \{1, ..., N_3\}$, where $C_3 = \frac{1}{\beta_3} \left(\ln A_3 + \sum_{n=1}^{N_3} \alpha_{3n} \ln \alpha_{3n} \right)$. Comparing this equation with log approximation of stage 3 input demands under CES specification in equation (4), there are two differences. The first difference is the magnitude of the coefficient on expected stage 3 production shock, $E_3[\varepsilon_3]$. The coefficient is positive in both cases; however, it is greater than one under Cobb-

[14] This approximation eliminates dependence of stage 1 input demands on expected stage 2 and stage 3 factor prices.

Douglas and less than one under CES when $\gamma_3 < 0$, that is, when stage 2 intermediate output and stage 3 production activities are complements.[15] The second difference is that relative prices of other stage 3 inputs have negative effect on input demand under Cobb-Douglas and positive effect under CES - again, assuming $\gamma_3 < 0$. If $\gamma_3 > 0$, so that intermediate stage 2 output and stage 3 production activities are substitutes, these qualitative differences with Cobb-Douglas specification go away. The same qualitative results hold for input demands in stages 1 and 2.

2.2 Timing Decision

Once planting commences, the timing of the second and third production stages, corresponding to intermediate growth and harvesting, is dictated by physiology of crop's development. This is confirmed by our data, which display consistency in duration of all three production stages across years[16]. Thus choice of timing in crop cultivation, and resulting variation of timing of production stages across households, are driven mainly by household's choice of when to plant, or, in our model, of the beginning of stage 1. The choice of planting date is, in turn, driven by observed and expected weather conditions. For successful seed germination and transplanting of seedlings, the soil has to contain a certain minimum amount of moisture available to the plant (Hasegawa, Sawano, Goto, Konghakote, Polthanee, Ishigooka, Kuwagata, Toritani, and Furuya, 2008). An optimum planting date can then be thought of as satisfying the following two conditions. First, the farmer must believe that current level of soil moisture is sufficient for planting. Second, the farmer must believe that rainfall and weather conditions in general in the upcoming month (the usual duration of stage 1 is one month) are sufficient for the newly planted plants to survive. In other words, the farmer starts planting when already realized level of soil moisture is sufficiently high and when his expectations of future weather realizations, conditional on observed history, are optimal for crop cultivation.

Formally, let $i = 0$ denote pre-planting stage, during which the farmer chooses a planting date, and let ε_0 denote the weather realization during this pre-planting stage. Farmer chooses ε_0 that maximizes his expected profits:

$$\max_{\varepsilon_0} E_0 \left[p y_3 \exp(\varepsilon_3) - \sum_{n=1}^{N_1} w_{1n} x_{1n} - \sum_{n=1}^{N_2} w_{2n} x_{2n} - \sum_{n=1}^{N_3} w_{3n} x_{3n} \right], \tag{9}$$

where expectation is conditional on the information set $I_0 = \{y_0\}$. By controlling the overall level

[15] A more explicit demonstration of the distinction between CES and Cobb-Douglas specifications that was raised in section 1 requires derivation of input demands under Cobb-Douglas specification for intermediate stages in addition to the final stage, which we do not include here for brevity of presentation.

[16] More details are provided in section 4.

of weather shock realized before planting, or ε_0, the farmer effectively chooses the planting date. For example, if we think of ε_0 as rainfall, by choosing a smaller or larger amount of rainfall to be realized before planting starts, the farmer effectively decides whether planting should take place earlier or later in a given year, conditional on (perceived) rainfall realizations. Note that input amounts for stages 1 through 3, as well as final yield, are functions of future production shocks $\varepsilon_1, \varepsilon_2$ and ε_3. Thus optimal planting timing defined by equation (9) maximizes expected profits conditional on realized and expected weather outcomes.

3 Production Shocks and Error Term

We have three directions of data variation: individual across households, spatial across plots and villages, and temporal across stages and years. Let k index plots, h index households, υ index villages, i index stages, and t index years. For each province, we have data on four villages over 11 years, with around 33 households per village, and around 2 plots cultivated on average by a given household per year. For each plot k in year t, we have three sets of production shocks and error terms, corresponding to three production stages.

Because we have two temporal dimensions, production shocks can potentially be autocorrelated across stages (over i) and across years (over t). Similarly, there are three levels of potential group error correlation, within a physical plot, within a household, and within a village. Let $\varepsilon_{kh\upsilon it}$ denote the overall production shock for plot k belonging to household h in village υ during production stage i in year t, and let $\xi_{kh\upsilon it}$ denote the total unobserved by econometrician error term, similarly defined. We now decompose the overall production shock and error term into observed and unobserved components.

Weather Shock

One of the main production shocks for rice cultivation is weather, and particularly rainfall and temperature. At a given point in time, weather is an aggregate shock at village level and is arguably spatially correlated across villages. In our data, the four sample villages in the same province are located very close to each other. Figures 1 and 2 illustrate this for Sisaket province. A similar picture holds for Burirum province. Correlations of monthly rainfall between villages[17] in our data range from 0.95 to 0.98. In addition, plots belonging to sample households from different villages are often adjacent to one another, and overall plots from all four villages in the same province are

[17] Our data contain daily village-level rainfall starting from 1998. More detailed description of the data is given in section 4.

spatially intermingled. This enables us to assume perfect spatial correlation of monthly weather across all sample plots in the same province as a good approximation. Although monthly weather is an aggregate shock, there is substantial variation among farmers in timing of production activities in a given year. This results in noticeable variation in weather, and in particular rainfall, between plots in a given stage, making stage weather plot-specific rather than aggregate. Let ρ_{khvit} denote weather shock realized on plot k belonging to household h in village v in stage i in year t.

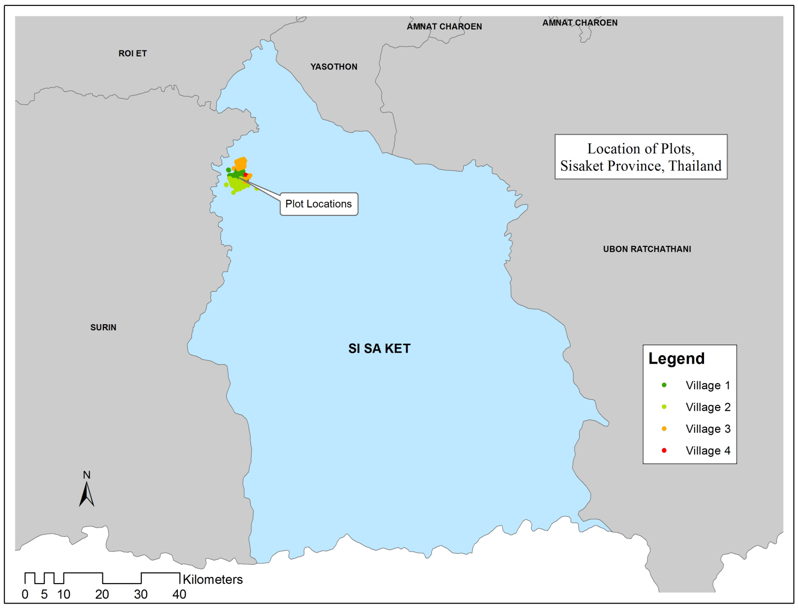

Figure 1: Location of Plots in Four Sample Villages in Sisaket Province

In terms of serial correlation, generally rainfall does not persist from year to year (Paxson, 1992). Weather is more likely to be serially correlated across stages. That is, covariance of ρ_{khvit} and $\rho_{khvi't}$ is generally different from zero for stages $i, i' \in \{1, 2, 3\}$, while covariance of ρ_{khvit} and $\rho_{khvit'}$ is zero for all years $t \neq t'$.

Farmers are able to predict, with varying success, the upcoming weather for future stages. Let $\bar{\rho}_{khvit}$ denote farmer h's weather expectation. The difference between realized and expected weather, $\rho_{khvit} - \bar{\rho}_{khvit}$, is the unanticipated weather shock. Let $\tilde{\rho}_{khvit}$ denote this difference. By construction, $\tilde{\rho}_{khvit}$ has zero mean and is uncorrelated with farmer's weather expectation, $\bar{\rho}_{khvit}$.

At any given point in time, the effect of rain on plant development would vary depending on the plot's soil, elevation and slope. We have a reasonable measure of soil quality, but not of elevation and slope. If elevation and slope vary substantially across plots, this would be a permanent plot-specific effect. Let u_{khv} denote this unmeasured effect for plot k belonging to household h in

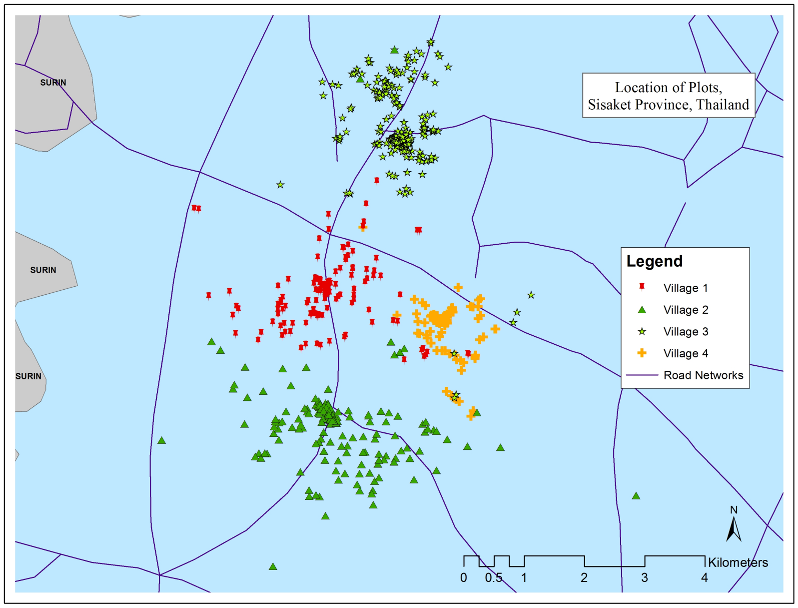

Figure 2: Zoom in on Plot Locations in Four Sample Villages in Sisaket Province

village υ, and $r_{kh\upsilon it}$ denote our measure of plot- and stage-specific weather shock during stage i in year t. Weather shock can be written as a sum of stage- and plot-specific observed shock and fixed plot-specific unobserved effect, $\rho_{kh\upsilon it} = r_{kh\upsilon it} + u_{kh\upsilon}$. As farmers know the characteristics of their plots, including plot's slope and elevation, $u_{kh\upsilon}$ is incorporated into farmer's weather expectation, $\bar{\rho}_{kh\upsilon it}$. We, as econometricians, on the other hand, do not observe plot's slope and elevation, and as a result our measure of a farmer's weather expectation, $\bar{r}_{kh\upsilon it}$, does not incorporate $u_{kh\upsilon}$. That is, while for a farmer realized weather shock can be decomposed as $\rho_{kh\upsilon it} = \bar{\rho}_{kh\upsilon it} + \tilde{\rho}_{kh\upsilon it}$, we can decompose it only as $\rho_{kh\upsilon it} = \bar{r}_{kh\upsilon it} + \tilde{\rho}_{kh\upsilon it} + u_{kh\upsilon}$, where $\tilde{\rho}_{kh\upsilon it}$ is unanticipated by farmer and observed by us weather shock[18] and $u_{kh\upsilon}$ is unobserved by us plot-specific fixed effect. Similarly, our measure of farmer's weather expectation is accurate only up to $u_{kh\upsilon}$: $\bar{\rho}_{kh\upsilon it} = \bar{r}_{kh\upsilon it} + u_{kh\upsilon}$.

