THE UNIVERSITY OF CHICAGO

EMPIRICAL ESSAYS IN DEVELOPMENT ECONOMICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

BY
KAMILYA TAZHIBAYEVA

CHICAGO, ILLINOIS
MARCH 2011

UMI Number: 3445054

DEDICATION

To my parents.

Contents

List of Tables

List of Figures

Acknowledgements

I want to thank my advisors - professors Robert M. Townsend, Robert E. Lucas, Jr., and Kevin M. Murphy - to whom I'm highly indebted, for their continuous support, encouragement, and honest criticism; my parents and sister for their unconditional love and care; my friends for sharing with me the various ups and downs of graduate student's life. I am very grateful for your presence in and influence on my life.

Abstract

This dissertation consists of two essays on decision making of individuals in developing economies with regards to earning opportunities. The common underlying theme is empirical analysis of processes which have large potential to affect individual earnings in developing economies: human capital accumulation (Chapter 1) and production adjustments by traditional farmers to potential climate change (Chapter 2).

Chapter 1 estimates individual's schooling decisions for senior high and college attendance as well as returns to different education levels with data on Indonesia. I find significant positive returns to vocational senior high and college education, as well as evidence of partial self-selection into different schooling choices based on unobserved personal characteristics. While effect of parents' education on individual schooling decision disappears by the time of senior high graduation, the opposite is true for the effect of family's finances.

Chapter 2 (joint with Professor Robert Townsend from MIT and John Felkner from National Opinion Research Center) models production behavior of rice growing farmers in Thailand and accesses their ability to adjust to different climate change scenarios. We specify a three-stage production function for rice cultivation which incorporates the sequential nature of both production shocks realizations, including rainfall, and input choices which are based on sequentially updated information sets of history of realized shocks and observed changes in crop growth. We integrate our economic model of rice production with soil science crop growth modeling, weather simulators, and global climate change models. We consider two alternative climate change scenarios for Southeast Asia. Comparison of yield changes predicted by the soil science model, which does not account for adjustments in input usage, with economic model predictions demonstrates the extent of farmers' ability to mitigate adverse effects of climate change.

Chapter 1

Schooling Choices and Returns to Schooling in Indonesia

1.1 Introduction

Beyond primary schooling, rates of school attendance are low in many developing countries. This is true not only in poorest economies, but also in lower middle income countries, where average enrollment rate for upper secondary schools was 34.19 percent in 1990-1999 and 42.12 percent in 2000-2009.[1] In other words, approximately two-thirds of 26-35 year olds[2] in lower middle income countries such as Indonesia, Thailand, China and India have not attended upper secondary school. For comparison, the corresponding number for the U.S. is 13 percent.

Tertiary enrollment rates are too low even for observed low upper secondary attendance. Table 1.1 demonstrates that in lower middle income economies, of those who attended upper secondary school in the last two decades, one-third continued to post-secondary schooling.

[1] Source: World Bank Education Statistics. Summary statistics on gross enrollment rates for upper secondary and tertiary education as well as levels of per capita GDP for several countries and for countries grouped by income are provided in table 1.1.

[2] Assuming upper secondary school starts at age 15.

The corresponding proportion is 88 percent in the U.S. and 61 percent for high income countries.

Table 1.1: Gross Enrollment Rates and Per Capita GDP

Country	Years[b]	Per capita GDP 2000 US$	Per capita GDP PPP 2005	Gross enrollment rate[a] Upper secondary	Gross enrollment rate[a] Tertiary
Indonesia	1990-1999	767	2,602	n.a.[c]	10.65
	2000-2009	945	3,204	49.45	17.33
Paraguay	1990-1999	1,432	4,106	41.12	10.55
	2000-2009	1,376	3,945	52.84	23.28
Philippines	1990-1999	901	2,385	63.37	26.76
	2000-2009	1,094	2,897	67.76	28.73
Thailand	1990-1999	1,792	5,070	n.a.[c]	22.71
	2000-2009	2,320	6,563	56.36	42.87
United States	1990-1999	30,265	34,145	87.08	75.02
	2000-2009	36,859	41,584	86.82	78.24
Low income[d]	1990-1999	239	729	19.76	3.58
	2000-2009	287	896	22.74	4.91
Lower middle income[d]	1990-1999	601	1,935	34.19	10.25
	2000-2009	1,030	3,237	42.12	15.14
Upper middle income[d]	1990-1999	3,090	7,977	65.76	24.55
	2000-2009	3,751	9,673	72.91	33.96
High income[d]	1990-1999	22,118	26,199	96.81	55.44
	2000-2009	26,551	31,912	97.52	63.27

Source: World Bank World Development Indicators and Education Statistics.

