The Impact of Climate Change on Rice Yields: the Importance of Heterogeneity and Family Networks

John Felkner

Kamilya Tazhibayeva

Robert Townsend

August 2012

1 Introduction

We quantify the impact of weather variation and climate change on the production of rain-fed rice, a major crop of Thailand both for domestic consumption and export[1], as it is for other countries in Southeast Asia. More generally rice and other staple crops are of great importance to the world economy. 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, and for climate shift scenarios.

Indeed this heterogeneity-induced distinction between aggregate and idiosyncratic shocks takes us to possibility of weather insurance for a given contemporary climate, and of insurance for climate shifts, and how these might be designed. This does carry us well beyond the scope of the current paper. Nevertheless, we take up the implications of what we have learned in a concluding epilogue section.

We take advantage in this paper of quite detailed agricultural data, gathered in the field in the Northeast of Thailand, in a semi-arid tropics zone, as part of Townsend Thai monthly survey, a panel from 1999-2009 covering 11 crop cycles. The data utilized here covers 268 households and 2,887 crop plots in two provinces, Sisaket and Burirum. The data itself includes the details of inputs including fertilizer, labor, land planted; harvests and hence yields; crop operations by name hence allowing multi stage of crop production; a suite of environmental data including initial plot soil measurements, monthly rainfall within village and temperature at nearby met station; economic variables such as consumption and borrowing/gifts; and demographic variables including household composition and kinship networks in the village (from a village census).

We proceed from aggregates, from the top down, so to speak, as we move into more and more micro heterogeneity. See figure 6, for histograms of actual yields versus various of the models, from reduced form to structural. All of these though are based in some degree on our baseline multi stage crop production model, at least in the organization of the data. First, we run "standard" regressions of observed sample rice yields onto monthly rain, rain squared and temperature, and interaction of rain and temperature. Here the data we have allows us to identify by crop plot the stage of operation, so, within a given year we have rain and temperature correctly assigned to the stage of production for each plot. We do this for each province one at time, to allow some local variation. Largely though, we use one province, Sisaket, to estimate and predict out of sample, to

[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).

Burirum, as in figure 7 using the data we have there, so we can quantify how well we would do if we extrapolate our results to rice growing regions in the larger Thai Kingdom.

Next we include information on tambon average soil and crop inputs through the lens of our multistage crop production model, which we describe below, but fixing household variables at common (average) values. As the figures reveal, the histogram of predicted yields is less peaked and picks up more of the actual variation, that are in the tails. Third we maintain tambon average in inputs but allow the measured soil variation of the sample. Finally, we use both soil and input variation and the full force of our actual model. The point of this exercise is to document that measured heterogeneity is substantial part of variation in yields, both in the exogenous soil part and in the endogenous input part. A companion table 4 shows that yields vary significantly by cation exchange capacity, organic matter, area cultivated, and usage of inputs in different production stages.

The baseline model we construct and estimate is a multistage crop production model in which farmers are imagined to maximize expected profits, as if they were risk neutral (see the discussion in the concluding section). Yields in the final stage are a function of rainfall at the end of that stage and inputs during that stage. The initial condition for the final stage is the condition of crop plot at that point in time. In turn that plot condition is a function of what happened in the previous stage, both inputs and shocks. We use a combination of a biophysical crop production model 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 take expectations of rainfall and prices into the future based on current information. This we summarize from regression analysis using historical data. Finally, the timing of planting is incorporated through its effect on the timing of stages and therefore on the rainfall realization for a given plot.

Our baseline model displays some key, reassuring features. There is limited substitutability between soil quality and planting activities, underlying the importance of variation in soil quality. There is more substitutability between planting stage and intermediate growing stage, and zero substitutability between growing stage and harvesting. Our estimates also indicate that the effect of weather, and rainfall in particular, is most pronounced during planting stage. Combined, these results suggest that the intermediate growing stage is the most opportune period for farmers to impact their yields, which is the stage when chemical fertilizer is used most extensively. 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 simulating soil models like DSSAT do, and at the same time demonstrate that using such simulation models without accounting for human input is not sufficient.

We study though the lens of the model as it stands the impact of variation in weather, for the given contemporary climate. Figure 10 shows simulated histograms of yields for distinct plots using 99 years of synthetic weather generated for the current climate scenario. One can see heterogeneity across plots in both mean yields and in mass in the tails.

On top of this we 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, Lal, Wu, Anokhin, Punsalmaa, Honda, Jafari, Li, and Ninh, 2007). We use an "ensemble-mean"[2] 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[3]. 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)[4], 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.

According to IPCC ensemble-mean predictions, results predict a net increase in yearly average temperature of between 1.32°C (lowest emissions scenario B1) and 2.01 $^\circ$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,

[2]"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).

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

[4]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 theoretically favored PPP exchanges rates, which correct for differences in purchasing power. However, for this micro-study, we accept these scenarios as given.

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 depending on CO2 regimes.

Again in our study, there is variation in climate shift impact across plots both in means and in entire distribution of yields. Though warming and varied patterns of rainfall make future scenarios worse than the current climate, there is heterogeneity in the impact of low versus high emissions, with some plots actually doing better under some circumstances under the latter.

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)[5]. 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.

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

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 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, after output from the previous stage is observed, before production shocks for the current stage are realized, and 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.

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 at the end of stage i. 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[6]:

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

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 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

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

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

$$f\left(y_0, \{x_i, \varepsilon_i\}_{i=1}^3\right) = \tag{2}$$

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

where $\delta_i = A_i \theta_i^{1/\gamma_i}$, $\rho_i = A_i (1-\theta_i)^{1/\gamma_i}$, 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[7]. Let p denote the price of final output and $w_i = \left(w_{i1}\ w_{i2}\ \ldots\ w_{i,N_i}\right)$ denote a vector of stage i input prices. Let the bar denote expectations, so $\bar{z}_j = E\left[z_j\right]$. 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: $I_i = \left\{\left\{y_j\right\}_{j=0}^{i-1}, \left\{\varepsilon_j\right\}_{j=1}^{i-1}, \left\{w_j\right\}_{j=1}^{i}\right\}$. Stage 3 information set also includes final output price p.

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, $I_3 = \left\{\left\{y_j\right\}_{j=0}^{2}, \left\{\varepsilon_j\right\}_{j=1}^{2}, \left\{w_j\right\}_{j=1}^{3}, p\right\}$, includes realization of stage 2 output and therefore of production shocks that occurred in stages 1 and 2. The farmer solves[8]

$$\max_{\{x_{3n}\}_{n=1}^{N_3}} p \underbrace{A_3 \left(\theta_3 \left(y_2 \exp(\varepsilon_2)\right)^{\gamma_3} + (1-\theta_3)\left(B_3 \prod_{n=1}^{N_3} x_{3n}^{\alpha_{3n}}\right)^{\gamma_3}\right)^{1/\gamma_3}}_{y_3} \exp(\bar{\varepsilon}_3) - \sum_{n=1}^{N_3} w_{3n} x_{3n},$$

with expectation of stage 3 production shock, $\bar{\varepsilon}_3$, conditional on information set I_3. The first order conditions are

$$p \exp(\bar{\varepsilon}_3) \frac{\partial y_3}{\partial x_{3n}} = w_{3n}\ \forall n \in \{1, \ldots, N_3\}$$

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 $j, k \in \{1, \ldots, N_3\}$.

[7] 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.

[8] We are approximating $\overline{\exp(\varepsilon)}$ with $\exp(\bar{\varepsilon})$, where the reasonableness of this approximation increases with the size of ε's mean relative to its variance.

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} $\forall k \in \{1, ..., N_3\}$:

$$x_{3k} = \left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3} \frac{\alpha_{3k} y_2 \exp(\varepsilon_2)}{B_3 w_{3k} \lambda_3} \left[(1-\theta_3)^{\frac{1}{\gamma_3 - 1}} \left(A_3 B_3 \lambda_3 p \exp(\bar{\varepsilon}_3)\right)^{\frac{\gamma_3}{\gamma_3 - 1}} - 1\right]^{-1/\gamma_3} \tag{3}$$

$\forall k \in \{1, ..., N_3\}$, 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

$$\ln x_{3k} \approx \ln \alpha_{3k} + C_3 + \ln y_2 + \varepsilon_2 - \frac{1}{1-\gamma_3} \ln \frac{w_{3k}}{p} + \tag{4}$$
$$+ \frac{\gamma_3}{1-\gamma_3} \sum_{n=1, n \neq k}^{N_3} \alpha_{3n} \ln \frac{w_{3k}}{w_{3n}} + \frac{1}{1-\gamma_3} \bar{\varepsilon}_3 \ \forall k \in \{1, ..., N_3\},$$

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, farmer's information set is $I_2 = \left\{\{y_j\}_{j=0}^{1}, \varepsilon_1, \{w_j\}_{j=1}^{2}\right\}$. Farmer solves

$$\max_{\{x_{2n}\}_{n=1}^{N_2}} \bar{p} \bar{y}_3 \exp(\bar{\varepsilon}_3) - \sum_{n=1}^{N_2} w_{2n} x_{2n} - E\left[\sum_{n=1}^{N_3} w_{3n} x_{3n}\right],$$

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 expected stage 3 input demands $\{\bar{x}_{3n}\}_{n=1}^{N_3}$, 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:

$$E\left[\sum_{n=1}^{N_3} w_{3n} x_{3n}\right] = \left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3} \frac{y_2 \exp(\bar{\varepsilon}_2)}{B_3 \bar{\lambda}_3 g_3^{1/\gamma_3}} \text{ and } \bar{y}_3 = A_3 \theta_3^{1/\gamma_3} y_2 \exp(\bar{\varepsilon}_2) \left(1 + \frac{1}{g_3}\right)^{1/\gamma_3},$$

where $g_3 = (1-\theta_3)^{\frac{1}{\gamma_3 - 1}} \left(A_3 B_3 \bar{\lambda}_3 \bar{p} \exp(\bar{\varepsilon}_3)\right)^{\frac{\gamma_3}{\gamma_3 - 1}} - 1$. The first order conditions for $x_{2k}, k = 1, \dots, N_2$, are

$$A_3 \left(\frac{\theta_3 (1+g_3)}{g_3}\right)^{1/\gamma_3} \frac{\partial y_2}{\partial x_{2k}} \exp(\bar{\varepsilon}_2 + \bar{\varepsilon}_3) = \frac{w_{2k}}{\bar{p}} + \left(\frac{\theta_3}{(1-\theta_3) g_3}\right)^{1/\gamma_3} \frac{\partial y_2}{\partial x_{2k}} \frac{\exp(\bar{\varepsilon}_2)}{B_3 \bar{\lambda}_3 \bar{p}}.$$

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. Concurrent marginal cost is the input's real price, $w_{2k}/\bar{p}$. 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, though 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, $\bar{\varepsilon}_2$, but also for the subsequent stage 3, $\bar{\varepsilon}_3$.