[18] $\tilde{\rho}_{ph\upsilon it} = \rho_{ph\upsilon it} - \bar{\rho}_{ph\upsilon it} = r_{ph\upsilon it} + u_{ph\upsilon} - \bar{r}_{ph\upsilon it} - u_{ph\upsilon} = r_{ph\upsilon it} - \bar{r}_{ph\upsilon it}$, and therefore we have an accurate measure of $\tilde{\rho}_{ph\upsilon it}$. This result comes from assumption that actual weather effect, $\rho_{ph\upsilon it}$, is additive in our measure of weather, $r_{ph\upsilon it}$, and the unobserved plot fixed effect, $u_{ph\upsilon}$.

Other Production Shocks and Measurement Errors

Household-level shocks are unlikely to be correlated across years[19,20], but are likely to persist from stage to stage in a given year. Let η_{hvit} denote these household- and stage-specific shocks, then $E\left[\eta_{hvit}, \eta_{hvi't}\right] \neq 0$ for stages $i, i' \in \{1, 2, 3\}$ and $E\left[\eta_{hvit}, \eta_{hvit'}\right] = 0$ for any years $t \neq t'$. Consequently, $E\left[\eta_{hvit}|I_1\right] = 0$ and $E\left[\eta_{hvi't}|I_i\right] \neq 0$ for $i = 2, 3$ and $i' > i$.

Two sources of measurement errors specific to our data should be mentioned. The first is related to the interaction of fertilizer application and soil quality. We have measured data on soil variables for a subset of plots, and these measurements were taken in the base year, 1998, and were not repeated. We have location coordinates for all plots in the sample; using these, plots with no soil data were assigned values of soil variables from geographically closest plots with soil data. If fertilizer application is measured accurately but soil quality is not, it will be hard to disentangle positive effect of fertilizer application on crop development for a given soil quality from the fact that poor soils require higher fertilizer use. In the latter case, higher fertilizer use would be an indicator of low soil quality, which has negative effect on crop development. Estimated effect of fertilizer application on yields will be the net of both positive direct effect of fertilizer use and negative effect of low soil quality. When soil quality is higher (lower) than reflected by soil variables, the direct effect of fertilizer will be overestimated (underestimated). This is a fixed plot-level measurement error and can be included together with unobserved slope and elevation into a permanent plot-specific effect u_{khv}.

The second source of measurement error is the structure of the questionnaires, which ask about activities performed since the last interview, not about dates on which they were performed. Because interviews are conducted monthly, all our variables measuring timing of production activities are accurate up to a month. As a result, the difference between the timing of production activity, in particular of planting, and of the timing of weather realization, can be measured only at month level. For example, if two plots were planted in May, one in the first week of May, another in the last week of May, and adequate rain started only in the middle of May, then second plot's timing is superior to that of the first plot, but we don't capture this in our data. This difference between the two plots ends up in the measurement error term and is specific to plot and stage. Let it be denoted by φ_{khvit}. As all plots are subject to this measurement error and it is random in nature, we can think that there is no systematic bias in this error term component.

[19] There are potential exceptions to this, such as cases of permanent disability.

[20] As part of preliminary data analysis, we have looked at each household's placement, by percentile, in the cross sectional yields distribution for each year, by province. The idea was to check whether some households consistently have higher yields than other households. The data do not display any patterns of household level persistency in yields between years.

Composite Production Shocks and Error Terms

Composite production shock realized on plot k during stage i is the combination of weather and household shocks, $\varepsilon_{khvit} = \rho_{khvit} + \eta_{hvit}$. Farmer's expectation of this shock is $\bar{\varepsilon}_{khvit} = \bar{\rho}_{khvit} + \bar{\eta}_{hvit}$. Both the realized shock and farmer's expectation of it are serially correlated across stages and within household and are serially uncorrelated across years.

Combining together all measurement errors and shocks that are unobserved by us, we can write the composite error term as $\xi_{khvit} = u_{khv} + \eta_{hvit} + \varphi_{khvit}$. This composite error term is autocorrelated within a plot across stages and years, and within a household across stages within a given year. Clustering of error terms at household level allows for unrestricted correlation of error terms corresponding to the same household, including correlation of error terms corresponding to the same plot. We have 141 distinct households in the Sisaket province sample, with an average of 18 observations per household. The number of households in our sample is large enough, and the number of observations per household is small relative to the number of households, to make clustering at household level a viable option in practice. It should be noted that clustering does not correct for potential bias in cases where household-specific shocks η_{hvit} or plot fixed effects u_{khv} are correlated with explanatory variables.

We model stage production shocks ε_{khvit} as joint normal variables. Let ζ_{khvit} be independent standard normal variables. Define stage 0, or pre-planting, shock for plot k belonging to household h in village v in year t as $\varepsilon_{khv0t} = \sigma_0 \zeta_{khv0t} + \mu_0$, where μ_0 and σ_0 are parameters. Then production shock in pre-planting months is a normal variable with mean μ_0 and variance σ_0^2. Parameters μ_0 and σ_0 are invariant across plots, households, villages and years. We postulate that for each of successive production stages 1 through 3 production shocks are correlated with the shock from the previous stage. That is, for $i = 1, 2, 3$, $\varepsilon_{khvit} = \sigma_i \left[\sigma_{i-1,i} \zeta_{khv,i-1,t} + \left(1 - \sigma_{i-1,i}^2\right)^{1/2} \zeta_{khvit} \right] + \mu_i$. Then each ε_{khvit} is a normal variable with mean μ_i, variance σ_i^2, and $\sigma_{i-1,i}$ is the correlation coefficient between production shocks in adjacent stages i and $i - 1$. This specification is flexible with respect to the level of correlation between production shocks in different stages, allowing for both zero and perfect correlation.

We can now write explicitly the expectation terms in linear approximations of input demands. Substituting in $\dfrac{\varepsilon_{khv2t} - \mu_2}{\sigma_2}$ for ζ_{khv2t}, expectation term $\ln E_3 \left[\exp\left(\varepsilon_{khv3t}\right)\right]$ in stage 3 input demand equation (4) is equal to

$$\ln E_3\left[\exp\left(\sigma_3\left[\sigma_{23}\underbrace{\frac{\varepsilon_{khv2t}-\mu_2}{\sigma_2}}_{\zeta_{khv2t}}+\left(1-\sigma_{23}^2\right)^{1/2}\zeta_{khv3t}\right]+\mu_3\right)\right]=$$

$$=\frac{\sigma_3\sigma_{23}}{\sigma_2}\varepsilon_{khv2t}+\mu_3-\mu_2\frac{\sigma_3\sigma_{23}}{\sigma_2}+\frac{\sigma_3^2\left(1-\sigma_{23}^2\right)}{2},$$

since shock ε_{khv2t} is already realized by the beginning of production stage 3 and is therefore included in information set I_3 on which expectation $E_3\left[\exp\left(\varepsilon_{khv3t}\right)\right]$ is conditional. Similarly, expectation terms in stage 2 input demand equation (7) and stage 1 input demand equation (8) are, respectively, equal to

$$\begin{aligned}\ln E_2\left[\varepsilon_{khv2t}+\varepsilon_{khv3t}\right] &= \frac{\sigma_2\sigma_{12}}{\sigma_1}\varepsilon_{khv1t}+\hat{\mu}_2\\ \text{and } \ln E_1\left[\varepsilon_{khv1t}+\varepsilon_{khv2t}+\varepsilon_{khv3t}\right] &= \frac{\sigma_1\sigma_{01}}{\sigma_1}\varepsilon_{khv0t}+\hat{\mu}_1,\end{aligned}$$

where $\hat{\mu}_2$ is a combination of parameters measuring means μ_i, variances σ_i^2, and correlation coefficients $\sigma_{i-1,i}$ of production shocks in stages 1 through 3, and $\hat{\mu}_1$ is a combination of parameters measuring means μ_i, variances σ_i^2, and correlation coefficients $\sigma_{i-1,i}$ of production shocks in stages 0 through 3. In words, expectations of upcoming production shocks incorporate two types of available information: realized production shocks and knowledge about the distribution from which future shock realizations are drawn. Most recent production shock realization embodies all relevant information that is available in preceding history of production shocks.

4 Data

Our data come from the Townsend Thai Project[21] (Binford, Lee, and Townsend, 2004). We focus on rice farmers in two provinces, Sisaket and Burirum, located in predominantly rural and poor northeastern region of the country. Figure 3 shows location of our sample provinces in Thailand. The northeastern region accounts for 57 percent of the total area under rice cultivation in Thailand and 46 percent of the total rice production (Naklang, 2005). In each province, a tambon[22] with four sample villages was selected at random. Data are collected monthly at a household-plot level,

[21] Detailed description of the project can be found at Thailand Database Research Archive, http://cier.uchicago.edu/.
[22] Thai equivalent of a U.S county.

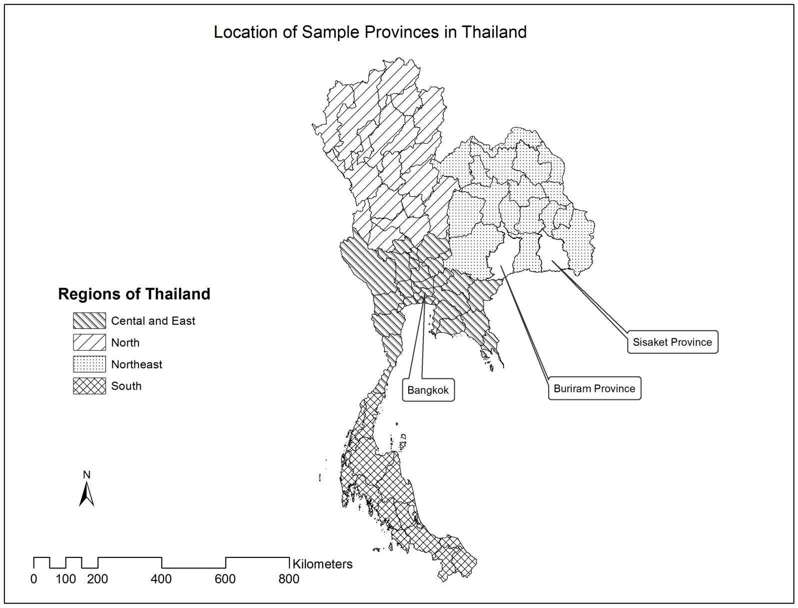

Figure 3: Location of Sample Provinces in Thailand

with many households cultivating several plots in a given year. We use an unbalanced eleven-year panel for 1999-2009. It includes 141 households in Sisaket province, with a total of 1,888 crop-plot observations over 11 years, and 127 households in Burirum province, with a total of 999 crop-plot observations. Table 1 shows village-level averages of number of years and plots per year in the data. The first column shows number of households per village, second column shows mean number of years per household, third column shows mean number of plots cultivated per year per household, and fourth column shows total number of observations. On average, we have data for seven years per household, with two crop-plots per cycle.