[a] Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown.
[b] Data are not available for all years for every country or group of countries.
[c] Data are not available.
[d] World Bank's definitions of income groups are available at http://data.worldbank.org/about/country-classifications/country-and-lending-groups.

The natural question is why do individuals in lower middle income countries abstain from upper schooling? Are they constrained out of schooling by lack of credit access and school availability? Or do they choose not to attend schooling beyond mandatory level because the returns to extra education are not there?

I use switching regressions framework to estimate schooling decision equations and returns to upper- and post-secondary education in Indonesia. While low enrollment rates for tertiary schooling make analysis of returns to college education alone irrelevant for large part of the population, focusing only on decision to enroll in senior high school, on the other hand, without taking into account information revealed by individual's decision about college enrollment, by construction leads to incomplete analysis of returns to schooling.

In the analysis, I differentiate between vocational and general upper secondary programs. I find significant positive returns to vocational senior high graduation and college attendance. An average individual from the sample would earn 4.5 times more if he attended college than as a non-vocational senior high graduate, and 1.3 times more as a vocational rather than non-vocational senior high graduate. There is evidence of financial constraints for college attendance decision but not for senior high school attendance. However, levels of wages in agriculture have significant negative effect on senior high attendance, for both vocational and non-vocational programs, suggesting that low enrollment in post-secondary education is at least partly explained by high opportunity cost of schooling.

The rest of the chapter is organized as follows. Section 1.2 provides overview of educational system in Indonesia. Section 1.3 presents estimation framework. Section 1.3 describes the data. Section 1.5 discusses estimation results. Section 1.6 concludes the paper.

1.2 Education in Indonesia

National education system in Indonesia is very young and at an early stage of development.[3] During the three-century Dutch colonial rule, from middle of the 17th century to 1949[4], schooling was oriented primarily towards Dutch population. Only in later years, towards the end of the 19th century, schooling system was extended to include Indonesian aristocracy in an effort to form lower-rank administration with local labor force.

[3] This section is based on Johnson, Gaylord, and Chamberland (1993) and Bjork (2005).
[4] Independence was declared in 1945 and recognized in 1949.

After Indonesia gained independence in 1949, the right of every Indonesian to education was declared by the National Law #4 in 1950. The brief history of Indonesia's national education system is intense and uneven. Its ups and downs echo political developments. Initial expansion and large degree of autonomy of schools in the first years of country's independence were followed by growth slowdown under the tightening dictatorship of General Sukarno's Guided Democracy. When in 1965 General Suharto ousted Sukarno, replacing Guided Democracy with the New Order, estimated 300,000-400,000 people were executed in the anti-communist cleansing, including large number of teachers.

The tide has changed in the 1970s with the windfall from high oil prices. School expansion was particularly strong during second Repelita, or five-year development plan, of 1974-75 to 1978-79, when government used oil money to fund a massive school construction program (INPRES). Duflo (2001) reports that during this period 61,000 primary schools were constructed, an average of two schools per 1,000 children of primary school age. Duflo finds that the INPRES increased average years of attained schooling by 0.12 to 0.19 and led to 1.5 to 2.7 percent increase in wages. National Education Law of 1989 established national education system and set standards for all schooling levels.

The structure of schooling system was inherited from the Dutch. It begins with six years of primary school, starting at the age of six. This is followed by three years of junior high school, three years of senior high school, four years of undergraduate studies, three years of master studies, and three years of doctorate studies. The 1989 education law increased mandatory education from complete primary to complete primary plus complete junior high education. It was implemented in phases during early 1990s.