From the first order conditions, $\alpha_{2j} w_{2k} x_{2k} = \alpha_{2k} w_{2j} x_{2j}$ for all $j, k \in \{1, \dots, N_2\}$. Again, we express all stage 2 inputs in terms of one stage 2 input, say, x_{21}, as $x_{2k} = \frac{\alpha_{2k} w_{21}}{\alpha_{21} w_{2k}} x_{21}$ for all $k \in \{1, \dots, N_2\}$. Solving first for the optimal x_{21}, we can then solve for optimal stage 2 input levels:

$$x_{2k} = \left(\frac{\theta_2}{1-\theta_2}\right)^{1/\gamma_2} \frac{\alpha_{2k} y_1 \exp(\varepsilon_1)}{B_2 w_{2k} \lambda_2} \left[(1-\theta_2)^{\frac{1}{\gamma_2 - 1}} (A_2 B_2 \lambda_2 P_3 \exp(\bar{\varepsilon}_2))^{\frac{\gamma_2}{\gamma_2 - 1}} - 1\right]^{-1/\gamma_2} \tag{5}$$

$\forall k \in \{1, \dots, N_2\}$, where $\lambda_{2n} = \prod_{n=1}^{N_2} \left(\frac{\alpha_{2n}}{w_{2n}}\right)^{\alpha_{2n}}$, $P_3\left(\bar{\varepsilon}_3, \{\bar{w}_{3n}\}_{n=1}^{N_3}, \bar{p}\right) = \theta_3^{1/\gamma_3} \bar{p} R_3^{\frac{\gamma_3 - 1}{\gamma_3}}$, and

$$R_3\left(\bar{\varepsilon}_3, \{\bar{w}_{3n}\}_{n=1}^{N_3}, \bar{p}\right) = (A_3 \exp(\bar{\varepsilon}_3))^{\frac{\gamma_3}{\gamma_3 - 1}} - \left((1-\theta_3)\left(B_3 \bar{\lambda}_3 \bar{p}\right)^{\gamma_3}\right)^{\frac{1}{1-\gamma_3}}.$$

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 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, $\partial x_{2k}/\partial\bar{\varepsilon}_3$, is positive, while the marginal effect of expected stage 3 real input prices on stage 2 input demands, $\partial x_{2k}/\partial\left(\bar{w}_{3k}/\bar{p}\right)$, 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

$$R_3\left(\bar{\varepsilon}_3, \{\bar{w}_{3n}\}_{n=1}^{N_3}, \bar{p}\right) = \frac{(1-\theta_3)^{\frac{1}{\gamma_3-1}}\left(A_3B_3\bar{\lambda}_3\bar{p}\exp\left(\bar{\varepsilon}_3\right)\right)^{\frac{\gamma_3}{\gamma_3-1}}-1}{(1-\theta_3)^{\frac{1}{\gamma_3-1}}\left(B_3\bar{\lambda}_3\bar{p}\right)^{\frac{\gamma_3}{\gamma_3-1}}}. \tag{6}$$

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. Using the approximation $\ln(x-1) \approx \ln x$, we can approximate $\ln R_3$ as $\ln R_3 \approx \frac{\gamma_3}{\gamma_3-1}\left(\ln A_3 + \bar{\varepsilon}_3\right)$.[9] Applying the log approximation to equation (5), we can write stage 2 input demands as

$$\ln x_{2k} \approx \ln\alpha_{2k}+C_2+\ln y_1+\varepsilon_1-\frac{1}{1-\gamma_2}\ln\frac{w_{2k}}{\bar{p}}+\frac{\gamma_2}{1-\gamma_2}\sum_{n=1,n\neq k}^{N_2}\alpha_{2n}\ln\frac{w_{2k}}{w_{2n}}+\frac{1}{1-\gamma_2}\left(\bar{\varepsilon}_2+\bar{\varepsilon}_3\right) \tag{7}$$

$\forall k \in \{1, ..., N_2\}$, where the common component of the constant term is $C_2 = \frac{1}{\gamma_3(1-\gamma_2)}\ln\theta_3 + \frac{1}{\gamma_2}\ln\theta_2 + \frac{1}{1-\gamma_2}\ln A_3A_2\left(1-\theta_2\right) + \frac{\gamma_2}{1-\gamma_2}\left(\ln B_2 + \sum_{n=1}^{N_2}\alpha_{2n}\ln\alpha_{2n}\right)$.

At the beginning of stage 1, farmer chooses profit-maximizing levels of stage 1 inputs, x_1, given initial state of the plot, y_0, and taking into account his anticipated stage 2 and stage 3 inputs demands. Farmer's information set at the beginning of production cycle includes only initial conditions and stage 1 factor prices: $I_1 = \{y_0, w_1\}$. Farmer solves

$$\max_{\{x_{1n}\}_{n=1}^{N_1}} \bar{p}\bar{y}_3\exp\left(\bar{\varepsilon}_3\right) - \sum_{n=1}^{N_1} w_{1n}x_{1n} - E\left[\sum_{n=1}^{N_2} w_{2n}x_{2n} - \sum_{n=1}^{N_3} w_{3n}x_{3n}\right],$$

where expectations are conditional on information set I_1, y_3 is given by equation (1) and is a function of expected stage 2 inputs demands (5) and stage 3 inputs demands (3). Substituting equations (3) and (5), respectively, for expected stage 3 input demands $\{\bar{x}_{3n}\}_{n=1}^{N_3}$ and stage 2 input demands $\{\bar{x}_{2n}\}_{n=1}^{N_2}$, we can express stage 3 production costs, $\sum_{n=1}^{N_3} w_{3n}x_{3n}$, stage 2 production costs,

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

$\sum_{n=1}^{N_2} w_{2n}x_{2n}$, and deterministic stage 2 output, y_2, in terms of deterministic component of stage 1 output, y_1:

$$E\left[\sum_{n=1}^{N_3} w_{3n}x_{3n}\right] = \left(\frac{\theta_3}{1-\theta_3}\right)^{1/\gamma_3} \frac{A_2\theta_2^{1/\gamma_2} y_1 \exp\left(\bar{\varepsilon}_2 + \bar{\varepsilon}_1\right)}{B_3\bar{\lambda}_3 g_3^{1/\gamma_3} g_2^{1/\gamma_2}} \left(g_2+1\right)^{1/\gamma_2} ,$$

$$E\left[\sum_{n=1}^{N_2} w_{2n}x_{2n}\right] = \left(\frac{\theta_2}{1-\theta_2}\right)^{1/\gamma_2} \frac{y_1 \exp\left(\bar{\varepsilon}_1\right)}{B_2\bar{\lambda}_2 g_2^{1/\gamma_2}} \text{ and } \bar{y}_2 = A_2\theta_2^{1/\gamma_2} y_1 \exp\left(\bar{\varepsilon}_1\right)\left(1+\frac{1}{g_2}\right)^{1/\gamma_2},$$

where $g_2 = (1-\theta_2)^{\frac{1}{\gamma_2-1}} \left(A_2B_2\bar{\lambda}_2\bar{P}_3 \exp\left(\bar{\varepsilon}_2\right)\right)^{\frac{\gamma_2}{\gamma_2-1}} - 1$. The first order conditions for $x_{1k}, k = 1, ..., N_1$, are

$$\begin{aligned} A_3\left(\frac{\theta_3\left(1+g_3\right)}{g_3}\right)^{1/\gamma_3} A_2\left(\frac{\theta_2\left(1+g_2\right)}{g_2}\right)^{1/\gamma_2} \frac{\partial y_1}{\partial x_{1k}} \exp\left(\bar{\varepsilon}_1+\bar{\varepsilon}_2+\bar{\varepsilon}_3\right) = \\ \frac{w_{1k}}{\bar{p}} + \left(\frac{\theta_2}{\left(1-\theta_2\right)g_2}\right)^{1/\gamma_2} \frac{\partial y_1}{\partial x_{1n}} \frac{\exp\left(\bar{\varepsilon}_1\right)}{B_2\bar{\lambda}_2\bar{p}} + \\ A_2\left(\frac{\theta_2\left(1+g_2\right)}{g_2}\right)^{1/\gamma_2} \left(\frac{\theta_3}{\left(1-\theta_3\right)g_3}\right)^{1/\gamma_3} \frac{\partial y_1}{\partial x_{1n}} \frac{\exp\left(\bar{\varepsilon}_1+\bar{\varepsilon}_2\right)}{B_3\bar{\lambda}_3\bar{p}}. \end{aligned}$$

The structure of the first order condition once again reflects the feedback between different production stages. For stage 1 inputs, anticipated future marginal cost has two components, one the effect on stage 2 input demands through stage 1 inputs' effects on stage 1 output, and another the effect on stage 3 input demands, through their indirect effect on stage 2 inputs and therefore on stage 2 output. The sum of this anticipated future marginal cost and concurrent marginal cost, or real stage 1 factor price, equals the marginal production of stage 1 input. As in stage 2, stage 1 input's marginal product and future marginal cost depend on expected production shocks in both current and subsequent production stages.

Once again, we use the first order conditions to express all stage 1 inputs in terms of one stage 1 input, say, x_{11}, as $x_{1k} = \frac{\alpha_{1k}w_{11}}{\alpha_{11}w_{1k}}x_{11}$ for all $k \in \{1, ..., N_1\}$. Solving first for the optimal x_{11}, we can then solve for optimal stage 1 input levels:

$$x_{1k} = \left(\frac{\theta_1}{1-\theta_1}\right)^{1/\gamma_1} \frac{\alpha_{1k}y_0}{B_1w_{1k}\lambda_1}\left[\left(1-\theta_1\right)^{\frac{1}{\gamma_1-1}}\left(A_1B_1\bar{\lambda}_1P_2\exp\left(\bar{\varepsilon}_1\right)\right)^{\frac{\gamma_1}{\gamma_1-1}} - 1\right]^{-1/\gamma_1} \tag{8}$$

$\forall k \in \{1, ..., N_1\}$, where $\lambda_1 = \prod_{n=1}^{N_1} \left(\frac{\alpha_{1n}}{w_{1n}}\right)^{\alpha_{1n}}$, $P_2\left(\left\{\bar{\varepsilon}_i, \{\bar{w}_{in}\}_{n=1}^{N_i}\right\}_{i=1}^{2}, \bar{p}\right) = \theta_2^{1/\gamma_2} \bar{p} R_2^{\frac{\gamma_2 - 1}{\gamma_2}}$, and

$$R_2\left(\left\{\bar{\varepsilon}_i, \{\bar{w}_{in}\}_{n=1}^{N_i}\right\}_{i=1}^{2}, \bar{p}\right) =$$
$$\left(A_2\left(\theta_3 R_3^{\gamma_3 - 1}\right)^{1/\gamma_3} \exp\left(\bar{\varepsilon}_2\right)\right)^{\frac{\gamma_2}{\gamma_2 - 1}} - \left((1-\theta_2)\left(B_2 \bar{\lambda}_2 \bar{p}\right)^{\gamma_2}\right)^{\frac{1}{1-\gamma_2}}.$$

Component R_2 captures the net indirect effect of a change in stage 1 input on production processes in stages 1 and 2. This indirect effect comes from the direct positive effect of stage 1 input use on stage 1 output. Consider first this effect on stage 2 production. On one hand, stage 1 output is used as input in stage 2 production, and so higher stage 1 output results in higher stage 2 output. On the other hand, recall from equation (5) that demand for stage 2 inputs increases in stage 1 output. This entails both additional increase in stage 2 output and higher stage 2 production costs. From equation (8), the marginal effect of expected stage 2 production shock on stage 1 input demands, $\partial x_{1k}/\partial \bar{\varepsilon}_2$, is positive, while the marginal effect of expected stage 2 real input prices on stage 1 input demands, $\partial x_{1k}/\partial \left(\bar{w}_{2k}/\bar{p}\right)$, is negative.

Consider next the effect of stage 1 inputs on stage 3 production. This effect has an additional level of indirectness to it, since a whole production stage separates stage 1 inputs from stage 3 production. Stage 1 inputs indirectly affect stage 2 output through their effect on stage 1 output and stage 2 inputs; subsequently, stage 2 output affects stage 3 output directly as an input into stage 3 production process and indirectly though its effect on stage 3 input demands. In other words, the indirect effect of stage 1 inputs on stage 3 output and production costs depends on the indirect effect of stage 1 inputs on stage 2 output. As a consequence of this double indirectness, the effects of stage 3 expected production shocks, $\bar{\varepsilon}_3$, and real factor prices, $\bar{w}_{3k}/\bar{p}$, on stage 1 input demands, x_{1k}, are less straightforward and depend on the sign of $\bar{R}_3$, with production shocks and factor prices having the opposing effect. As in stages 2 and 3, stage 1 input demand is increasing in initial conditions, in expected stage 1 production shocks, and in relative prices of other stage 1 inputs, and decreasing in its own real price.