The data include information on usage and cost of labor and non-labor inputs used in separate production operations. We also have sets of measures of plot soil quality, some household socio-economic characteristics, and environmental data such as daily rainfall and temperature. During each monthly interview, households are asked in detail about all their rice cultivation activities. For each plot on which they grow rice, households report which operations were performed on the plot since the last interview, which inputs were used and in which quantities.

The fact that data were gathered monthly for each plot enables us to avoid imposing uniform bounds on stage timing and duration. Rather, we allow for plot-specific timing and duration of stages. That is, not all farmers are doing the same thing at the same time. The fact that timing of stages and of the overall production cycle vary across households and plots has several important

Table 1: Number of Observations per Household, Village, and Province

		Average number of:		
	Hhds, total	years per hhd	plots per hhd per year	Obs., total
		Sisaket province		
Village 1	38	9.0	2.6	598
Village 6	43	8.7	2.0	534
Village 9	38	8.3	2.3	434
Village 10	22	9.2	2.3	322
Province total	141	8.8	2.3	1,888
		Burirum province		
Village 2	24	7.2	1.3	135
Village 10	37	7.2	1.9	316
Village 13	30	6.2	1.9	227
Village 14	36	7.8	2.0	321
Province total	127	7.2	1.8	999

implications. Stage timing reflects variation in a number of plot-specific phenomena that determine it, such as plot characteristics, current state of the crop, effects of the unobserved production shocks, expectations of future production shocks, and the farmer's approach to rice cultivation. By incorporating variation in stage timing we take advantage of this additional information contained in the data. Moreover, aggregate production shocks such as rainfall have different effects on different plots because they may hit these plots during different production stages. Thus using plot-specific stage timing enables us to estimate the effects of changes in rainfall and temperature on rice cultivation with increased accuracy. When computing amounts of inputs used in each cultivation operation in each stage, we aggregate input usage over plot- and cycle-specific stage periods. For the model estimation, we use variation in stage timing between plots that is already present in the data and do not separately estimate the choice of planting date.

To map growth phases of rice plant into production stages, we look at the timing of cultivation operations required at different stages of plant growth. At different stages of growth the rice plant requires different types of care and so calls for performance of different operations. Operations involved in rice production can be divided into three groups. The first group involves preparatory operations necessary for initiation of plant growth. These include soil preparation, plowing, and planting. The final group involves terminal operations that take place at the end of production cycle, when plant growth nears conclusion. These include harvesting and preparation of harvest for sale and/or storage. The timing of both preparatory and terminal operations in production cycle is fairly intuitive: preparatory operations are performed at the beginning of production cycle in stage 1, and terminal operations are performed at the end of production cycle in stage 3. The intermediate group involves operations aimed at plant care during plant development, such as fertilizing and weeding. The timing of these intermediate operations is less intuitive.

For each plot, we determine the timing of stages 1 and 3 by looking at the timing of operations that intuitively correspond to each of these stages. That is, the timing of stage 1 is determined by farmer's timing of preparatory operations, and the timing of stage 3 is determined by farmer's timing of terminal operations. Time period between stages 1 and 3 constitutes stage 2.

Table 2 shows variation in stage duration and timing across years. As noted earlier, we determine the timing of stages individually for each plot in each cycle. The first column of table 2 shows the mean value across plots of the starting month for stage 1. Columns two through four show the standard deviation for the starting month of each stage by province, and the remaining columns show the province mean and standard deviation of duration of each stage in calendar months. It is clear from table 2 that while stage durations are fairly consistent over years, there is pronounced variation in stage timing across years. This suggests that there is effectively one timing decision in a given year, namely, the choice of starting month for stage 1.

Our weather data consist of village-level daily rainfall data from 1998, and province-level daily rainfall, temperature and solar radiation data from 1972. Temperature data include daily mean, minimum and maximum temperature measures.

Rainfall shocks are of high significance for rice cultivation. Rice is a very water-demanding plant. Most rice cultivation in Northeast Thailand is rainfed and makes little use of irrigation. According to the report by the International Rice Research Institute, rainfed rice is grown on approximately 92 percent of the area under rice cultivation in northeastern Thailand (Naklang, 2005). Farmers have to take the possibility of adverse rainfall shocks into account when making input decisions. The effect of rain on the rice crop also depends on temperature, as higher temperature can cause higher evapotranspiration and therefore lower soil moisture, the key latent variable. Both the direct effect of rain on crop and its integration with temperature are nonlinear. We use village-level total daily rainfall and province-level maximum daily temperature to construct a measure of rainfall shock as an estimated linear combination of rain, square of rain, temperature, and interaction of rain and temperature. Although weather is an aggregate shock, realized and expected weather varies across plots due to variation in stage timing.

Land variables describe the area used for rice cultivation as well as inherent characteristics of land that affect rice cultivation, such as quality of soil. In any given cycle households typically use several land plots. Land plots belonging to the same household need not be adjacent or even located close to each other. Typically, smaller plots are located close to the house and larger plots are spread around the village. As was illustrated in section 3, distributions of plots for villages

Table 2: Timing and Duration (in months) of Stages

Year	Stage 1 first month, mean	Starting month, st. dev. Stage 1	Stage 2	Stage 3	Stage 1 length mean	st. dev.	Stage 2 length mean	st. dev.	Stage 3 length mean	st. dev.
Sisaket province										
1999	6.28	0.55	0.58	0.99	1.09	0.29	1.84	0.95	2.89	0.92
2000	7.13	0.61	0.62	0.85	1.04	0.21	1.48	0.88	2.30	0.81
2001	7.35	0.59	0.67	0.99	1.09	0.30	1.73	0.92	2.11	1.24
2002	7.50	0.89	0.89	0.74	1.07	0.29	1.62	0.96	2.00	0.92
2003	7.20	0.84	0.84	0.83	1.02	0.16	1.63	0.95	2.09	0.68
2004	6.91	0.69	0.75	0.88	1.12	0.32	1.75	0.96	2.08	0.88
2005	7.62	0.92	0.93	0.79	1.05	0.22	1.58	0.86	1.63	0.68
2006	6.13	0.94	0.95	0.84	1.01	0.08	2.11	1.06	2.80	0.94
2007	6.43	1.01	1.01	0.92	1.01	0.09	2.06	0.89	2.66	0.95
2008	6.72	0.75	0.75	0.71	1.00	0.00	1.68	0.80	2.54	0.73
2009	6.63	0.87	0.87	0.73	1.00	0.00	1.49	0.62	2.80	0.74
Burirum province										
1999	6.11	0.56	0.56	0.76	1.00	0.00	1.85	0.76	3.35	0.84
2000	6.58	0.62	0.62	0.94	1.03	0.31	1.89	0.90	2.63	1.05
2001	6.68	0.82	0.85	0.92	1.02	0.15	2.00	0.96	2.48	0.92
2002	6.79	1.15	1.14	0.92	1.00	0.00	2.15	0.99	2.34	1.08
2003	6.42	0.73	0.73	0.81	1.00	0.00	2.53	0.85	2.04	0.76
2004	6.29	0.91	0.91	0.80	1.02	0.13	2.19	0.80	2.54	0.84
2005	5.82	0.65	0.67	1.22	1.05	0.28	3.05	1.38	2.41	1.11
2006	5.95	0.67	0.67	0.54	1.00	0.00	1.90	0.77	3.22	0.60
2007	5.71	0.87	0.87	0.72	1.00	0.00	2.81	0.95	2.47	0.81
2008	6.34	0.71	0.71	0.88	1.00	0.00	1.94	1.03	2.58	0.89
2009	5.10	0.89	0.89	0.73	1.00	0.00	3.21	0.99	2.62	0.73

in the same province overlap. It is even possible that plots belonging to the same household may actually be further apart than plots belonging to different households. Similarly, plots belonging to households from the same village may actually be further apart than plots belonging to households from different villages. Thus, whether plots belong to the same household or even the same village is not a good indicator of similarities in soil quality. Rather, soil quality is better captured by the location of plots relative to one another.

Variables that describe soil quality include measures of chemical composition of soil and its density. They indicate the soil's ability to provide nutrients to plants and to retain water and nutrients after rains and fertilizing. Soil variables describe initial conditions of rice production, corresponding to y_0 in terms of section 2 notation. We use two soil variables, cation exchange capacity (CEC) and organic matter. CEC measures soil's capacity to hold cation nutrients. It is determined by the amounts of clay and humus in the soil, which improve its nutrient and water-holding capacity. Organic matter helps the soil hold water and supplies nutrients.[23] In terms of section 2 notation, initial condition y_0 is an estimated linear combination of two soil quality measures and area under cultivation.

To construct a measure of intermediate "outputs", we use DSSAT - a powerful computer crop growth model.[24] The DSSAT system takes in amounts and timing of application of non-labor production factors such as seeds and chemical fertilizer, as well as detailed data on inherent soil quality and climatic conditions. The latter include actual historical data on daily variation in precipitation, maximum and minimum temperature, and solar radiation. DSSAT then employs physical and biophysical models of soil-plant-atmosphere interactions to simulate, day by day, the biological growth of the plant by computing crop-specific growth responses, measured precisely in laboratory conditions, to physical inputs and changes in soil, water, carbon, and nitrogen. DSSAT tracks plant's growth with 30 dynamic indicators, such as number of leaves per stem, root density, and stem weight.

The big advantage of DSSAT is the great level of detail and accuracy in modeling nonlinear crop response due to purely climatic and soil conditions. Note, however, that DSSAT does not take into account labor inputs or idiosyncratic shocks. In other words, DSSAT simulates plant growth due to exogenous climatic and soil conditions but does not consider all factors and shocks under which rice cultivation occurs in the field. DSSAT simulations are thus not exact measures of the actual crop state. Rather, they are approximations of the crop state that should occur under observed soil parameters, climatic conditions and non-labor crop inputs, as a result of quantified

[23]CEC is measured in meq/100g, or milliequivalents of hydrogen per 100 grams of dry soil. Organic matter content is measured in percent.

[24]DSSAT, or Decisions Support System for Agrotechnology Transfer, has been maintained and supported by the International Consortium for Agricultural Systems Applications (ICASA).

crop-specific growth responses measured precisely in laboratory conditions.