National education system was modified on the go as it developed, reflecting the learning by doing approach the government had to adopt to facilitate rapid expansion. In 1988 the Ministry of Education and Culture (MOEC)[5] decreed to close all vocational junior high

[5] The MOEC was reformed into the Ministry of National Education in late 1990s.

schools, basing this decision on the body of international research suggesting that junior high level was too early for separation into vocational and academic trainings. However, vocational training is very popular among low-income population group, as it is perceived as direct skill acquisition for specific occupation.

The quality of Indonesian schooling system suffers from the side-effects of rapid expansion from a very small base. At the moment of independence in 1949, there were only two tertiary state institutions in the country. Thirty more were established during 1960s. In mid-eighties, about half of all towns had a junior high school, and about a third had a senior high school. In rural areas, situation was much worse. Significant expansion was necessary at every education level. However, very low literacy rates required particularly large increase in number of primary and junior high programs, which, as a result, were most affected by the strain on human and physical resources. Between 1945 and 1984, primary school attendance jumped from 2,523,000 to 26,567,688.

To this day, demand for education drops significantly as one moves to higher schooling levels. In 1989, 24 percent of junior high graduates who qualified academically to continue to senior high level did not do so. The drop is even higher beyond secondary schooling. Only 5.77 percent of all 19 to 24 year-olds were enrolled in a tertiary program in 1991. As a result, problems that plague Indonesian schooling at all levels are particularly stark at lower schooling levels.

State expansion of schooling system was followed, with about ten-year lag, by emergence and growth of private education. Private schools multiplied rapidly, outnumbering state schools at each level by at least a factor of ten. A number of private schools are well-established, with highly competitive admissions, challenging curricula, and faculty and reputation rivaling those of the best state schools.[6] However, the majority of private schools are small, with underqualified personnel unable to fully cover the MOEC-required curricu-

[6] Elite private schools are usually linked to religious foundations.

lum. It is not uncommon for a private school to consist solely of several rented apartment floors located above a supermarket. Unlike western private schools, private schools in Indonesia are generally a backward option to obtain graduation papers for those who do not qualify for admission to more prestigious state schools.

All schools, whether public, private, or religious, have to follow centralized MOEC-approved curriculum. The MOEC certifies all programs offered by private schools at every level. Accreditation is done by the smallest program unit. For example, if a private college has two departments each offering three specialization fields, these six specialization fields are accredited individually. Each private program is given one of three statuses: equalized, recognized, or registered. Equalized programs are considered to be on a par with state programs of the same education level. Recognized and registered programs are considered inadequate by the MOEC. Graduation certificates and diplomas from private programs with recognized or registered status are not recognized by the state. Consequently, they cannot be used to apply to programs at the next level of education in any school, either state or private, or for employment by the state. Graduates from recognized or registered programs can take state certification exam, passage of which gives their certificates equalized status.

The overwhelming majority of private programs do not have equalized status. In 1990, the number of equalized programs was 524 out of almost 14,000, or less than 4 percent, for junior high schools, 455 out of 9,200, or 5 percent, for senior high schools, and 8 percent for tertiary programs.

There are several reasons for this quality gap between state and private schools. Establishment of private schools generally lagged state schools by 10-15 years. This gave state schools more time to accumulate physical assets and invest into training of their personnel. State schools established right after the independence inherited the campuses of colonial Dutch schools, with large classroom buildings, libraries and sports grounds. State schools benefited from direct government funding, particularly so during the three decades from 1968 to 1997 when Indonesian economy grew at an average annual rate of 7.4 percent. In addition to money from the oil price windfall, government had sufficient resources to invest in public education from such sources as the World Bank, Asian Development Bank, United States Agency for International Development (USAID) and United Nations Educational, Scientific and Cultural Organization (UNESCO). This enabled state schools to send their teachers and lecturers overseas for advanced studies and qualification improvement.

As a result, state schools have better facilities, higher teacher-student ratios and better qualified lecturers who had first-hand experience of western education methodology. State schools have more stringent admission criteria and more rigorous curricula. This attracts the best students and leads to more able student body and higher graduation rates. In addition, state schools are considerably cheaper than private schools, which charge two to three times higher tuition fees as well as a series of other fees such as 'construction' and 'administration' fees.