Rewrite R_2 as

$$R_2\left(\left\{\bar{\varepsilon}_i, \{\bar{w}_{in}\}_{n=1}^{N_i}\right\}_{i=1}^{2}, \bar{p}\right) = \frac{(1-\theta_2)^{\frac{1}{\gamma_2 - 1}}\left(A_2 B_2 \bar{\lambda}_2 \bar{p}\left(\theta_3 R_3^{\gamma_3 - 1}\right)^{1/\gamma_3} \exp\left(\bar{\varepsilon}_2\right)\right)^{\frac{\gamma_2}{\gamma_2 - 1}} - 1}{(1-\theta_2)^{\frac{1}{\gamma_2 - 1}}\left(B_2 \bar{\lambda}_2 \bar{p}\right)^{\frac{\gamma_2}{\gamma_2 - 1}}}. \quad (9)$$

For $\gamma_2 < 0$, it appears that the effect of expected increase in stage 2 output due to higher stage

1 input use is very pronounced, while the effect of higher stage 2 production costs is negligible. Note that $\partial R_2/\partial\gamma_2 < 0$, so the stronger is complementarity between production stages 1 and 2 (that is, the more negative γ_2 is), the more prevalent is the former effect over the latter. Using the approximation $\ln(x-1) \approx \ln x$, we can approximate $\ln R_2$ as

$$\ln R_2 \approx \frac{\gamma_2}{\gamma_2 - 1}\left(\ln A_2 + \frac{1}{\gamma_3}\ln\theta_3 + \frac{\gamma_3 - 1}{\gamma_3}\ln R_3 + \bar{\varepsilon}_2\right)$$

. Applying the log approximation to equation (8) and using approximations[10] of $\ln R_2$ and $\ln R_3$, we can write stage 1 input demands as

$$\ln x_{1k} \approx \ln\alpha_{1k} + C_1 + \ln y_0 - \frac{1}{1-\gamma_1}\ln\frac{w_{1k}}{\bar{p}} + \tag{10}$$
$$+\frac{\gamma_1}{1-\gamma_1}\sum_{n=1,n\neq k}^{N_1}\alpha_{1n}\ln\frac{w_{1k}}{w_{1n}} + \frac{1}{1-\gamma_1}\left(\bar{\varepsilon}_1 + \bar{\varepsilon}_2 + \bar{\varepsilon}_3\right)\ \forall k \in \{1, \dots, N_1\},$$

where the common component of the constant term is $C_1 = \frac{1}{\gamma_3(1-\gamma_1)}\ln\theta_3 + \frac{1}{\gamma_2(1-\gamma_1)}\ln\theta_2 + \frac{1}{\gamma_1}\ln\theta_1 + \frac{1}{1-\gamma_1}\ln A_3A_2A_1(1-\theta_1) + \frac{\gamma_1}{1-\gamma_1}\left(\ln B_1 + \sum_{n=1}^{N_1}\alpha_{1n}\ln\alpha_{1n}\right)$.

Rice cultivation process is described by a system of equations consisting of production function equations (1) and input demand equations for each of three stages (8), (5), and (3). Final output can also be expressed as a cumulative production function (2), and input demands can be approximated by equations (10), (7), and (4).

Because stage 2 and stage 3 input demands depend on realized outputs from previous stages, they depend on realized production shocks from earlier stages: stage 2 input demands in equation (5) depend on realized stage 1 production shocks ε_1, and stage 3 input demands in equation (3) depend on realized stage 1 and stage 2 production shocks ε_1 and ε_2. Final output depends on realizations of production shocks in all three stages. This is expressed explicitly in the composite production function equation (2), and implicitly in the stage-specific production function equation (1) though the dependence of y_i on y_{i-1}.

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

[10] These approximations of $\ln R_2$ and $\ln R_3$ eliminate dependence of stage 1 input demands on expected stage 2 and stage 3 factor prices.

than potentially complements as allowed for with our CES specification. Let stage production functions be

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

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:

$$\ln x_{3k} = \ln \alpha_{3j} + C_3 + \ln y_2 + \varepsilon_2 - \frac{1}{\beta_3} \ln \frac{w_{3k}}{p} + \frac{1}{\beta_3} \sum_{n=1, n\neq k}^{N_3} \alpha_{3n} \ln \frac{w_{3k}}{w_{3n}} + \frac{1}{\beta_3} \bar{\varepsilon}_3$$

$\forall k \in \{1, ..., N_3\}$, 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. First, relative prices of other stage 3 inputs have positive effect on input demand under Cobb-Douglas and negative effect under CES when $\gamma_3 < 0$, that is, when stage 2 intermediate output and stage 3 production activities are complements. The second difference is the magnitude of the coefficient on expected stage 3 production shock, $\bar{\varepsilon}_3$. The coefficient is positive in both cases; however, it is greater than one under Cobb-Douglas and less than one 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.

3 Error Structure

We have three levels of data variation: individual across households, spatial across villages and plots, and temporal across stages and years. Let h index households, k index plots, v 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 years (over t) and across stages (over i). Similarly, there are three levels of potential group error correlation, within a physical plot, within a household, and within a village. Let ε_{khvit} denote the overall production shock for plot k belonging to household h in village v during production stage i in year t, and let ξ_{khvit} 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.

Rainfall Shock

One of the main production shocks for rice cultivation is rainfall. At a given point in time, rainfall 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, as illustrated in figures 1 and 2 for Sisaket province and figures 3 and 4 for Burirum province. Correlations of monthly rainfall between villages[11] 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 spatially intermingled. This enables us to assume perfect spatial correlation of monthly rainfall across all sample plots in the same province as a good approximation. Although monthly rainfall 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 rainfall between plots in a given stage, making stage rainfall plot-specific rather than aggregate. Let ρ_{khvit} denote rainfall shock realized on plot k belonging to household h in village v in stage i in year t.

In terms of serial correlation, generally rainfall does not persist from year to year (Paxson, 1992). Rainfall 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 rainfall for future stages. Let $\bar{\rho}_{khvit}$ denote farmer's rainfall expectation. The difference between realized and expected rainfall, $\rho_{khvit} - \bar{\rho}_{khvit}$, is the unanticipated rainfall shock. Let $\tilde{\rho}_{khvit}$ denote this difference. By construction, $\tilde{\rho}_{khvit}$ has zero mean and is uncorrelated with farmer's rainfall expectation, $\bar{\rho}_{khvit}$.

At any given point in time, the effect of rain on plant development would vary depending on 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 and r_{khvit} denote our measure of plot- and stage-specific rainfall. Rainfall shock can be written as a sum of stage- and plot-specific observed shock and fixed plot-specific unobserved effect, $\rho_{khvit} = r_{khvit} + u_{khv}$. As farmers know the characteristics of their plots, including plot's slope and elevation, u_{khv} is incorporated into farmer's