DSSAT traces the state of the crop throughout the whole production cycle, something we do not observe in the survey data. This allows us to use DSSAT simulations as imperfect estimates of intermediate stage outputs. We use measures of leaf weight and number of tillers as indicators of intermediate output from stage one, and measure of the progress of grain filling as indicator of intermediate output from stage two. Because DSSAT does not incorporate labor input, we use DSSAT indicators of intermediate output together with measures of labor inputs in previous stages to provide a more accurate proxy for intermediate output. In terms of section 2 notation, intermediate output $y_1 \exp(\varepsilon_1)$ is an estimated linear combination of two DSSAT measures of intermediate output from stage 1 and labor used in stage 1, and intermediate output $y_2 \exp(\varepsilon_2)$ is an estimated linear combination of one DSSAT measure of intermediate output from stage 2 and labor used in stage 2.

Typically there are several groups of laborers working on a given plot: household members who work on their own plot; villagers outside of the household, both relatives and non-relatives, who work for free, for labor exchange, or for pay; and workers hired through a broker, usually in teams. In other words, there are several ways for a household to adjust its labor input at a given point of rice production in response to realization of intermediate output from the previous stage or production shocks. We measure labor input as total hours by all workers for each production stage. Apart from labor, other production inputs are seeds and seedlings for planting and chemical fertilizer. Table 3 provides summary statistics for yields, production inputs, cultivated area, soil quality measures, factor and final output prices, and rain and temperature for each province. There are three inputs in stage 1: seeds, seedlings, and labor. Chemical fertilizer and labor are the two inputs in stage 2. Stage 3 uses only labor input. In total, there are six input demand equations corresponding to crop cultivation in our data.[25] For factor and final output prices, we use the village average as the measure of actual price[26]. This reflects the assumption that all households are price takers in inputs and output markets.

Finally, we have data on whether the household is a member of informal social network in the village. It is a direct measure of kinship that is not based on transactions, but on whether or not individuals in a given household are related by blood or marriage to the individuals in any other household, as of the time of the initial 1998 village census. As mentioned in the introduction, there is evidence that these networks are effective smoothing mechanisms for households and approxi-

[25]We do not include equipment use in our estimation as we do not have good data on utilization of household's equipment.

[26]For each price, we first impose lower and upper caps on plot-specific observations. These caps are the first and the 99th percentiles of the province distribution for a given production stage.

Table 3: Summary Statistics

	Sisaket province		Burirum province	
	Mean	St. Dev.	Mean	St. Dev.
Yield (kh/ha)	1,777.23	803.24	1,984.68	891.00
Initial conditions				
Area (ha)	1.12	0.92	1.30	0.87
CEC	2.32	1.20	2.56	1.53
Organic matter	0.52	0.36	0.43	0.25
Amounts of inputs (per hectare)				
Stage 1 chemical fert. (kg/ha)	11.32	41.90	40.67	117.27
Stage 2 chemical fert. (kg/ha)	147.15	142.52	113.35	103.93
Seeds (kg/ha)	35.24	66.03	96.46	77.98
Seedlings (sets/ha)[a]	788.69	703.06	280.63	559.72
Stage 1 labor (hrs/ha)	141.39	139.76	73.49	94.30
Stage 2 labor (hrs/ha)	34.82	68.96	22.96	60.20
Stage 3 labor (hrs/ha)	220.65	157.49	230.30	166.02
Factor and final output prices (in Thai Baht)				
Stage 1 chemical fert. (Baht/kg)	10.71	4.44	9.32	3.64
Stage 2 chemical fert. (Baht/kg)	10.76	4.42	9.34	4.03
Seeds (Baht/kg)	8.09	3.64	7.88	2.73
Seedlings (Baht/set)[a]	1.70	1.66	5.22	5.93
Stage 1 labor (Baht/hr)	101.37	93.39	73.59	59.89
Stage 2 labor (Baht/hr)	69.06	57.26	45.74	45.42
Stage 3 labor (Baht/hr)	127.61	85.79	48.85	21.27
Rice (Baht/kg)	7.35	2.73	7.44	2.27
Rainfall and temperature				
Stage 0 total rainfall (cm)	49.94	25.48	36.22	20.98
Stage 1 total rainfall (cm)	25.55	13.28	20.01	8.31
Stage 2 total rainfall (cm)	29.24	16.24	40.08	21.60
Stage 3 total rainfall (cm)	2.92	4.96	6.54	8.20
Stage 0 daily temperature[b] (C)	33.85	0.49	34.13	0.57
Stage 1 daily temperature[b] (C)	32.95	0.91	33.86	0.83
Stage 2 daily temperature[b] (C)	32.19	0.67	32.88	0.59
Stage 3 daily temperature[b] (C)	31.09	1.32	30.79	1.43

[a]One set contains about 100 seedlings.
[b]Temperature measure is stage mean of maximum daily temperature.

mate closely the full insurance.[27] In Sisaket province, 87 percent of sample households are part of their village's informal networks; the number for Burirum is 80 percent. To be consistent with our assumption of profit maximization, we estimate our model with Sisaket sample of plots belonging to households that are part of such networks.

5 Production Function Estimation

We estimate the composite production function and input decision rules as a system. The system approach to estimation accounts for common components of error terms in different equations and delivers estimates of the parameters of the composite production function as well as decision rules for all production inputs. We use iterative feasible general nonlinear least squares (IFGNLS) estimator. We estimate the model with Sisaket province data. We then use Burirum data to evaluate model's accuracy in predicting out of estimation sample.[28]

Because the equations in the system are sequential in nature, the feedback of the error terms goes only in one direction and the system is not simultaneous. Stage 1 inputs do not depend on any realized production shocks, and only contain household or plot-level unobservables that can potentially correlate with future production shocks. Stage 2 inputs depend on realization of stage 1 production shocks only. Stage 3 inputs depend on realization of production shocks in stages 1 and 2. Composite production function, or yield equation, depends on realization of production shocks in all three stages. Within a given stage, input amounts are determined simultaneously but do not influence each other, as is evident from equations (3) and (5).[29] So in nature the system is recursive across stages and seemingly unrelated within stages. However, in practice the system becomes seemingly unrelated across stages as well because of possible unobserved error components discussed in section 3.

System estimation with the error terms clustered by household improves efficiency of the estimator and reliability of inference. To address consistency of estimation, we need to examine the necessity to explicitly account for unobserved household or plot effects in the estimation, via household fixed effects (FE). This would be the case if either different households use different production technologies, so that parameters of production function in equation (1) vary across

[27] See Bonhomme, Chiappori, Townsend, and Yamada (2012), Chiappori, Samphatharak, Schulhofer-Wohl, and Townsend (2012), and Kinnan and Townsend (2012).

[28] We plan in future research to calibrate our model to predict the effects of weather shocks and climate change for the whole of Thailand. We use data from Burirum province, also located in Northeast Thailand but not adjacent to Sisaket province, to test model's ability to predict out of estimation sample.

[29] The same is true for stage 1 inputs. We have omitted stage 1 input equations from section 2 for the sake of brevity of presentation, as they have the same structure as equations for stage 2 inputs.

households, or if these unobserved effects correlate with our explanatory variables. The former is an unlikely scenario, given spatial compactness of our sample. Suppose there is some unobserved household-level effect, such as a shock to household's demographics or economic affluence, that cannot be promptly countered. Such shock could affect household's input choices and timing of planting decision, and therefore introduce endogeneity into equations. In this case, explicitly accounting for household fixed effects reduces estimation bias. An unobserved "managerial ability" of a household would have similar consequences.

There are several disadvantages to using fixed effects. First, FE reduce variation in the data, focusing on within variation only and not taking into account between variation. This leads to the estimation of only the short run effects of the explanatory variables, but not the long run effects. Among our explanatory variables, rainfall (or weather shock in general) and fertilizer use are of main interest. Within-household or within-plot variation in rainfall shock would capture rainfall variation across years and therefore is sufficient to estimate the effect of rainfall on yields that we are interested in. For fertilizer use, estimation based only on within variation will not capture the overall effect fertilizer has on yields that is characteristic for the population as a whole. However, given small number of plots cultivated per household, it will eliminate or at least substantially mitigate the effect of unobserved plot-level slope and elevation data and measurement error in plot soil quality, denoted by u_{khv} in the previous section. The average number of observations per household is 20.24 for Sisaket province and 12.96 for Burirum province, which is arguably not enough for consistent household FE or plot FE estimation. An important disadvantage of FE estimation is that it rules out use of explanatory variables that are fixed at the household level. As section 3 describes, in our data soil quality measures are fixed at plot level. Given small number of plots cultivated per household, it is possible that low variation in soil quality variables among plots belonging to the same household will lead to underestimation of the effect of soil quality on yields in the household FE estimation. We therefore do not use FE in estimation but bear in mind their potential presence.

Table 4 shows estimates of the structural coefficients of the model. First three rows show estimates of elasticities of substitution. Note the low elasticities for stages 1 and 3, both of which are not significantly different from zero and are significantly different from unity (the Cobb-Douglas case). Low stage 1 elasticity underlines the decisive importance of initial conditions such as soil quality for crop cultivation, suggesting that low soil quality is hard to compensate for with use of inputs. Thus heterogeneity in soil quality will inevitably lead to heterogeneity in yields. Low elasticity for stage 3 is intuitive. Stage 3 corresponds to harvesting operations, and if the crop hasn't developed, no amount of harvesting labor can compensate. Similarly, if no labor is employed to harvest the crop, the yields will be zero despite the presence of developed crop. Stage 2 elasticity estimate is negative, although it is different from zero only at 10 percent significance level. A negative elasticity estimate implies γ parameter greater than one, while $\gamma = 1$ is the upper limiting case of perfect substitution. The 5 percent confidence interval for γ_2 includes $\gamma_2 = 1$. We can interpret this estimate as suggesting that there is a fair amount of trade-off between stage 1 intermediate output, or how well young seedlings are developing after (trans)planting, and stage 2 production operations such as fertilizer application and weeding, as well as stage 2 weather realizations. Overall, elasticity estimates suggest that stage 2 is the most opportune period for farmer to impact his crop's development.