Factors that are behind the high prestige of state schools compared to private schools also contribute to considerable regional variation in schooling quality within state programs. The more prestigious, older schools are concentrated in large urban areas constituting traditional political and cultural centers. The same is true for schools based on Dutch campuses which are situated close to administrative centers.

This tendency is true on both regional and country-wide scales. In every province, most schools are located in urban areas, and best schools are close to city downtown. School rank decreases with its distance from the city center.[7] Correspondingly, prestigious schools, whether public or private, attract students from well-educated families, with parents employed in high-ranking civil posts or in prosperous private enterprises. As the quality of school goes down, parents' education declines and their employment shifts to blue collar and

[7] Bjork (2005) provides example of lower secondary schools in the city of Malang (East Java province), comparing 19 percent admission rate in elite central public school to 58 percent in average public school on the edge of town to universal admission in several private schools.

menial service, and then further on to agriculture.

Similarly, out of the 29 provinces, Java was always the administrative, cultural and economic center of Indonesia. One consequence is that Java has both more schools than any other province, as well as almost all elite schools, both state and private. Out of total of 63 tertiary state schools in Indonesia in 1993, 27 were on Java.[8] In contrast, many other islands have only one or two. Major institutes and universities on Java admit top 5 percent of all students nationwide.

Two important conclusions emerge. First, at each level of education, school quality and availability are strongly correlated with proximity to administrative centers, resulting in large variation in both availability and quality of schools between rural and urban areas. Second, state expansion of schooling system is focused on mandatory schooling levels - primary and junior high. As a result, expansion at senior high and tertiary levels comes mostly from private sector.

1.3 Estimation Framework

I use switching regression framework to model schooling choice. Individual's choice between available options is represented by decision equation. Resulting outcomes under different choice options are represented by a set of outcome equations. Individual uses decision equation to make his choice among available options and then switches to an outcome equation to collect the outcome corresponding to his choice.[9]

Let $j \in \{1, ..., J\}$ index the J available choices. For a junior high graduate making decision about senior high attendance, there are three options to choose from: no senior high school, vocational senior high school, and non-vocational senior high school. For a

[8]Table A.1 in appendix A shows availability and summary statistics of public tertiary institutions by province.

[9]This setup is also known as a latent index model, where the latent variable corresponds to the decision equation (Heckman, Tobias, and Vytlacil, 2001)

senior high graduate making decision about whether to attend college, there are two options to choose from, college and no college. Let

$$y_{ij} = X_i \beta_j + u_{ij} \tag{1.1}$$

represent wages of individual i who chooses option j, where X_i are factors affecting individual's productivity. Wage equations are choice-specific because occupations corresponding to different schooling choices may vary in optimal usage and desirable combinations of personal characteristics. Equations (1.1) are outcome equations for choices $j = 1, ..., J$. Let

$$V_{ij} = Z_i \gamma_j + \varepsilon_{ij} \tag{1.2}$$

denote individual i's indirect utility from choice j, where Z_{ij} are factors that affect individual i's preferences for choice j. Disturbance terms u_{ij} include unobserved personal characteristics that affect individual's outcome for each option, while disturbance terms ε_{ij} include unobserved personal characteristics that affect individual's indirect utility from each option. As long as there exist unobserved personal characteristics that affect both individual choice and individual outcome under different choice options, u_{ij} and ε_{ij} are correlated. For example, a gifted mechanic is both more likely to choose vocational training and to be more productive in a vocational profession. Individual i chooses option $s \in \{1, ..., J\}$ if it provides him with maximum indirect utility, that is,

$$s \text{ is chosen } \Leftrightarrow V_{is} > \max_{\substack{j=1,...,J \\ j \neq s}} V_{ij}. \tag{1.3}$$

Equation (1.3) is decision equation, which depends on preference equations (1.2).