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

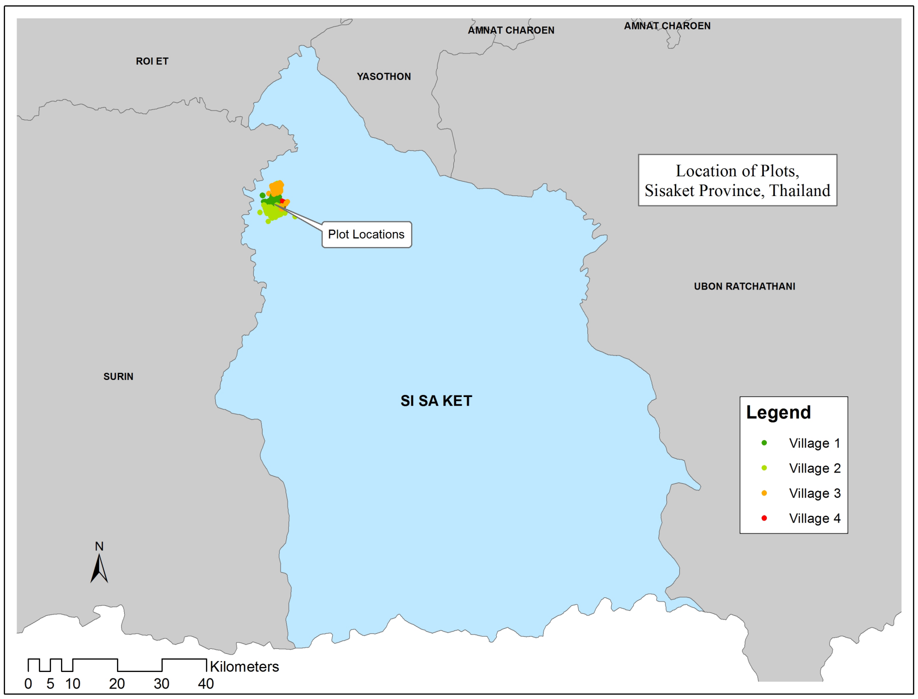

Figure 1: Location of Plots in Four Sample Villages in Sisaket Province

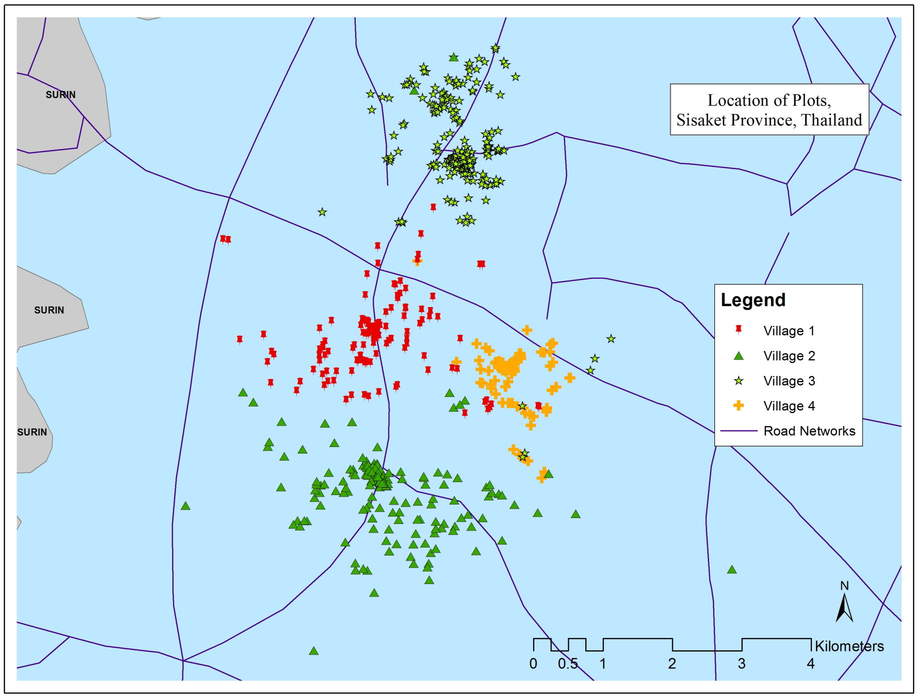

Figure 2: Zoom in on Plot Locations in Four Sample Villages in Sisaket Province

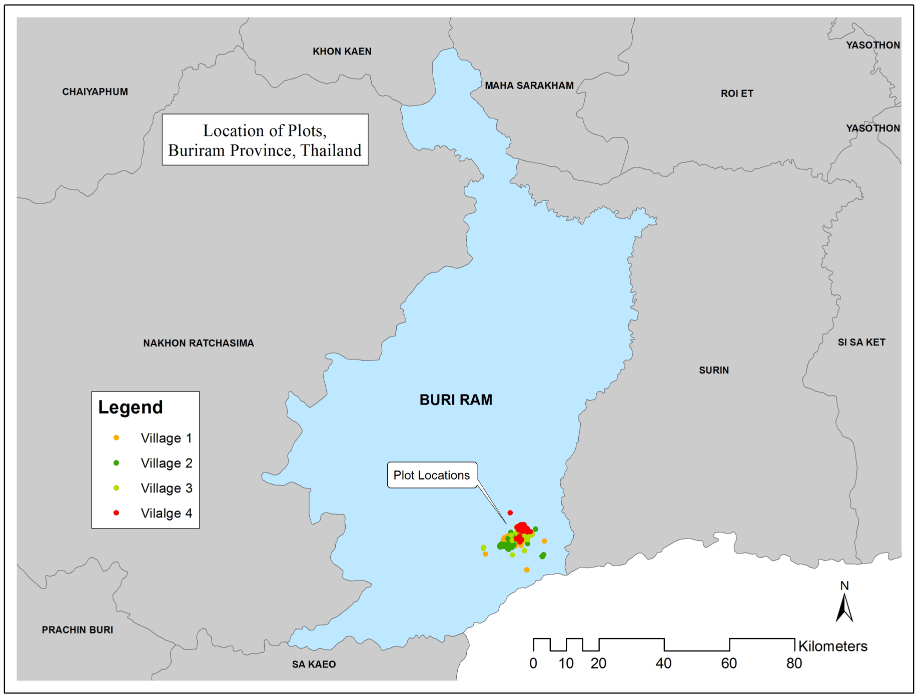

Figure 3: Location of Plots in Four Sample Villages in Burirum Province

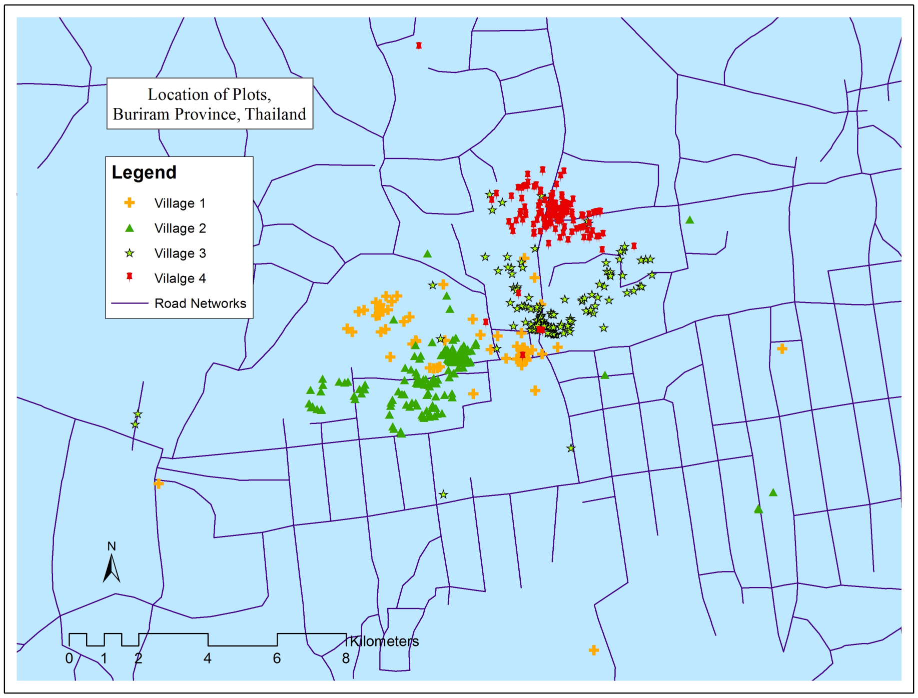

Figure 4: Zoom in on Plot Locations in Four Sample Villages in Burirum Province

rainfall expectation, $\bar{\rho}_{khvit}$. We, on the other hand, do not observe plot's slope and elevation, and as a result our measure of farmer's rainfall expectation, $\bar{r}_{khvit}$, does not incorporate u_{khv}. That is, while for a farmer realized rainfall shock can be decomposed as $\rho_{khvit} = \bar{\rho}_{khvit} + \tilde{\rho}_{khvit}$, we can decompose it only as $\rho_{khvit} = \bar{r}_{khvit} + \tilde{\rho}_{khvit} + u_{khv}$, where $\tilde{\rho}_{khvit}$ is unanticipated by farmer and observed by us rainfall shock[12] and u_{khv} is unobserved by us plot-specific fixed effect. Similarly, our measure of farmer's rainfall expectation is accurate only up to u_{khv}: $\bar{\rho}_{khvit} = \bar{r}_{khvit} + u_{khv}$.

Other Production Shocks and Measurement Errors

At the household level, a number of demographic and economic shocks, such as family member sickness, death, pregnancies, employment lay offs or promotions, are hard to anticipate. Assuming some market frictions, these shocks are difficult to fully smooth right away. Consequently, they potentially can affect production decisions through their impact on household's credit constraint and availability of household labor. These household-level shocks are unlikely to be correlated across years[13], 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 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}.

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

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

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 rainfall 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.

Composite Production Shocks and Error Terms

Composite production shock realized on plot k during stage i is the combination of rainfall 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.

We observe a subset of the production shock, $r_{khvit} = \rho_{khvit} - u_{khv}$, and, as a result, plot-specific fixed effect u_{khv} is part of the composite error term. Similarly, we have an imperfect measure of farmer's prediction of rainfall, $\bar{r}_{khvit} = \bar{\rho}_{khvit} - u_{khv}$. 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 heteroskedastic, serially correlated across stages and years, autocorrelated within a plot, and within a household for a given year. One way to partially account for this error structure in estimation is to use cluster error terms at household level. We have 141 distinct households in the Sisaket province sample, with an average of 2 observations per household per year, since many households cultivate more than one plot in a given year. 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. Existing variation in number of crop-plots[14] per household indicates that ignoring clustering of error terms will have a large effect on estimated standard errors, making inference unreliable. This approach will not fully take care of the unobserved plot-specific fixed effects u_{khv} or of potential bias in cases where household-specific shocks η_{hvit} are correlated across production stages in a given year.

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

4 Data

Our data come from the Townsend Thai Project[15] (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 5 shows location of our sample provinces in Thailand.

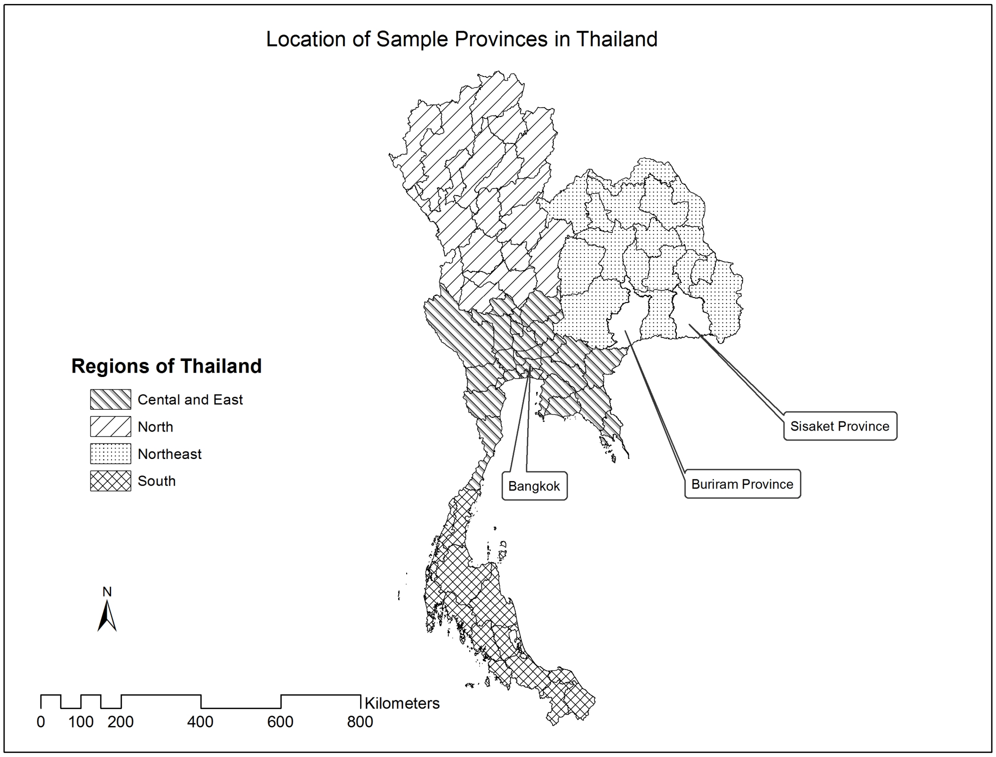

Figure 5: Location of Sample Provinces in Thailand

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[16] with four sample villages was selected at random. Data are collected monthly at a household-plot level, 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

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

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.

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
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
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	1888

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 and chemical composition of water sources.

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 and duration of stages and of the overall production cycle vary across households and plots has several important 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 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. We do not endogenize the planting decision, however, nor the length and timing of stages for each farmer.

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. First three columns of table 2 show the province mode for the starting month of each stage, and last three columns show the province mode of duration of each stage in calendar months. It is clear from table 2 that while stage durations are fairly constant over years, there is pronounced variation in stage timing across years. Little variation in stage duration implies 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 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

Table 2: Timing and Duration of Stages

	Starting month, mode			Length (months), mode		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Burirum province						
2000	7	8	9	1	1.9	2.6
2001	7	8	9	1	2	2.5
2002	6	7	10	1	2.2	2.3
2003	6	7	10	1	2.5	2
2004	6	7	9	1	2.2	2.5
2005	6	7	9	1.1	3	2.4
2006	6	7	9	1	1.9	3.2
2007	6	7	9	1	2.8	2.5
2008	6	7	9	1	2	2.5
2009	5	6	9	1	3.3	2.6
Sisaket province						
1999	6	7	9	1.1	1.8	2.9
2000	7	8	9	1	1.5	2.3
2001	7	8	10	1.1	1.7	2.1
2002	8	9	10	1.1	1.6	2
2003	8	9	10	1	1.6	2.1
2004	7	8	9	1.1	1.7	2.1
2005	8	9	10	1	1.6	1.6
2006	6	7	9	1	2.1	2.8
2007	6	7	10	1	2.1	2.7
2008	7	8	10	1	1.7	2.5
2009	7	8	9	1	1.5	2.8

to take the possibility of adverse rainfall shocks into account when making input decisions. The effect of rain on crop also depends on temperature, as higher temperature can cause faster 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 monthly total of village-level daily rainfall and monthly mean of province-level maximum daily temperature to construct a measure of rainfall shock as a linear combination of rain, square of rain, temperature, and interaction of rain and temperature. We use historic rainfall and temperature data to construct a measure of expected future rainfall and temperature at the beginning of each production stage as a function of monthly indicators. Although rainfall is an aggregate shock, realized and expected rainfall 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. The four villages in each province are located near each other, and distributions of plots for villages overlap, as was illustrated in figures 1 and 2 for Sisaket province and figures 3 and 4 for Burirum province. As a result, plots belonging to households from the same village may actually be further apart than plots belonging to households from different villages. Similarly, plots belonging to the same household may actually be further apart than plots belonging to different households. Thus, whether plots belong to the same village or even the same household 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 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.[17] In terms of section 2 notation, initial condition y_0 is a linear combinations 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.[18] The DSSAT system takes in amounts and timing of application of non-labor pro-

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

[18]DSSAT, or Decisions Support System for Agrotechnology Transfer, has been maintained and supported by the

duction 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 nor 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 crop-specific growth responses measured precisely in laboratory conditions. However, despite the high precision and accuracy of DSSAT crop-growth simulations, the software typically is not able to model certain particular and idiosyncratic environmental stresses that reduce crop growth from the optimal predicted amounts.

The advantage of our economic model of rice production over DSSAT is that economic model takes into account farmers' decisions on timing and labor inputs. Again, the advantage of DSSAT over our economic model is that DSSAT has information on the way plant develops biologically and therefore can trace 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 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 a linear combinations 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 a linear combinations of one DSSAT measure of intermediate output from stage 2 and labor used in stage 2.

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, and

International Consortium for Agricultural Systems Applications (ICASA).

soil quality measures 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.

Table 3: Summary Statistics on Yields, Soil, and Production Inputs

	Burirum province		Sisaket province	
	Mean	St. Dev.	Mean	St. Dev.
Yield (kh/ha)	1,983.75	887.60	1,777.02	803.56
Area (ha)	1.30	0.88	1.12	0.92
CEC (meq/100g)	2.54	1.50	2.32	1.20
Organic matter (%)	0.43	0.25	0.52	0.36
Stage 1 chemical fert. (kg/ha)	40.41	116.82	11.31	41.91
Stage 2 chemical fert. (kg/ha)	113.19	103.72	147.18	142.55
Seeds (kg/ha)	96.48	77.94	35.28	66.05
Seedlings (sets/ha)[a]	281.58	559.19	788.62	703.41
Stage 1 labor (hrs/ha)	74.21	94.86	141.36	139.83
Stage 2 labor (hrs/ha)	23.53	61.21	34.78	68.96
Stage 3 labor (hrs/ha)	232.87	178.83	220.59	157.53

[a]One set contains about 100 seedlings.
***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.

5 Production Function Estimation

To account for endogeneity of input decisions, we estimate the composite production function and input decision rules as a system of simultaneous equations. The system approach to estimation 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. 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 truly 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. However, the system is not recursive because of possible unobserved error components discussed in section 3. We estimate the model with Sisaket province data. We then use Burirum data to evaluate model's accuracy in predicting out of estimation sample.