Measures of initial condition include cultivated area and soil quality measures. Significant positive coefficient estimate for organic matter variable confirms the expectation that soil quality has positive effect on crop cultivation. On the other hand, estimate of the coefficient for CEC is negative and significant. One possible explanation of this result is that our variables are imperfect measures of actual soil quality, as discussed in section 3. It also appears that larger plots enjoy higher yields. Stage 1 intermediate output is approximated by DSSAT measures of plot development and stage 1 labor, as discussed in section 4. Similarly, stage 2 intermediate output is approximated by DSSAT measures of plot development and stage 2 labor. For both stage 1 and stage 2 intermediate outputs, DSSAT measures are positive and significant. This is intuitive and means that healthier crop at the beginning of production stage contributes to better crop development during each stage. Note that in stage 1, labor used in previous stage is also statistically significant as a proxy for previous stage output. These results suggest that, on the one hand, DSSAT measures accurately crop's development and nonlinear interactions of non-labor inputs, soil quality, and weather realizations. On the other hand, accounting for only non-labor inputs into crop cultivation is not sufficient to capture farmer- and plot-specific crop development. Variation in labor input plays a significant role. This result underlines the danger of relying on biophysical simulations of yields alone when analyzing variations in yields in general, and in response to changes in weather realizations and climate change in particular. Stage 2, or crop maintenance stage, is much less labor-intensive than stage 1, when planting and transplanting - a very labor-intensive activity - are performed. With this in mind, it is not surprising that in stage 3 labor used in previous stage is not statistically significant as a proxy for previous stage output.

The weather shock is approximated by rain, rain squared, maximum daily temperature, and interaction of rain and temperature, by stage. As expected, the effect of weather shock is particularly pronounced in stage 2, when crop is developing. Estimates also confirm the importance of accounting for temperature when measuring the effect of rainfall on crop development, rather than rainfall alone, and the nonlinear nature of their interaction. Interestingly, estimates for stage 1 are not individually significant. Table 4 also reports results of the joint F test for all weather variables

Table 4: Model Parameters, Estimated with Sisaket Data

Variable group	Variable	Coefficient[a]	St. error[b]
Elasticity	Stage 1 elasticity	0.084	0.066
	Stage 2 elasticity	-0.454*	0.239
	Stage 3 elasticity	0.056	0.059
Initial condition	Area (ha)	0.896***	0.023
	CEC (meq/100g)	-0.084***	0.031
	Organic matter (%)	0.043**	0.019
Stage 1 output	Number of tillers (DSSAT)	0.414***	0.054
	Leaf weight (kg/ha) (DSSAT)	0.204*	0.118
	Stage 1 labor (hrs/ha)	0.277***	0.08
Stage 2 output	Grain filling (DSSAT)	0.391***	0.07
	Stage 2 labor (hrs/ha)	0.038	0.024
Stage 1 weather	Rainfall (cm)	0.079	0.048
	Rainfall squared	1.06E-05	5.31E-05
	Max daily temperature (C/10)	0.097	0.301
	Rain*temperature	-0.024	0.015
	F test for joint significance	12.990**	0.011
Stage 2 weather	Rainfall (cm)	0.123***	0.036
	Rainfall squared	-5.90E-06	5.97E-05
	Max daily temperature (C/10)	0.589*	0.346
	Rain*temperature	-0.038***	0.011
	F test for joint significance	13.420***	0.009
Stage 3 weather	Rainfall (cm)	-0.013	0.105
	Rainfall squared	6.25E-04	4.11E-04
	Max daily temperature (C/10)	0.344**	0.144
	Rain*temperature	0.001	0.034
	F test for joint significance	12.270**	0.015
Stage 1 inputs	Labor (hrs/ha)	0.155***	0.023
	Chem. fert. (kg/ha)	-0.013	0.02
	Seedlings (sets/ha)[c]	-0.094***	0.036
	Seeds (kg/ha)	-0.091***	0.021
Stage 2 inputs	Labor (hrs/ha)	-0.655***	0.096
	Chem. fert. (kg/ha)	-0.321***	0.086

[a] Chi-squared value for joint weather F tests.
[b] P-value for joint weather F tests.
[c] One set contains about 100 seedlings.
***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.

in each stage, which indicate that the joint effect of weather variables in each stage on final yields is highly significant.

The last two parts of table 4 show coefficient estimates for stage 1 inputs and stage 2 inputs. Positive and significant coefficient on stage 1 labor confirms earlier discussion on the high significance of labor during particularly labor-intensive stage 1 operations to successful crop cultivation. Coefficient estimate on stage 1 chemical fertilizer is not significant. Negative and significant coefficient estimate for stage 2 chemical fertilizer can be interpreted as indication of negative relationship between soil quality and yields. As we discussed earlier in section 3, our measures of soil quality are imperfect and might have poor accuracy for plots on which soil quality variables were not measured directly. Measurement error might also be an explanation for negative estimates of the coefficients on seeds and seedlings. Another possible explanation for these negative estimates is that we specify all inputs in per hectare terms. Planted area does not factor out completely with CES specification, and we account for this explicitly in estimation. However, as discussed in section 2, we use linear approximations of input equations rather than explicit input demands derived from the farmer's profit maximization problem. It is possible that per hectare input measures and planted area do not separate fully in these linear approximations, with some contamination remaining between the two. In this case, negative estimates on seeds and seedlings can reflect the fact that higher planting density is associated with low survival rate and weakness of young plants.

We next use the model to gauge the importance of rainfall versus other factors in yield variation between crop-plots and households. There are three main sources of yield variation. Weather, and rainfall in particular, is one. Predetermined heterogeneity in soil quality is another. Household's choice of planting timing and input amounts is the third source. Household's choice of timing affects the effective weather shock for a given crop-plot. We perform the following exercise. We first plot kernel density of actual sample yields. This is the solid line in figure 4 for Sisaket province and in figure 5 for Burirum province. It is obvious that there is a lot of heterogeneity in yields in each province. We next predict yields with our estimated model and plot the kernel density of these predictions. This is the long dash (_ _ _) line in figures 4 and 5. We see that the model does a good job in explaining existing heterogeneity in yields, both for Sisaket sample, which was used to estimate the model, and, notably, for Burirum sample as well. This latter result suggests that our model estimates can be used to make plausible yield predictions out of sample and for other rainfed rice growing regions in Thailand. Figure 6 illustrates the accuracy of model fit with actual data by plotting model prediction versus actual yield.

Next we exclude part of the third source of heterogeneity, namely, variation in household's choice of input amounts, by using province-year means of labor, seeds, seedlings, and chemical fertilizer to predict yields with the model. This is the dash-dot (_..._) line in figures 4 and 5. It captures variation in yields due to difference in timing, weather realizations, and soil quality. We see that it explains slightly less of the yield variation. Next, we also exclude variation in soil quality, by using province-year means of area, soil variables, labor, seeds, seedlings, and chemical fertilizer to predict yields with the model. This is the dotted (...) line in figures 4 and 5. It captures variation in yields due to difference in timing and rainfall realizations only. It is clear that while this is an important source of yield variation, it fails to explain large part of heterogeneity present in the data in both provinces. This exercise illustrates the significance of soil quality as a source of yield heterogeneity, even relative to variation in weather shocks.

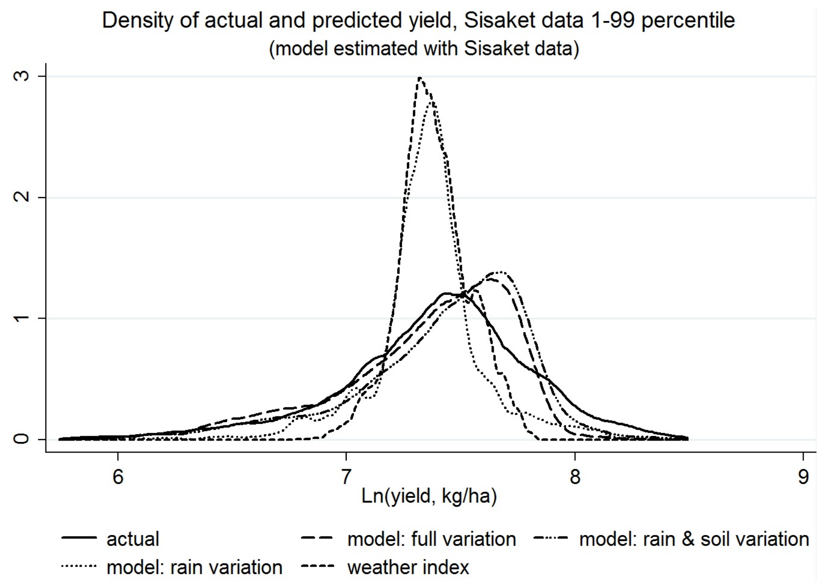

Figure 4: Kernel Density of Actual and Predicted Yields, Sisaket Sample

Finally, we construct a "weather index" for our sample, by regressing yields on measures of rain, rain squared, temperature, and interaction of rain and temperature in all three production stages, 12 weather variables in total, as well as village indicators, for each of the two provinces. We then predict yields with this simple OLS regression. This "weather index" measure of predicted yields is based on standard ways weather indices are constructed in weather-based index insurance contracts. Note that our "weather index" benefits from plot-specific timing of production stages, information that would not typically be available or utilized in standard weather indices. For Burirum province, our "weather index" measure has an additional advantage over model predictions, since the model was estimated with Sisaket sample while "weather index" regression was estimated separately for each province. The short dash (- - -) line in figures 4 and 5 corresponds to our version of "weather index". We see that, compared to our most constrained model-based prediction, the

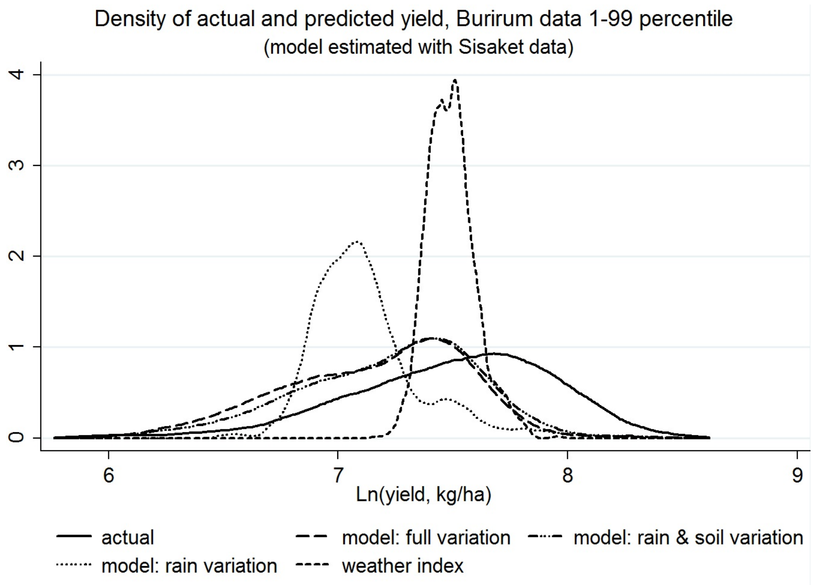

Figure 5: Kernel Density of Actual and Predicted Yields, Burirum Sample

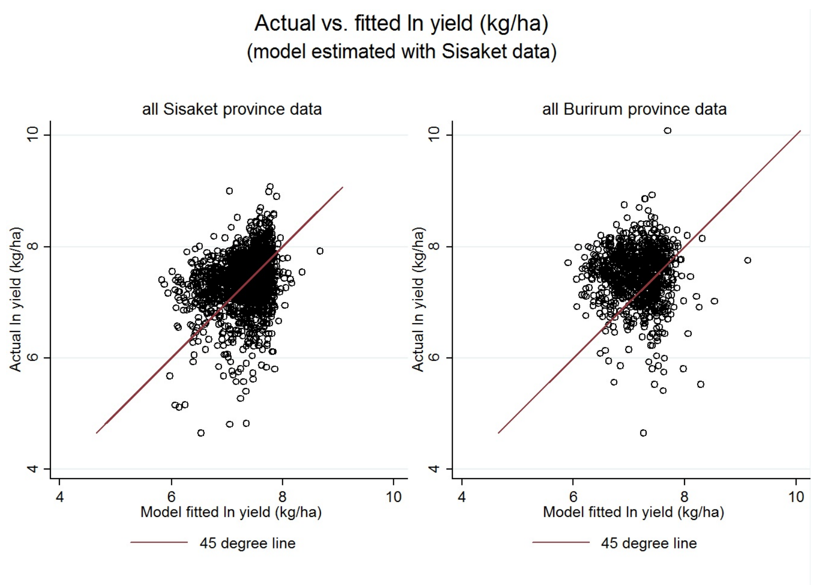

Figure 6: Accuracy of Model Fit with Actual Data

"weather index" explains slightly less variation for Sisaket, with a stronger result for Burirum. This last result illustrates that even when focusing only on one source of yield variation, namely, rainfall and timing, simple linear approximation can capture noticeably less of yield variation present in the data than structural approach.