Before we present the estimates, we look at yield variation present in the data. Table 4 shows variation of yields in each province depending on the initial conditions and the amounts of inputs used in production. For each variable listed in column 1, we split the sample into two groups, one with the value of that variable below 40th percentile for the province and the other above 60th percentile for the province. We then do mean comparison tests of mean yields for these two groups. Column 2 shows mean yield for the above 60th percentile group, column 3 shows mean yield for the below 40th percentile group, and column 4 shows the difference in means between the two groups. Columns 5, 6 and 7 show, respectively, the t statistic, one-sided p-value for null hypothesis that mean yield for above 60th percentile group is greater than mean yield for below 40th percentile group, and one-sided p-value for the reverse null hypothesis of test comparing group one mean yield and group two mean yield.[19] The results of this data summary exercise are both intuitive and informative. All inputs correspond to expected and significant variation in yields: yields are higher on plots with better soil, more fertilizer use in both stages 1 and 2, planted with seedlings rather than seeds, and more labor use in all stages. Higher use of labor in stage 3 - that is, for harvesting - corresponds to particularly large increase in yields. Interestingly, yields are lower on plots with larger cultivated area.

Table 5 performs the same data summary exercise, this time looking at variation of yields across rain and temperature realizations. The structure of table 5 is the same as that of table 4. Stage measures of rain and temperature are plot- and year-specific and depend on stage timing on a given plot in a given year. Rain is measured as mean of daily rain, in millimeters, over each stage. For example, if on a given crop-plot stage 1 took place in June of 2004, then stage 1 rain for that plot is mean of daily rain in June 2004. Temperature is measured as mean of maximum daily temperature, in degrees Celsius, over all months in a given stage. Comparison of mean yields is performed once again between two groups, one with weather variable's measure above 60th percentile for the province, and the other with weather variable's measure below 40th percentile for the province. Several results here are of note. First is the large correspondence between higher stage 1 rain and higher yields in Sisaket province, and lack of it in Burirum province. Second is the pronounced negative correspondence between temperature and yields, present in both provinces. Third is the absence of significant correspondence between stage 2 rainfall and yields. These preliminary results suggest that the effect of rainfall on yields is less straightforward than could be expected. Comparing tables 4 and 5, it is interesting to note that variations in weather realizations correspond to noticeably smaller differences in yields than variations in input amounts.

Table 6 shows estimates of the structural coefficients of the model. First three rows show esti-

[19] The exception are rows "seedlings vs. seeds" and "used chem. fert. in stage 1". These variables are one-zero indicators. Correspondingly, "below" group for these rows has indicator equal to zero (meaning "seeds" and "no chem. fert. in stage 1"), and "above" group has indicator equal to one (meaning "seedlings" and "did use chem. fert. in stage 1").

Table 4: Mean Yield Comparison Tests, by Input Usage and Soil

	Mean yield				p-value	
	> 60th pctile	< 40th pctile	Difference	t stat.	Diff. < 0	Diff. > 0
Burirum province						
Area (ha)	1948	2151	-203	-2.88***	0.002	0.998
CEC	2213	1885	328	4.52***	1	0
Organic matter	2057	2028	28	0.37	0.645	0.355
Seedlings vs. seeds	2309	1928	381	5.09***	1	0
Used chem. fert. in stage 1	2106	1971	135	1.98**	0.976	0.024
Stage 2 chem. fert. (kg/ha)	2197	1872	325	4.29***	1	0
Stage 1 labor (hrs/ha)	2176	1933	243	3.25***	0.999	0.001
Stage 2 labor (hrs/ha)	2072	2007	65	0.88	0.811	0.189
Stage 3 labor (hrs/ha)	2393	1708	685	9.37***	1	0
Sisaket province						
Area (ha)	1706	1967	-261	-6.33***	0	1
CEC	1853	1828	25	0.56	0.712	0.288
Organic matter	1858	1759	99	2.30**	0.989	0.011
Seedlings vs. seeds	1928	1620	308	6.87***	1	0
Used chem. fert. in stage 1	1970	1826	144	2.07**	0.981	0.019
Stage 2 chem. fert. (kg/ha)	2000	1665	335	7.73***	1	0
Stage 1 labor (hrs/ha)	2024	1656	367	7.96***	1	0
Stage 2 labor (hrs/ha)	1939	1758	181	4.03***	1	0
Stage 3 labor (hrs/ha)	2118	1615	503	11.00***	1	0

***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.

Table 5: Mean Yield Comparison Tests, by Stage Rainfall and Temperature

	Mean yield				p-value	
	> 60th pctile	< 40th pctile	Difference	t stat.	Diff. < 0	Diff. > 0
Burirum province						
Stage 1 rain	2052	1993	59	0.80	0.789	0.211
Stage 2 rain	2082	2042	40	0.52	0.700	0.300
Stage 3 rain	2057	2055	2	0.03	0.512	0.488
Stage 1 temperature	1968	2109	-141	-1.92**	0.028	0.972
Stage 2 temperature	2036	2040	-4	-0.06	0.475	0.525
Stage 3 temperature	1964	2109	-145	-1.90**	0.029	0.971
Sisaket province						
Stage 1 rain	1986	1720	266	5.80***	1	0
Stage 2 rain	1805	1857	-52	-1.12	0.131	0.869
Stage 3 rain	1876	1874	2	0.04	0.515	0.485
Stage 1 temperature	1698	1945	-246	-5.63***	0	1
Stage 2 temperature	1794	1879	-85	-1.88**	0.030	0.970
Stage 3 temperature	1815	1873	-58	-1.21	0.114	0.886

***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.

mates of elasticities of substitution, which are significantly different from unity (the Cobb-Douglas case) in all three stages. Note the low elasticity of 0.2 for stage 1, which suggests that initial conditions, such as soil quality, are hard to compensate for with use of inputs. Thus heterogeneity in soil quality will inevitably lead to heterogeneity in yields. Stage 2 elasticity of 0.5 suggests a fair 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. Stage 3 low elasticity of 0.02, which is not significantly different from zero, is also 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. This last result underscores large effect of harvesting labor on yields observed in preliminary data analysis in table 4. Overall, elasticity estimates suggest that stage 2 is the most opportune period for farmer to alter the course of his crop's development and impact expected yields.

Measures of initial condition include cultivated area and soil quality measures. Positive and significant coefficient estimate for organic matter confirms the expectation that soil quality has positive effect on crop cultivation. 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 both stages, labor used in previous stage is also statistically significant as a proxy for previous stage output. These results suggest that, on 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.

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 1, when crop is planted. 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 2 are not significant. Our estimates suggest that elasticity of substitution in stage 2 is highest out of all three

Table 6: Coefficient Estimates of Model Parameters

Variable group	Variable	Coefficient	St. error
Elasticity	Stage 1 elasticity	0.197***	0.040
	Stage 2 elasticity	0.488***	0.055
	Stage 3 elasticity	0.018	0.056
Initial condition	Area (ha)	0.770***	0.019
	CEC (meq/100g)	-0.066	0.044
	Organic matter (%)	0.082***	0.029
Stage 1 output	Number of tillers (DSSAT)	0.474***	0.069
	Leaf weight (kg/ha) (DSSAT)	0.175	0.107
	Stage 1 labor (hrs/ha)	0.300***	0.055
Stage 2 output	Grain filling (DSSAT)	0.405***	0.042
	Stage 2 labor (hrs/ha)	0.043**	0.019
Stage 1 weather	Rainfall (mm)	-0.660**	0.287
	Rainfall squared	0.008***	0.002
	Max daily temperature	-2.621***	0.637
	Rain*temperature	0.155*	0.081
Stage 2 weather	Rainfall (mm)	0.052	0.220
	Rainfall squared	-0.001	0.001
	Max daily temperature	-0.303	0.498
	Rain*temperature	-0.017	0.066
Stage 3 weather	Rainfall (mm)	-0.443	0.312
	Rainfall squared	0.010**	0.004
	Max daily temperature	0.299**	0.130
	Rain*temperature	0.120	0.101
Stage 1 inputs	Labor (hrs/ha)	-0.258***	0.035
	Chem. fert. (kg/ha)	-0.065	0.045
	Seedlings (sets/ha)[a]	0.065**	0.028
	Seeds (kg/ha)	-0.211***	0.038
Stage 2 inputs	Labor (hrs/ha)	2.471***	0.673
	Chem. fert. (kg/ha)	-0.856***	0.180

[a]One set contains about 100 seedlings.
***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.

stages, while effect of weather shock on yields is lowest in stage 2. This result, again, suggests that stage 2 gives most opportunities for the farmer to improve his yields beyond levels that can be expected given his plot's soil quality and concurrent year's aggregate weather realizations.

The last two parts of table 6 show coefficient estimates for stage 1 inputs and stage 2 inputs. As expected given preliminary data exercise in table 4, planting of seedlings results in higher yields and planting of seeds in lower yields. 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.

We next examine whether model prediction errors, measured as actual yield minus predicted, correlate with main socio-economic data. We have data on whether the household is a member of informal social network in the village, on household monthly per capita consumption, and on household's monthly borrowing. In Sisaket province, 87 percent of sample households are part of their village's informal network; the number for Burirum is 80 percent. For network, we use 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. We examine how model prediction error correlates with per capita household consumption and borrowing at different periods of the crop growth cycle, and whether these correlation patterns differ for households that are part of the informal networks and those that are not. We first remove the household fixed effect from model prediction error, to eliminate any potential household-level bias from correlation analysis.

Table 7 shows correlations of model prediction error with borrowing variables. We consider three types of borrowing: any incidence of borrowing, borrowing used to rollover debt, and "bridge" borrowing. Rollover and "bridge" borrowing are the same in nature, in that new loan is acquired to pay off maturing debt (these measures come from Sripakdeevong and Townsend, 2012). The difference is that "bridge" borrowing involves an additional party. Rollover borrowing means making new loan with the same lender to cover maturing debt, and "bridge" borrowing means taking out new debt with a third party to pay off maturing debt with original lender. "Bridge" loans constitute a more extreme version of borrowing and can be thought of as most indicative of credit constraints. We aggregate household's monthly borrowing data over five stages. Stages 1 through 3 correspond in time and order to the three production stages. Stage 0 covers three months before beginning of crop cultivation. For example, if for a given crop-plot production stage 1 started in June, stage 0 for this household is March through May of that year. Stage 4 covers three months after the end of crop cultivation activities. Columns 1 and 2 show correlations for households not in the village network. Column 1 shows correlations with fractions of rollover and "bridge" borrowing out of total amount borrowed over a given time period. Column 2 shows correlations with zero-one indicators for overall borrowing, rollover, and "bridge" borrowing. Columns 3 and 4 show correlations with, respectively, borrowing fractions and indicators, for household in the village network.

For no-network households, there is no correlation of model prediction error with borrowing variables. One stark exception is large positive correlation of 0.2 with overall borrowing indicator in stage 3. It implies that no-network households that have higher yields than model predicts are quite likely to borrow during harvest stage. One possible interpretation is that households that experience large potential yields are in need of a lot of labor for harvesting. If these households are not in the village network, they cannot rely on the informal network to provide the necessary resources either directly in form of exchange labor or indirectly in form of financial resources to use for hiring labor. Their only recourse is to borrow money. If they do borrow, they are able to hire labor for harvesting, and their potential high yields are realized. Note that we observe a significant and positive correlation of model prediction error with borrowing indicator in stage 3 for households in the village network (column 4); however, the magnitude of this correlation for network households is three times lower than that for no-network households. This suggests that both network and no-network households experience same resource demands in the harvesting stage, however, network households are able to satisfy these demands to a large extent through the informal network, while no-network households have no such alternative to borrowing. For households in village network, there is also significant negative, albeit small in magnitude, correlation of model prediction error with overall and rollover borrowing in stage 1. It implies that network households that have lower yields than model predicts are likely to rollover loans during planting stage. In other words, these households were relatively credit constrained during planting stage, and they subsequently had lower yields. These results seem to indicate that informal networks provide an important but incomplete mechanism for alleviating credit constraints.