In sum, yields data contain a lot of heterogeneity, a substantial part of which is due to plot-specific soil quality. An important and interesting question is how different sources of this heterogeneity interact with climate change. We next use the model to analyze the effect of climate change on yields.

6 Climate Change Impact

To analyze the effect of climate change on rice yields, we integrated our estimated economic model with DSSAT, climate change models, and a weather generation model.[30] We first describe the process we used to predict yields under climate change and characterize the climate change scenarios we considered. We then present our results.

We simulated future "synthetic" weather from the widely used WGEN weather simulation model (Richardson, 1981). The WGEN model begins by first calculating an extensive set of statistical parameters describing the observed, historical 1972-2002 daily weather data, including mean monthly amounts for all key input variables, as well as probabilities of wet days, probabilities of dry days, and within-year precipitation variation. WGEN then generates daily values for precipitation, maximum and minimum air temperature, and solar radiation for an N-year period at a given location. The precipitation component of WGEN is a Markov-chain–gamma-distribution model. The occurrence of wet or dry days is generated with a first-order Markov-chain model in which the probability of rain on a given day is conditioned on whether the previous day was wet or dry. Using statistics computed for the historical, 1972-2002 observed weather data, we generated 100 stochastic weather realizations for a "neutral" scenario, which assumes that future climate will be a direct, linear extension of the late 20th century. These weather realizations approximate weather conditions that would prevail in 2040-2069 under a neutral scenario. To generate future weather with SRES climate change scenarios, we inputted future changes to monthly precipitation and temperature and drew 100 realizations for each scenario.

For the analysis of climate change effects, we selected a random subsample of 83 plots in

[30] For all analysis in this section, we have used province-level rainfall data, rather than village-level data. The reason is that weather generator requires long time series of historical weather data in order to generate synthetic weather for the area and such data are available only on province level.

Sisaket province[31]. For each of these 83 plots, we used estimated model together with DSSAT to predict yields under each weather realization in each of the three considered climate scenarios.

We first used actual data to estimate farmers' timing decisions. Recall that equation (9) expresses farmer's choice of ε_0, or cumulative realizations of weather shocks at the time of planting, as maximizing expected profits conditional on information set $I_0 = \{y_0\}$. Thus the calendar month of planting that is observed in the data is a function of initial conditions y_0, such as area planted and soil quality, and cumulative weather realizations ε_0. We approximate this relationship as a linear combination of our measures of y_0 and ε_0. Using the whole sample and actual weather realizations, we regressed actual calendar month corresponding to the beginning of stage 1 on the plot's cultivated area[32], rain and temperature in January through March of current calendar year, rain and temperature in April through December of the previous calendar year, and household dummies. We then used the resulting equation to predict the starting month for a given plot for each synthetic weather realization under each climate scenario. Because this approach uses lagged data, we end up with 99 observations per plot per climate scenario. Table 5 shows the distribution of the starting month for the crop cultivation process in actual data, and predicted starting month for each of the three climate scenarios. It appears that cultivation activities shifted to the right in time and start on average one month later under the high emissions scenario compared to the neutral climate scenario, with no substantial difference between low emissions and neutral climates. Using these timing predictions, we constructed plot-specific timing of stages for each of the three climates, assuming that stage 1 lasts one month, stage 2 lasts two months, and stage 3 lasts two and a half months as in the actual data despite variation in rain over years. Using actual data, we estimated output prices as a function of concurrent monthly measures of rain and temperature, 11 monthly lags of rain and temperature, all measured at the provincial level, and indicators for calendar month.[33] Using actual data, we also estimated factor prices as functions of calendar month indicators, and interaction of month indicators with monthly rain and temperature at the provincial level. We then used these estimates to predict output and factor prices with synthetic weather data for each of the three climate scenarios we consider.

We next used our estimated production model together with DSSAT to predict inputs amounts

[31] We could not perform the analysis on the entire sample due to computational constraints. To ensure that final subsample contained no more than one observation per household, we first selected at random one plot per household, and then drew at random 83 plots from the remaining subsample.

[32] Again, we initially included soil quality measures in this regression, but they were highly statistically insignificant, both individually and jointly. Prediction of planting month with this regression is the first of several sequential computational steps in our climate change analysis. For this reason, we considered it advisable to use the most parsimonious specification of the timing equation and excluded soil measures due to their low explanatory power.

[33] Section 2 assumes that, for a given climate, output and factor prices are independent from weather shocks in the village. It is possible that the shift from one climate scenario to another alters the distributions of output and factor prices, which we here approximate by historical experience at the provincial level.

Table 5: Distribution of the First Month of Production Process

Calendar month	Actual data	% of sample with a given first month Neutral climate	High emissions	Low emissions
3		0.72	0.09	0.33
4	0.05	4.41	1.68	2.08
5	5.29	36.30	17.52	25.08
6	25.82	35.26	42.78	40.95
7	40.05	18.85	25.42	22.34
8	26.46	4.24	10.90	8.31
9	2.33	0.23	1.58	0.90
10			0.02	

for all three production stages. This step required alternation between the model and DSSAT, since DSSAT simulations are used in our model as measures of intermediate output and enter input equations, while DSSAT requires measures of non-labor production inputs in order to simulate plant growth. If we pre-specify amounts of seeds, seedlings and fertilizer, DSSAT alone can be run in one batch for the entire cropping reason. However, in our CES specification DSSAT interacts with labor input to form a measure of intermediate output. Consequently, input equations depend on DSSAT measures of intermediate output. We therefore needed model predictions of stage 1 physical non-labor inputs to generate DSSAT measure of stage 1 output; we needed DSSAT measure of stage 1 output to use the model to predict stage 2 fertilizer and labor use; we then needed model predictions of stage 1 and stage 2 non-labor inputs to generate DSSAT measure of stage 2 output; we finally needed this DSSAT measure of stage 2 output to use the model to predict stage 3 labor use. Finally, having now plot-, stage-, and weather realization-specific predictions of inputs and weather shocks, we used the model to predict yields. We repeated these steps for our subsample of 83 plots for 99 weather realizations under each of the three climate scenarios. In sum, this approach combines DSSAT's detailed modeling of complex soil and weather interactions with crop growth, and our model's structure of farmer's production choices. It also allows for variation in timing of production activities both across climates and with a given climate across different weather realizations.

We consider two alternative climate change trajectories - the high and low emissions. We also use a neutral climate scenario as a reference base. Table 6 uses 100 weather realizations generated by WGEN for each climate scenario and compares high and low emission climate scenarios to the neutral scenario. Panel A of table 6 compares amounts of daily precipitation and panel B compares average temperature during daylight hours. In each panel, the second column contains mean daily values for each month under the no change, neutral climate scenario. The next three columns

Table 6: Comparison of Neutral to Alternative High and Low Emissions Climates

Panel A: Daily amount of precipitation, in mm

	Neutral	Neutral to high emissions shift			Neutral to low emissions shift			Low to high emissions shift		
Month	Mean	Mean change	Percent	P-value	Mean change	Percent	P-value	Mean change	Percent	P-value
1	0.123	0.003	2.285	0.01	0.005	4.413	0.01	-0.003	-2.038	0.01
2	0.226	0.003	1.342	0.00	0.007	3.252	0.00	-0.004	-1.850	0.00
3	1.119	0.035	3.157	0.00	0.034	3.062	0.00	0.001	0.092	0.00
4	3.329	0.102	3.053	0.00	0.102	3.053	0.00	0.000	0.000	0.00
5	4.882	0.152	3.111	0.00	0.150	3.066	0.00	0.002	0.044	0.00
6	6.402	-0.059	-0.914	0.00	0.024	0.375	0.00	-0.083	-1.285	0.00
7	5.068	-0.001	-0.021	0.01	0.050	0.984	0.00	-0.051	-0.995	0.00
8	5.691	0.030	0.522	0.00	0.055	0.967	0.00	-0.025	-0.441	0.00
9	8.127	-0.082	-1.013	0.00	0.080	0.990	0.00	-0.163	-1.983	0.00
10	4.391	-0.042	-0.967	0.00	0.042	0.967	0.00	-0.085	-1.915	0.00
11	1.160	-0.014	-1.210	0.00	0.007	0.629	0.00	-0.021	-1.827	0.00
12	0.023	-0.001	-3.467	0.00	0.001	5.270	0.00	-0.002	-8.300	0.00

Panel B: Daily temperature during daylight hours, in degrees Centigrade

	Neutral	Neutral to high emissions shift			Neutral to low emissions shift			Low to high emissions shift		
Month	Mean	Mean change	Percent	P-value	Mean change	Percent	P-value	Mean change	Percent	P-value
1	26.697	2.300	8.615	0.00	1.300	4.869	0.00	1.000	3.572	0.00
2	29.083	2.300	7.909	0.00	1.300	4.470	0.00	1.000	3.291	0.00
3	31.391	2.300	7.327	0.00	1.300	4.141	0.00	1.000	3.059	0.00
4	32.357	2.300	7.108	0.00	1.300	4.018	0.00	1.000	2.971	0.00
5	31.339	2.327	7.426	0.00	1.300	4.148	0.00	1.027	3.147	0.00
6	30.464	2.078	6.821	0.00	1.300	4.267	0.00	0.778	2.449	0.00
7	29.894	2.100	7.024	0.00	1.300	4.349	0.00	0.800	2.563	0.00
8	29.414	2.199	7.478	0.00	1.300	4.420	0.00	0.899	2.928	0.00
9	30.350	1.205	3.971	0.00	1.300	4.283	0.00	-0.095	-0.300	0.00
10	28.434	1.298	4.567	0.00	1.300	4.572	0.00	-0.002	-0.005	0.00
11	27.191	1.172	4.311	0.00	1.300	4.781	0.00	-0.128	-0.449	0.00
12	25.733	2.420	9.405	0.00	1.300	5.052	0.00	1.120	4.144	0.00

address the shift from neutral to high-emissions climate. Column three shows the corresponding change in mean daily values, column four expresses this change in percent, and column five shows the probability value for the test of the equality means of daily precipitation (of average daily temperature in panel B) with the null of no change under neutral and high-emissions climates. In the same manner, columns six through eight address the shift from neutral to low emissions climate, and columns nine through 11 address the shift from low emissions climate to high emissions climate.

Climate change is more extreme under high emissions scenario. While daily temperatures increase under both climate scenarios, the magnitude of increase under the high emissions climate is about 40 percent higher. Daily precipitation increases throughout the year under low emissions climate. On the other hand, under the high emissions climate there is less rain in the second half of the year, starting in June, which is exactly the period of rice cultivation. Thus low emissions scenario brings moderate increase in temperature and more rain, while high emissions climate brings both higher increase in temperature and less rain for rice cultivation.

Figure 7 plots the kernel densities of means of household yields[34] for each of the three climates. There is a noticeable shift of the high emissions distribution to the left of neutral and low emissions distributions, and to a less extent of the low emissions distribution to the left of neutral. These are standard graphs of impact made in the spirit of what is reported in the literature. Figure 7 gives the impression that mean yields are falling substantially for most households. However, as will become evident in the discussion below, the diagram mixes up cross sectional change in heterogeneity and changing uncertainty facing the households.

We compared predicted yields under the three climate scenarios for each of the 83 plots, one at a time. We find noticeable variation from plot to plot. Before presenting numerical evidence, we first illustrate this point visually in figures 8 and 9. Each of these figures shows kernel densities of yields under neutral, high emissions, and low emissions climate scenarios for one of the plots in our sample. Each of these two plots is representative of approximately a third of our sample, with the remaining third of the sample somewhere in between. The difference in the effect of climate change on the yield distribution for these two plots is obvious. The plot in figure 8 has lower mean yields under high emissions scenario compared to neutral and low emissions scenarios, due to higher probability concentration on lower yield realizations. Distributions for both neutral and low emissions scenarios have yields concentrated at the same higher level; however, this concentration is much higher under neutral scenario. In other words, while the overall range of yield realizations doesn't change much with climate, the probability of the yield distribution shifts to the left, with higher probabilities associated with lower yield levels. This shift is more pronounced for the high

[34] All analysis in this section uses yields measured as natural log of kg per hectare.

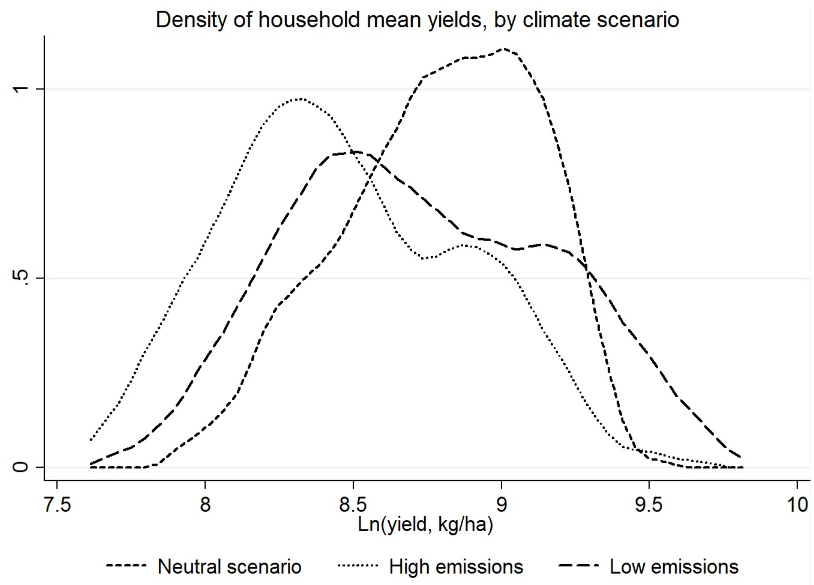

Figure 7: Kernel Density of Predicted Yields, by Climate Scenario

emissions scenario, and so the overall distribution for high emissions is skewed to the right.

The situation in figure 9 is different. Here mean yields are similar for the neutral and high emissions scenarios, with the latter climate having a slightly wider yield range. Under low emissions scenario, the probability concentration is highest at higher yield levels compared to neutral and high emissions climates.

Figures 8 and 9 also illustrate heterogeneity in yields across plots under the same climate scenario. Comparing these two plots for any given climate, it is clear that, despite the same aggregate weather shocks, these plots from the same geographical area face completely different yield distributions, both in terms of yield levels and of distribution shape

Given our panel data, we can consider variation in levels and shape of yields distributions across households for a given climate, across households given a shift from one climate to another, and within a household given a shift from one climate to another. We start with the former, looking at cross sectional variation between households under a given climate scenario. We first compute mean and coefficient of variation of yields for each household under each climate scenario. We then look at the distribution of these household means and coefficients of variation across all households, moving in the cross section so to speak, for a given climate. The results are summarized in table 7. The first three rows describe coefficient of variation, the normalized measure of uncertainty. For each climate, column two of table 7 reports fifth percentile of the cross sectional distribution of household-specific coefficients of variation of yields in our sample, and column

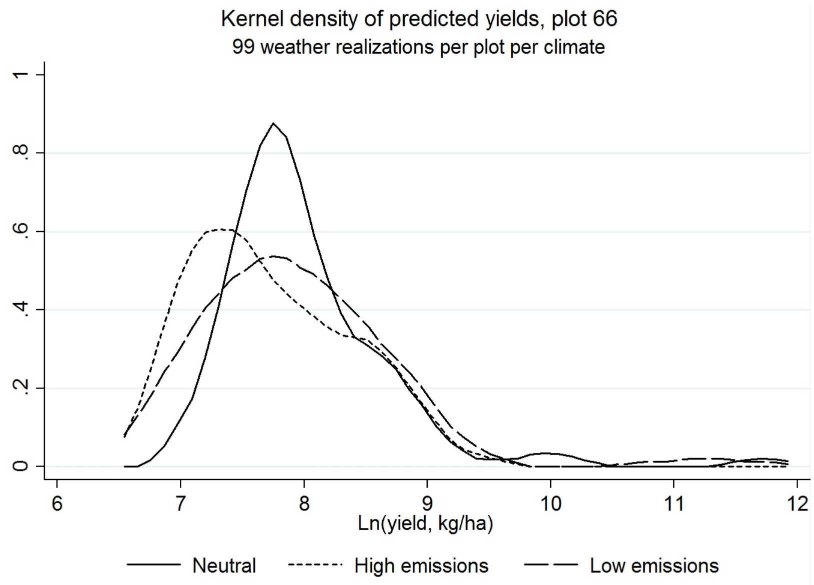

Figure 8: Kernel Density of Predicted Yields, by Climate Scenario

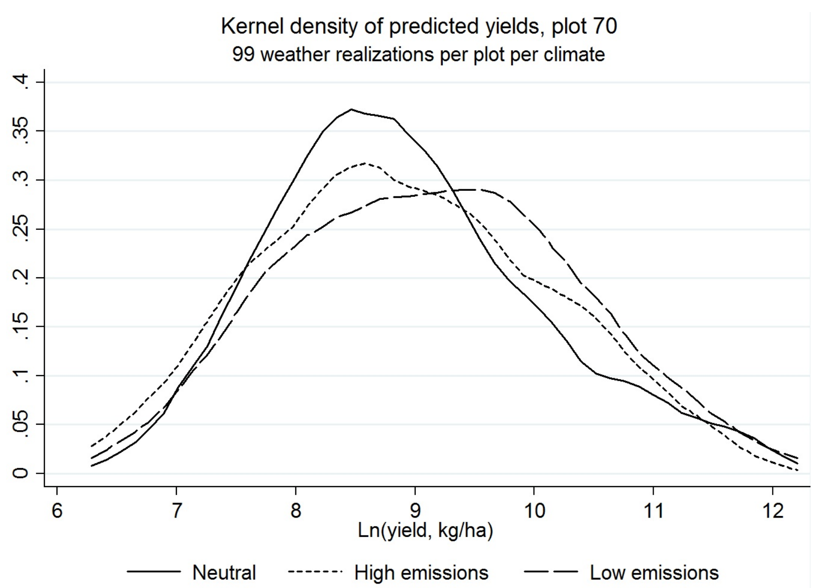

Figure 9: Kernel Density of Predicted Yields, by Climate Scenario

three reports the 95th percentile, similarly defined. Column four reports the percentage change from column two to column three. In the cross sectional distribution of the household-specific coefficient of yield variation, the fifth percentile describes low-variation households facing relatively little uncertainty about their yields, while 95th percentile describes high-variation households facing the highest uncertainty in our sample. Column four then measures the relative difference in uncertainty between households on these two opposite ends of uncertainty spectrum. We see from table 7 that this percentage increase in uncertainty is 35 percent under neutral climate, increases to 68 percent under high emissions climate, and is 41 percent under low emissions. These numbers demonstrate two points. First, it appears that heterogeneity across households in amount of uncertainty they face increases substantially with a shift to the high emissions climate. Second, even under neutral and low emissions climates, the heterogeneity in the cross section is non-trivial. Columns five through seven repeat the same exercise, comparing 25th and 75th percentiles. That is, columns five though seven present a milder comparison of households with differing but relatively moderate uncertainly levels (less sensitive to extremes). The same qualitative conclusions emerge, although now the increase in uncertainty between the two groups is more comparable and only modestly higher under the high emissions climate.

Table 7: Within-Climate Heterogeneity of Yield Distributions Across Households

Climate	5th pctile	95th pctile	% diff.	25th pctile	75th pctile	% diff.
			Household coefficient of variation			
Neutral	0.129	0.175	35.23	0.140	0.167	19.26
High emissions	0.100	0.169	68.13	0.129	0.160	23.69
Low emissions	0.126	0.177	41.31	0.142	0.166	16.86
			Household mean			
Neutral	8.255	9.194	11.37	8.567	9.065	5.82
High emissions	7.924	9.084	14.63	8.186	8.807	7.59
Low emissions	8.112	9.427	16.21	8.404	9.071	7.93

The last three rows of table 7 describe the cross sectional distribution of household mean yields and are constructed in the same manner. We see that, under a given climate, relative variation in mean yield levels across households, as measured by the percent difference, is much lower than relative variation in uncertainty. Interestingly, for mean yields the low emissions climate has largest cross sectional difference between low-level and high-level households, as presented in columns four and seven, though the difference in these numbers between climates is small. So, in sum, the high emissions scenario increases the cross sectional dispersion in uncertainty and, to the much lesser extent, increases the cross sectional dispersion in means.