Table 8 shows correlations of model prediction error with per capita household consumption. Table 8 has the same structure as table 7. Consumption is aggregated over five stages, stages 1 through 3 corresponding to the three crop cultivation stages, and stages 0 and 4 corresponding, respectively, to three months before the start and after the end of cultivation process. Column 1 shows correlations for households not in the village network, and column 2 shows correlations for household in the village network. We consider three measures of household consumption: consumption of household's own production, consumption expenditure by household on goods not produced by the household, and total household consumption which is the sum of the two. All three consumption measures are per capita. There is strong positive correlation of model prediction error

Table 7: Correlation of Model Prediction Error with Borrowing, by Stage

		No network		In network	
	Statistic	Fraction	Indicator	Fraction	Indicator
Stage 0 (three months before planting)					
Borrowing	corr.		0.0762		0.0086
	p-value		0.2909		0.7497
Rollover	corr.	-0.0221	-0.0221	-0.0071	0.0152
	p-value	0.7598	0.7598	0.7939	0.5735
"Bridge"	corr.	0.0128	0.0128	-0.0024	-0.0127
	p-value	0.8599	0.8599	0.9300	0.6392
Stage 1 (planting)					
Borrowing	corr.		-0.088		-0.0729*
	p-value		0.2212		0.0068
Rollover	corr.	-0.0053	-0.0053	-0.049	-0.0536*
	p-value	0.9415	0.9415	0.0694	0.0468
"Bridge"	corr.	-0.0053	-0.0053	-0.0414	-0.0414
	p-value	0.9415	0.9415	0.1250	0.1250
Stage 2 (intermediate crop growth stage)					
Borrowing	corr.		0.0254	-0.0209	
	p-value		0.7248		0.4399
Rollover	corr.	-0.0484	-0.0569	0.0111	0.0103
	p-value	0.5016	0.4291	0.6798	0.7029
"Bridge"	corr.	-0.0558	-0.1004	-0.0103	-0.0088
	p-value	0.4386	0.1626	0.7015	0.7457
Stage 3 (harvesting)					
Borrowing	corr.		0.1970*		0.0660*
	p-value		0.0058		0.0144
Rollover	corr.	-0.0012	0.0055	0.0134	0.0212
	p-value	0.9866	0.9387	0.6184	0.4322
"Bridge"	corr.	0.0465	0.0748	0.0101	0.0297
	p-value	0.5189	0.2986	0.7089	0.2719
Stage 4 (three months after end of harvesting)					
Borrowing	corr.		0.0901		0.0114
	p-value		0.2227		0.6832
Rollover	corr.	0.0292	0.0151	-0.0621*	-0.0457
	p-value	0.6932	0.8380	0.0261	0.1019
"Bridge"	corr.	0.0559	-0.0080	-0.0499	-0.0390
	p-value	0.4497	0.9140	0.0736	0.1623

* denotes significance at 5% level.

with consumption for no-network households in four out of five year segments, with correlations approximately around 0.14-0.19. In contrast, for network household there is significant but much lower in magnitude (up to four times lower) correlation for stage 0, significant negative correlation of -0.05 for stage 1, and no significant correlation for remaining year segments. It seems clear that for no-network households consumption correlates positively with yields, while for network households this correlation is much weaker and present only in pre-planting stage. One interpretation is that network households are able to smooth consumption through the informal network, while no-network households have no alternative smoothing mechanism. Ability of informal networks to smooth consumption is weakest at pre-planting stage, when expected future yields are most removed in time.

Table 9 also shows correlations of model prediction error with per capita household consumption; here, consumption is aggregated by calendar quarters rather than by stages. The main difference between tables 8 and 9 is that the former takes into account variation in timing of crop cultivation activities between households and across years, while the latter aggregates household consumption over the same time periods for all households. The latter table also includes correlation of model prediction error with aggregate quarterly village consumption. Columns 1 and 2 of table 9 show correlation with per capita household consumption, respectively for no-network and network households. Columns 3 and 4 show correlation of model prediction error with aggregate village consumption, respectively for no-network and network households. These results confirm results from table 8. Per capita household consumption correlates positively and significantly with model prediction error for no-network households, with correlation coefficients ranging from 0.15 to 0.22. In contrast, there is no significant correlation for network households. Moreover, aggregate village consumption does not correlate with model prediction error for no-network households, but correlates significantly for network households. These correlation results provide evidence that informal networks are a substantial consumption smoothing mechanism for rice cultivating households. We return to this in section 7, at the end.

There is also a correlation of model prediction error with yields, of approximately 0.11. We surmise that household know something about the evolution of the state of the plot which we do not see despite the richness and detail of our data. But our data are too limited to incorporate this new aspect into the model without putting an enormous amount of structure on the problem.

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 rainfall shock for a given crop-plot. We perform the following exercise. We

Table 8: Correlation of Model Prediction Error with Per Capita Hhd Consumption, by Stage

	Statistic	No network	In network
Stage 0 (three months before planting)			
Total cons.	corr.	0.1888*	0.0528*
	p-value	0.0041	0.0434
Own cons.	corr.	0.0369	-0.0193
	p-value	0.5774	0.4622
Cons. expenditure	corr.	0.1847*	0.0554*
	p-value	0.005	0.0344
Stage 1 (planting)			
Total cons.	corr.	0.0317	-0.0102
	p-value	0.6321	0.695
Own cons.	corr.	0.1174	-0.0517*
	p-value	0.0749	0.0472
Cons. expenditure	corr.	-0.0058	-0.0024
	p-value	0.9298	0.9253
Stage 2 (intermediate crop growth stage)			
Total cons.	corr.	0.1643*	-0.0084
	p-value	0.0124	0.7474
Own cons.	corr.	0.0925	-0.0137
	p-value	0.1613	0.6006
Cons. expenditure	corr.	0.1567*	-0.0059
	p-value	0.0172	0.8215
Stage 3 (harvesting)			
Total cons.	corr.	0.1077	-0.0203
	p-value	0.1026	0.4367
Own cons.	corr.	0.1355*	-0.0269
	p-value	0.0397	0.3019
Cons. expenditure	corr.	0.0669	-0.0182
	p-value	0.3112	0.4866
Stage 4 (three months after end of harvesting)			
Total cons.	corr.	-0.0513	-0.0303
	p-value	0.4377	0.2477
Own cons.	corr.	0.0416	0.0116
	p-value	0.529	0.6573
Cons. expenditure	corr.	-0.0614	-0.0325
	p-value	0.3529	0.2145

* denotes significance at 5% level.

Table 9: Correlation of Model Prediction Error with Consumption, by Quarter

	Statistic	Hhd per capita No network	Hhd per capita In network	Village aggregate No network	Village aggregate In network
1st quarter					
Total cons.	corr.	0.0521	-0.0013	0.0355	0.0483
	p-value	0.4306	0.9615	0.5918	0.0638
Own cons.	corr.	0.0864	-0.0126	-0.0451	-0.0109
	p-value	0.1908	0.6306	0.4951	0.6767
Cons. expenditure	corr.	0.043	0.0003	0.0537	0.0583*
	p-value	0.5159	0.9923	0.4167	0.0255
2nd quarter					
Total cons.	corr.	0.2210*	0.0333	0.067	0.0517*
	p-value	0.0007	0.2021	0.3106	0.0473
Own cons.	corr.	0.0833	-0.0508	-0.0097	-0.0095
	p-value	0.2071	0.052	0.8833	0.7153
Cons. expenditure	corr.	0.2074*	0.0379	0.0754	0.0607*
	p-value	0.0015	0.1466	0.2538	0.0199
3rd quarter					
Total cons.	corr.	0.0867	-0.0328	-0.0355	-0.0538*
	p-value	0.1892	0.2088	0.5918	0.0392
Own cons.	corr.	0.0908	-0.0303	-0.0285	-0.0194
	p-value	0.169	0.2461	0.6665	0.4567
Cons. expenditure	corr.	0.0567	-0.0256	-0.0249	-0.0502
	p-value	0.3909	0.3268	0.7067	0.0543
4th quarter					
Total cons.	corr.	0.1560*	-0.0178	0.0862	-0.0434
	p-value	0.0177	0.4963	0.1918	0.0958
Own cons.	corr.	0.0902	-0.0508	-0.0036	-0.0159
	p-value	0.1721	0.0515	0.9561	0.5431
Cons. expenditure	corr.	0.1486*	-0.0095	0.1093	-0.049
	p-value	0.0239	0.7146	0.0974	0.0605

* denotes significance at 5% level.

first plot kernel density of actual sample yields, after taking out the household fixed effect. This is the black solid line in figure 6 for Sisaket province and in figure 7 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, take out the household fixed effects, and plot the kernel density of these predictions. This is the green long dash (_ _ _) line in figures 6 and 7. 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.

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 red dotted (...) line in figures 6 and 7. It captures variation in yields due to difference in timing, rainfall realizations, and soil quality. We see that it explains noticeably 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 blue dash-dot (_..._) line in figures 6 and 7. 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 that while factors largely out of farmer's control, such as rainfall and soil quality, are important sources of yield variation, other factors that are directly controlled by the farmer, namely, input amounts used, contribute significantly to yield heterogeneity. An analysis of variation in yields that does not account for farmers' response to production conditions through their choice of input amounts is likely to significantly underestimate the resulting variation in yields.

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, and take out the household fixed effect. 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. The orange short dash (- - -) line in figures 6 and 7 corresponds to our version of "weather index". We see that it explains substantially less variation than even our most constrained model-based prediction, which allowed only for rainfall and timing variation, both in Sisaket and especially in Burirum province. This last result illustrates that

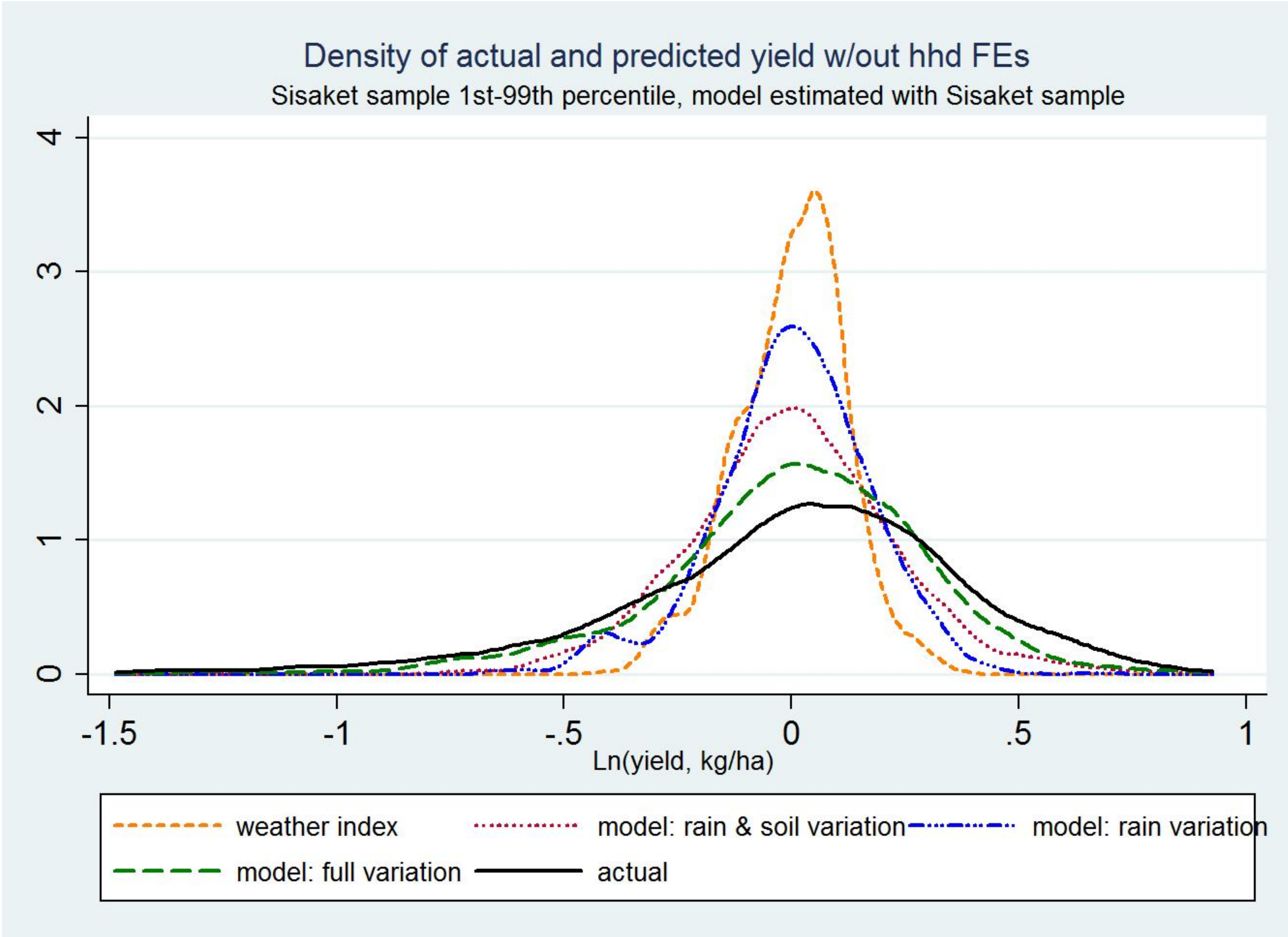

Figure 6: Kernel Density of Actual and Predicted Yields, Sisaket Sample

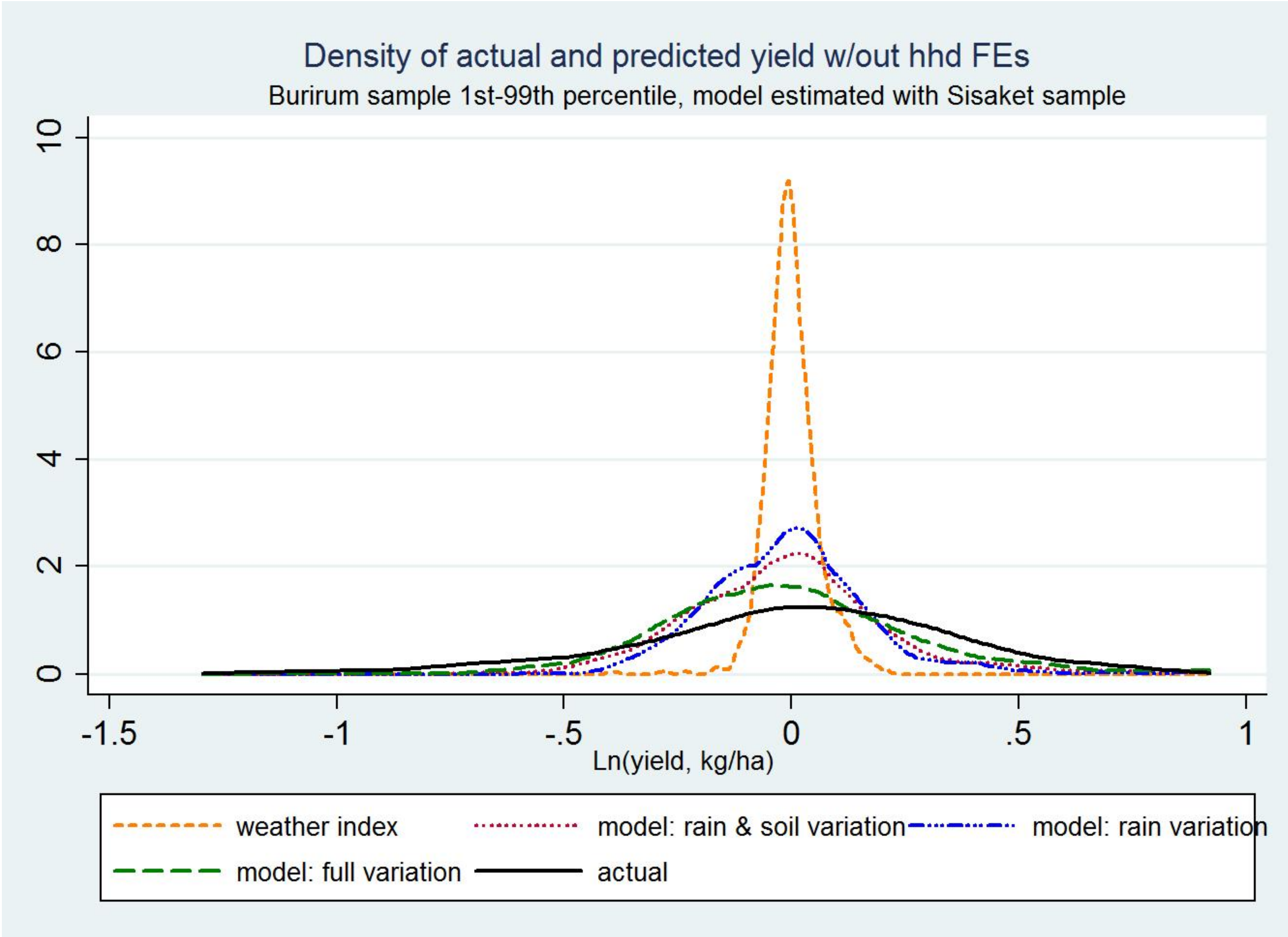

Figure 7: Kernel Density of Actual and Predicted Yields, Burirum Sample

even when focusing only on one source of yield variation, namely, rainfall and timing, simple linear approximation captures 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 household-specific production choices and plot-specific soil quality. An important and interesting question is how these sources of 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 weather generation model.[20]

We simulated future "synthetic" weather from the widely used WGEN weather simulation model (Richardson, 1981). The WGEN weather generation 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 including 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. We generated 100 stochastic weather year realizations based directly on the statistics computed for the historical, 1972-2002 observed weather data. We refer to these weather realizations as describing a "neutral" scenario, assuming that future climate will be a direct, linear extension of the late 20th century. 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 effect, we selected a subsample of 83 plots in Sisaket province[21]. For each of these 83 plots, we used estimated model together with DSSAT to predict yield under each weather realization in each of the three considered climate scenarios.

We first used actual data to estimate farmers' timing decision. As discussed in section 4, in

[20] 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 weahter for the area, and such data are available only on province level.

[21] We could not perform the analysis on the whole sample due to computational constraints.

our data timing choice effectively reduces to the choice of starting month for the crop cultivation process. We postulated that farmer's choice of when to start crop cultivation depends on cultivated area and on observed rain and temperature realizations at the beginning of the current year and during previous crop growth cycle. Using the whole sample and actual weather realizations, we regressed actual calendar month corresponding to beginning of stage 1 on plot's cultivated area, rain and temperature in January through March of current calendar year, and rain and temperature in April through December of the previous calendar year. We then used the resulting equation to predict 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 10 shows the distribution of starting month for crop cultivation process in actual data, and predicted starting month for each of the three climate scenarios. It appears that cultivation activities started on average one month later under high emissions scenario compared to neutral climate scenario, with no substantial difference between low emissions and neutral climates. Using these timing predictions, we constructed plot-specific timing of stages, assuming that stage 1 lasts one month, stage 2 lasts two months, and stage 3 lasts two and a half months. 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, and indicators for calendar month. Using actual data, we also estimated factor prices as functions of calendar month indicators, and interaction of month indicators with monthly rain and temperature. We then used these estimates to predict output and factor prices with synthetic weather data for each of the three climate scenarios we consider.

Table 10: Distribution of First Month of Production Process

	Number of obs. with a given month out of total (percent)			
Calendar month	Actual data	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	

We next alternated between the model and DSSAT, first predicting stage 1 inputs with the model, then using these predictions to simulate DSSAT with synthetic weather, then using these DSSAT predictions as measures of intermediate stage 1 output to predict stage 2 inputs with the model, then using these predictions to simulate DSSAT with synthetic weather, and then using these DSSAT predictions as measures of intermediate stage 2 output to predict stage 3 inputs with the model. Having now plot-, stage-, and weather realization-specific predictions of inputs and weather shocks, we use the model one final time to predict yields. We repeated these steps for our subsample of 83 plots for 99 weather realizations under each of the three climate scenarios. As a result, we had 99 yield realizations per plot per climate scenario. 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 for different weather realizations.

We first provide a summary of the two alternative climate changes that we consider - the high and low emission scenarios. Table 11 uses 100 weather realizations generated by WGEN for each climate scenario to compare high and low emission climate scenarios to the neutral scenario.

Panel A of table 11 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 address 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 mean test on the equality of daily precipitation under neutral and high-emissions climates. In the same manner, columns six through eight address shift from neutral to low emissions climate, and columns nine through 11 address 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 high emissions climate is about 40 percent higher. Daily precipitation increases throughout the year under low emissions climate. On the other hand, under 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 climate change brings moderate increase in temperature and more rain, while high emissions climate bodes both higher increase in temperature and less rain for rice cultivation.

It would be intuitive to expect decrease in yields under both climate change scenarios relative to neutral climate, and to expect larger decrease in yields under high emissions scenario than under low emissions scenario. We compared predicted yields under the three climate scenarios for each of the 83 plots. While the former intuitive conjecture appears to hold, the latter does not apply for large part of our sample. We find that while for some plots yields are substantially lower with high emissions climate compared to both neutral and low emissions climates, for other plots there is no substantial difference between yield distributions under high and low emissions climates. Figures 8

Table 11: 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

and 9 illustrate this point. 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. The difference in the effect of climate change on yield distribution for these two plots is starking. While there is progressive decrease in mean yields from neutral to low emissions, and from low emissions to high emissions climates for the plot in figure 8, there is no decrease in mean yields between low and high emissions climates for plot in figure 9. The difference between the effects of the same climate change for the two plots doesn't stop there. The way the shape of yield distribution changes between climates is noticeably deferent for the two plots. For both plots, yield distributions under neutral and low emissions climates are somewhat centered, and become skewed to the right under high emissions climate. However, this skewness to the right is barely noticeable for the plot in figure 9, while being very pronounced for the plot in figure 8. The patterns displayed in both figures 8 and 9 are typical for large parts of our sample, with some in-between cases. It is clear that heterogeneity in yields for plots experiencing common aggregate weather shock that is present in actual data is going to persist under climate change.

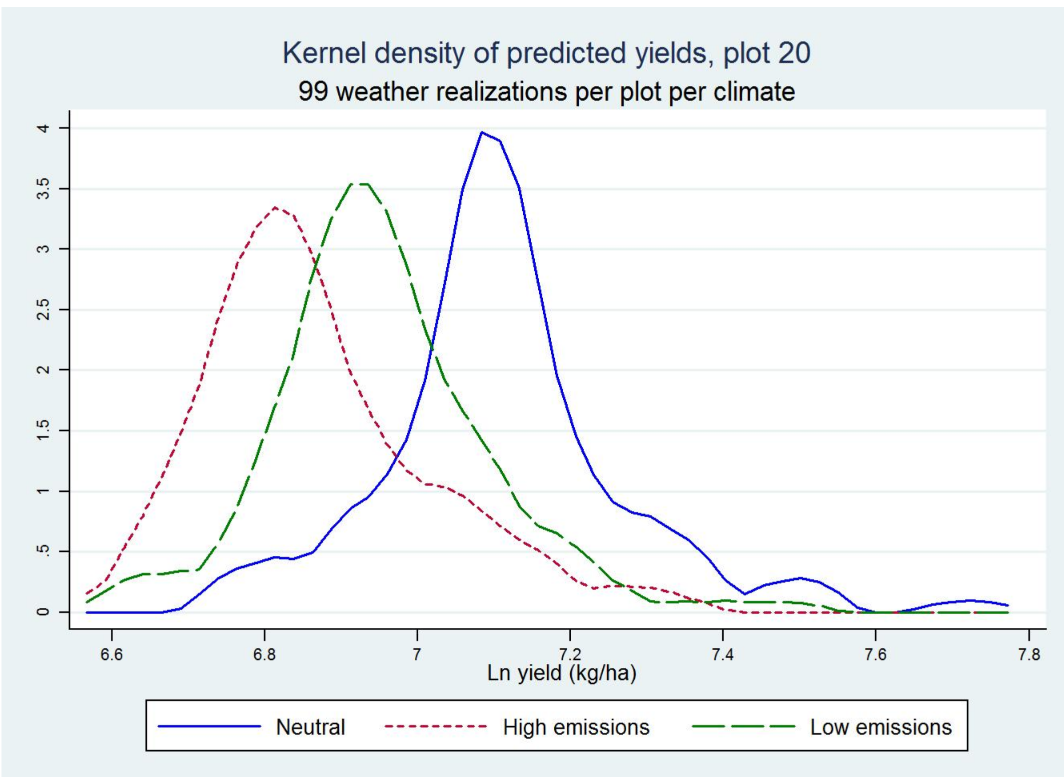

Figure 8: Kernel Density of Predicted Yields, by Climate Scenario

Figure 10 provides additional evidence of heterogeneity in yields under a given climate sce-