We now examine the variation in yield levels and uncertainty for a given household across climates. For each household, we have measures of its mean yield and coefficient of yield variation for each of the three climates, and can compare how these statistics change with a climate shift. We first look at changes in coefficient of variation. We consider shifts from neutral to high emissions climate, from neutral to low emissions climate, and from low to high emissions climate. For each of these climate shifts, we compute the percent change in coefficient of variation for each household in our sample. We then compute the percent of the sample that experiences change in coefficient of variation less that -10 percent, between -10 and -5 percent, between 5 and 10 percent, and greater than 10 percent. These numbers are reported in table 8. For example, row one tells us that, given a shift from neutral to high emissions climate, 14.5 percent of households experience more than a 10 percent decrease in their coefficient of yield variation, and 32.5 percent of households experience between five and 10 percent decrease in their coefficient of yield variation. It is clear from table 8 that household-level uncertainty decreases under the high emissions climate and is the highest under low emissions climate. This result is not surprising given lower rainfall under high emissions climate and higher rainfall under low emissions climate relative to the neutral scenario. Tables 7 and 8 together illustrate that a more extreme climate change can lead to lower uncertainty of yields from household's perspective, at the same time contributing to higher heterogeneity across households in both uncertainty about and, to a smaller degree, levels of yields.

Table 8: Between-Climate Change in Uncertainty at Household Level

	% of sample with given change in c.v.			
Climate shift	< -10%	[-10%, -5%]	[5%, 10%]	> 10%
Neutral to high emissions	14.46	32.53	4.82	0.00
Neutral to low emissions	3.61	4.82	8.43	10.84
Low to high emissions	27.71	36.14	3.61	0.00

We conclude by comparing mean yields under different climates for each household. These results are presented in table 9. Columns two and three describe results significant at 5 percent level, and columns five and six describe results significant at 10 percent level. Columns two and four report the percent of the sample with significant change in mean yields, and columns three and five report the corresponding percent change in yields for plots with significant change. Column one indicates whether the significant change was a decrease or an increase in yields. Four results are evident. First, shift from either neutral or low emissions to high emissions climate results in statistically significant decrease in mean yields for majority of the sample. Second, these decreases in yields, although widespread, are relatively small in magnitude. Third, there is almost no significant increase in yields under any alternative climate scenario relative to neutral climate. Even a shift from neutral to a milder, low emissions climate results in higher yields only for 2.4 percent of the sample, and this result is not significant at 5 percent level. Fourth, only a small fraction of households experience significant difference in yields between neutral and low emissions climates, and for this small fraction the difference in yields is significant only at 10 percent level.[35]

Table 9: Comparison Test of Mean Yield by Household Across Climates

	5% significance		10% significance	
	% of sample	Yield change, %	% of sample	Yield change, %
		Neutral to high emissions		
Decrease	68.67	-3.79	73.49	-3.73
Increase	0	0	0	0
		Neutral to low emissions		
Decrease	0	0	8.43	-2.26
Increase	0	0	2.41	3.75
		Low to high emissions		
Decrease	22.89	-3.24	68.67	-2.81
Increase	0	0	0	0

7 Conclusion

We develop a three stage structural model of rice cultivation, integrate it with a biophysical model and a weather generator, and use this framework to evaluate the effect of climate change on rice yields in Northeast Thailand. We estimate the economic model with data from Sisaket province and then use data from Burirum province, also located in the Northeast of the country but not adjacent to Sisaket, to predict out of sample. The results suggest that our model can be calibrated to estimate the effect of variations in weather and climate on rainfed rice for all of Thailand with the use of Agricultural Census data. The census data contain key variables that can be matched to our study, such as soil data, and rainfall and temperature from meteorological stations, all of these geolocated. In turn, such an extended study would be an input into considerations of rainfall or weather insurance and into micro founded macro models.

[35]We repeat the same exercise with profits, comparing mean profits for each household between climate scenarios. There are no statistically significant differences in mean profits for households in our sample. One exception is that profits decrease when moving from low emissions to high emissions climate for one household in our sample with this difference in mean profits significant at 5 percent level, and for two households with this difference in mean profits significant at 10 percent level.

Household labor constitutes a non-trivial part of total labor input for all households; however, we do not have an adequate measure of wages for family labor. For this reason, we forego any further analysis of profits. This issue is common in the literature (see, for example, Karlan, Osei, Osei-Akoto, and Udry (2012)).

One result that permeates throughout the paper is substantial cross sectional heterogeneity in yields, both in actual data and in data predicted for alternative climates. Model estimation results indicate that variation in soil quality is a substantial source of this heterogeneity in yields. Another source is variation in timing of planting. As a result, a "common" weather shock within the village has differing effect on rice crops cultivated by different farmers on plots with different soil. Again, thinking ahead, this has large implications for the provision of insurance and the next generation of models.

When analyzing the effects of climate change on yields, it is important to distinguish the cross sectional effect from the effect from the farmer's perspective. As the results in our paper illustrate, uncertainty, as measured by coefficient of variation, from the farmer's perspective decreases noticeably under the high emissions climate. However, it decreases much more for farmers at the "lower" uncertainty end, that is, farmers who low uncertainty levels to begin with, relative to farmers at the "high" uncertainty end. As a result, even though individual farmer-level uncertainty in yields decreases, cross sectional heterogeneity in uncertainty, as well as, to a lesser extent, mean yields, go up.

Finally, our results illustrate the complementary nature of production activities for two out of three crop cultivation stages, underlining the significance of the choice of planting date. This also means that simpler reduced form estimates of weather variation on yields can be misleading.

References

ALEM, M., AND R. M. TOWNSEND (2007): "An Evaluation of Safety Nets and Financial Institutions in Crisis and Growth," unpublished manuscript, University of Chicago.

ANTLE, J. M. (1983): "Sequential Decision Making in Production Models," *American Journal of Agricultural Economics*, 65, 282–290.

ANTLE, J. M., AND S. M. CAPALBO (2001): "Econometric-Process Models for Integrated Assessment of Agricultural Production Systems," *American Journal of Agricultural Economics*, 83(2), 389–401.

ANTLE, J. M., AND S. A. HATCHETT (1986): "Dynamic Input Decisions in Econometric Production Models," *American Journal of Agricultural Economics*, 68, 939–949.

ANTLE, J. M., AND J. J. STOORVOGEL (2006): "Incorporating Systems Dynamics and Spatial Heterogeneity in Integrated Assessment of Agricultural Production Systems," *Environment and Development Economics*, 11(01), 39–58.

BINFORD, M. W., T. J. LEE, AND R. M. TOWNSEND (2004): "Sampling Design for an Integrated Socioeconomic and Ecological Survey by Using Satellite Remote Sensing and Ordination," *Proceedings of the National Academy of Sciences*, 101(31), 11517–11522.

BONHOMME, S., P. A. CHIAPPORI, R. M. TOWNSEND, AND H. YAMADA (2012): "Sharing Wage Risk," Working paper.

CHIAPPORI, P. A., K. SAMPHATHARAK, S. SCHULHOFER-WOHL, AND R. M. TOWNSEND (2012): "Heterogeneity and Risk-Sharing in Village Economies," Working paper.

CRUZ, R. V., H. HARASAWA, M. LAL, S. WU, Y. ANOKHIN, B. PUNSALMAA, Y. HONDA, M. JAFARI, C. LI, AND N. H. NINH (2007): "Asia," in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. by M. Parry, O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson, pp. 469–506. Cambridge University Press, Cambridge, UK.

GIORGI, F., AND L. O. MEARNS (2002): "Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations Via the Reliability Ensemble Averaging (REA) Method," *Journal of Climate*, 15(10), 1141–1158.

HASEGAWA, T., S. SAWANO, S. GOTO, P. KONGHAKOTE, A. POLTHANEE, Y. ISHIGOOKA, T. KUWAGATA, H. TORITANI, AND J. FURUYA (2008): "A Model Driven by Crop Water Use and Nitrogen Supply for Simulating Changes in the Regional Yield of Rainfed Lowland Rice in Northeast Thailand," *Paddy and Water Environment*, 6, 73–82.

JUST, R. E., AND R. D. POPE (1978): "Stochastic Specification of Production Functions and Economic Implications," *Journal of Econometrics*, 7, 67–86.

——— (2001): "The Agricultural Producer: Theory and Statistical Measurement," *Handbook of Agricultural Economics*, 1, 629–741.

KARLAN, D., R. OSEI, I. OSEI-AKOTO, AND C. UDRY (2012): "Agricultural Decisions after Relaxing Credit and Risk Constraints," Working paper.

KINNAN, C., AND R. M. TOWNSEND (2012): "Kinship and Financial Networks, Formal Financial Access and Risk Reduction," *The American Economic Review*, 102(3), 289–93.

NAKICENOVIC, N., J. ALCAMO, G. DAVIS, B. DE VRIES, J. FENHANN, S. GAFFIN, K. GREGORY, A. GRUBLER, T. Y. JUNG, AND T. KRAM (2000): *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

NAKLANG, K. (2005): "Managing Soil Fertility for Sustainable Rice Production in Northeast Thailand," in *Rice is Life: Scientific Perspectives for the 21st Century*, ed. by K. Toriyama, K. L. Heong, and B. Hardy, pp. 357–359. International Rice Research Institute (IRRI), Proceedings of the World Rice Research Conference held at Tsukuba, Japan on 5-7 Nov 2004.

PARRY, M., C. ROSENZWEIG, A. IGLESIAS, M. LIVERMORE, AND G. FISCHER (2004): "Effects of Climate Change on Global Food Production under Sres Emissions and Socio-Economic Scenarios," *Global Environmental Change*, 14(1), 53–67.

PAXSON, C. H. (1992): "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand," *The American Economic Review*, 82(1), 15–33.

RICHARDSON, C. W. (1981): "Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation," *Water Resources Research*, 17(1), 182–190.

ROSENZWEIG, C., A. IGLESIAS, X. B. YANG, P. R. EPSTEIN, AND E. CHIVIAN (2001): "Climate Change and Extreme Weather Events: Implications for Food Production, Plant Diseases, and Pests," *Global Change & Human Health*, 2(2), 90–104.

STOORVOGEL, J. J., J. M. ANTLE, C. C. CRISSMAN, AND W. BOWEN (2004): "The tradeoff Analysis Model: Integrated Bio-Physical and Cconomic Modeling of Agricultural Production Systems," *Agricultural Systems*, 80(1), 43–66.

VALDIVIA, R. O., J. M. ANTLE, AND J. J. STOORVOGEL (2012): "Coupling the Tradeoff Analysis Model with a Market Equilibrium Model to Analyze Economic and Environmental Outcomes of Agricultural Production Systems," *Agricultural Systems*, 110(C), 17–29.