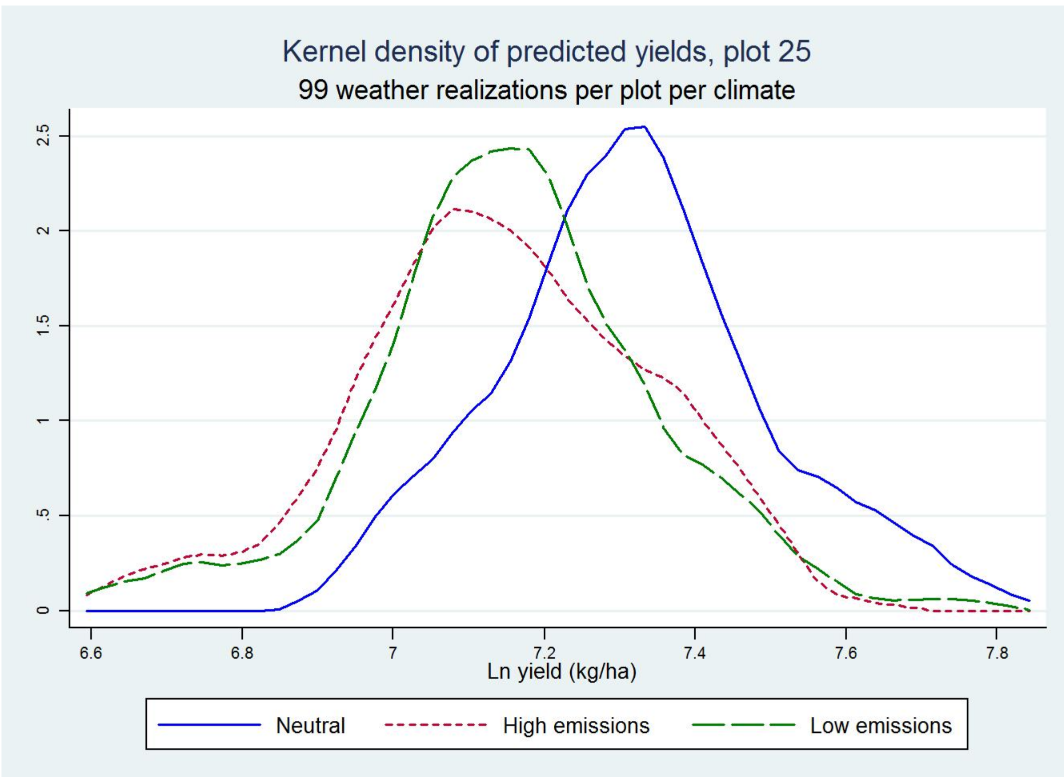

Figure 9: Kernel Density of Predicted Yields, by Climate Scenario

nario. It shows kernel densities of yields under neutral climate for three plots in our sample. Faced with the same aggregate weather shocks, these three plots from the same geographical area experience completely different yield distributions. Clearly the mean yields are different across different plots, though this may be due in part to plot specific fixed effect. Beyond that, one can see the plot in solid lines has more mass in both tails, and comparing the other two, one has more mass at least in the right tail.

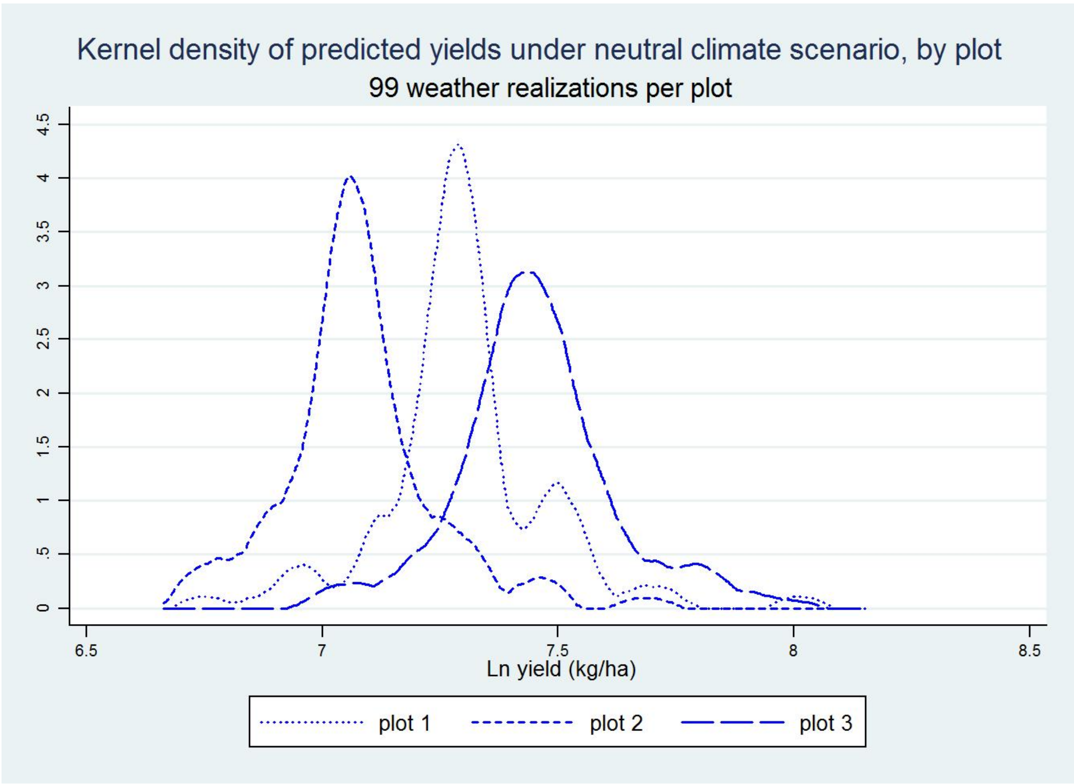

Figure 10: Kernel Density of Predicted Yields under Neutral Climate Scenario, by Plot

7 Specific Comments on Insurance

The meteorological variables we used in figures 6 and 7 are often used in the design of rainfall insurance products to determine how yields vary with rain/temp, or more specifically to determine stress thresholds (typically, these vary with crop and agricultural zone). The appeal of these risk products is that they are based on exogenous variables over which an individual farmer has no control, seemingly mitigate the well know moral hazard problems of direct crop insurance (Skees, Hazell, and Miranda, 2012).

If we were to generalize the model of this paper to allow explicit consideration of insurance, in future work, we would not take households as firms to be risk neutral, as a primitive. A deeper premise/rationalization of the current "risk neutral" specification, though, is that there is an as-if-equivalent-complete set of markets and institutions with a village, or better put with a tambon (in particular the four villages which constitute for us a provincial sample). Thus up to but not including aggregate shocks, production decisions separate from household decisions. That is, for within tambon idiosyncratic shocks, it is as if the crop produced and sold forward at state contingent input and output prices, with enough variation across idiosyncratic shocks, or the way seemingly aggregate shocks impact individual plots and households, that prices are equal to probabilities. That is, as in the standard CAPM all idiosyncratic risks are pooled to have zero mean and there is no risk premium. There is evidence in the papers of Bonhomme, Chiappori, Townsend, and Yamada (2012), Chiappori, Samphatharak, Schulhofer-Wohl, and Townsend (2012), and Kinnan and Townsend (2012) that there is something close to full insurance in consumption and labor supply, especially for households in kinship networks, controlling for aggregate shocks and evidence from Samphantharak and Townsend (2010) that the CAPM does remarkable well in pricing aggregate risk. As for what constitutes an aggregate shock, this is not a seemingly common rainfall shock itself, but rather a move in aggregate consumption, something which the tambon itself cannot insure. That part of risk should be included in consumption based stochastic discount factors, unlike the current model which uses expected outside market prices.

As described in the paper, we analyzed the correlation of predictions errors of our current model with consumption and other variables, to see if some of our maintained assumptions are a reasonable approximation and where the model will need to be improved. Here the results are somewhat reassuring especially for the collections of households in a village that are part of a kinship networks. Relative to households not in networks, household specific consumption is far less correlated with prediction errors, if at all. Further for those in networks it seems that the correlation is with aggregate consumption. For those without networks the correlation with own household consumption is substantial, and as one might predict for households with limited financial markets, there is no pattern with the aggregation consumption (and no reason to be in risk sharing since the null is that such households do not share risk). One reading of the correlation with debt and roll over loans provides a similar story. Household out of networks might be credit constrained, while those in networks less so.

The current model thus offers some possible lessons for the design of rainfall insurance products, assuming that the heterogeneity results would be robust to a consumption based utility model. This discussion here begins as if households were in financial autarky or faced limited credit and insurance markets. Again, this is most relevant for households not in a kinship network. We then move the discussion toward insurance at the tambon level for those in internal risk sharing networks.

Even with the advantage of knowing stages of operation, the simple rainfall index model is designed around the median or modal household in the sample, missing the large mass of household who do better or worse, not just on average but with different impact of weather (i.e., each with his/her own histogram). The same is true for prediction for the out of sample Burirum projection in figure 7.

Thus, extrapolating even further out of sample to new provinces with the Agricultural Census, using rainfall and temperature data, begs the question of variables that are available and how well one could do. Data on area planted and fertilizer use is helpful, and this is available in Thailand at the level of "holder" in a cross section. But as model comparisons indicate, these predictions are like too peaked. And we cannot even control for household fixed effects without a reasonable panel.

Another point about insurance and our no-network households. First, labor and other costly inputs determine output, not just rain, and if households respond to rain in a given stage with endogenously chosen labor, then there is a human response. Unlike Goodwin and Mahul (2004), Kapphan (2012), Mahul (2001), and Musshoff, Odening, and Xu (2009), one cannot simply regress yields on weather and expect that part to remain fixed when part of the yield variation due to rain is introduced in earlier stages. According to our current model, households change inputs.[22]

Again, insurance products have to be designed for each farmer, or at least for groups of farmers with identical technologies that experience weather shocks in the same way. One product does not fit all. Indeed consistent with the mixed experience on take up, it may be this is the real "basis" risk which explains limited take up. In other words it is not that the product uses measurements from some distant rain gauge. Rather the problem is the heterogeneous impact of a “common” shock within the village.

With risk sharing networks it’s tempting to jump to the conclusion that one should aggregate yields and use simple index products based on the correct mean. However, this intuition is not correct. What needs to be insured is the variation on individual crop plots (or variation on groups of crop plots which are otherwise identical). It would be as if there were a representative household farming all of the plots which has an objective the maximization of expected utility. Then one can aggregate consumption and aggregate labor supply in terms of utility consequences and the modeling of choices. But one should not aggregate up over technologies which are distinct due to

[22]Likewise a well design insurance product should take into account variation in input and state of the crop plot at each stage, as final yields are not a simple additive function of all the variation. To the extent that some insurance product vary payoff by rain by stage, with multiple stages, this we agree that is a move in the right direction.

heterogeneity in soil and so on. That is, it would be a mistake to use average soil quality since that is an input into few if any production technology. The representative household assigns its "household" labor and purchases input to various distinct technologies. In short one still needs good approximations to technologies, to take into account production heterogeneity which is apparent in this paper.

The discussion on insurance against climate change draws on the above discussion. For a given, new climate, the demand for insurance will be a function of heterogeneity in the distribution of yields, as above. Insurance bought ex ante against future climate change has to do with the direction of shift of distributions. Again there is heterogeneity in at least the magnitude of downward shifts, with some plots (household) experiencing a more severe impact from high emissions than others who experience relatively little.

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

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.

GOODWIN, B., AND O. MAHUL (2004): "Risk Modeling Concepts Relating to the Design and Rating of Agricultural Insurance Constracts," *World Bank Policy Research, Working Paper*, (3392).

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.

KAPPHAN, I. (2012): "Weather Risk Management in Light of Climate Change Using Financial Derivatives," Doctoral dissertation, ETN Zurich.

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.

MAHUL, O. (2001): "Optimal Insurance against Climatic Experience," *American Journal of Agricultural Economics*, 85(3), 594–603.

MUSSHOFF, O., M. ODENING, AND W. XU (2009): "Management of Climate Risks in Agriiculture - Will Weather Derivatives Permeate?," *Applied Economics*, 1, 1–11.

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.

SAMPHANTHARAK, K., AND R. M. TOWNSEND (2010): "Risk and Return in Village Economies," working paper.

SKEES, J., P. HAZELL, AND M. MIRANDA (2012): "New Approaches to Pubic/Private Crop Yield Insurance," EPTD Discussion Paper, No. 55, International Food Policy Research Institute, Washington, D.C.

SRIPAKDEEVONG, P., AND R. M. TOWNSEND (2012): "Bridge Loans as a Risk Sharing Mechanism," work in progress.