Consider wage equations (1.1). Available options are mutually exclusive, so wage y_s is observed only for individuals who choose option s using decision equation (1.3). Because unobservables in decision and outcome equations are correlated, this nonrandom self-selection into subsample for which y_s is observed is a source of selection bias:

$E(y_s|y_s \text{ is observed}, X, Z) = E\left(y_s|V_s > \max\limits_{\substack{j=1,\dots,J \\ j\neq s}} V_j; X, Z\right) =$

$= X\beta_s + E\left(u_s|\max\limits_{\substack{j=1,\dots,J \\ j\neq s}} V_j - \varepsilon_s < Z\gamma_s\right) \neq X\beta_s = E(y_s)$, since expectation term in the sum is not equal to zero due to correlation of u_{ij} and ε_{ij}. The difference between population mean $E(y_s)$ and observed conditional mean $E(y_s|y_s \text{ is observed}, X, Z)$ is the measure of selection bias. In other words, distribution of outcomes for a given choice option differs for individuals who chose that option from that for the whole population. Including a measure of selection bias as an explanatory variable when estimating wage equations (1.1) would account for existing self-selection and produce unbiased estimates of β_s.

Lee (1983) develops parametric method of accounting for selection bias in multiple-choice switching regression models. Define $\eta_s = \max\limits_{\substack{j=1,\dots,J \\ j\neq s}} V_j - \varepsilon_s$. Under the assumption that disturbances ε_j are independently and identically Gumbel distributed, the distribution of η_s is $F(\eta_s) = \exp(\eta_s) / \left[\exp(\eta_s) + \sum\limits_{\substack{j=1, \\ j\neq s}}^{J} \exp\left(X\beta_j\right)\right]$, and the probability that option s is chosen is

$$P_s = \frac{\exp(X\beta_s)}{\sum\limits_{j=1}^{J} \exp\left(X\beta_j\right)},$$

which is the multivariate logit model. As Lee shows, this implies that selection bias $E\left(u_s|\max\limits_{\substack{j=1,\dots,J \\ j\neq s}} V_j - \varepsilon_s < Z\gamma_s\right)$ for option s is equal to

$$-\sigma_s\rho_s\frac{\phi\left(\Phi^{-1}(P_s)\right)}{P_s},$$

where σ_s is standard deviation of disturbance u_s, ρ_s is correlation coefficient of u_s and

$\Phi^{-1}(F(\eta_s))$, which is a standard normal variable, ϕ is standard normal density function, and Φ is standard normal cumulative density function. Wage equations (1.1) can then be rewritten as

$$y_{ij} = X_i\beta_j - \sigma_j\rho_j \frac{\phi\left(\Phi^{-1}\left(P_{ij}\right)\right)}{P_{ij}} + e_{ij}, \tag{1.4}$$

where $E(e_{ij}|s \text{ is chosen}) = 0$.

Estimation is done in two stages. In the first stage, decision equations (1.3) are estimated with multinomial logit, and the resulting estimates are used to construct $\hat{P}_{ij}$ for $j = 1, ..., J$. In the second stage, constructed $\hat{P}_{ij}$ are used to estimate wage equations (1.4). Correction of the second stage covariance matrix for the fact that explanatory variables measuring selection bias were constructed from stage one estimates can be performed following, with appropriate modifications, Lee, Maddala, and Trost (1980). Constructed $\hat{P}_{ij}$ variables are in essence probabilities that an individual with characteristics Z_i will choose option j. Estimated coefficients on $\hat{P}_{ij}$ variables in wage equations demonstrate two things. Statistical significance of the estimated coefficient supports the assumption of correlation between error terms in decision and wage equations, indicating that there indeed exists statistically significant difference in distribution of the error term in the wage equation between the whole sample and the subsample of individuals that self-selected into option j. As Heckman, Urzua, and Vytlacil (2006) note, quality of available data affect incidence of correlation between error terms in decision and wage equations. The sign of estimated coefficient on the $\hat{P}_{ij}$ variable indicates whether this correlation of the error terms in decision and wage equation is positive or negative.

I estimate two schooling decision equations, one made at the time of junior high graduation about senior high attendance, and the other made at the time of senior high graduation about college attendance. The sample of senior high graduates used for estimation of college decision equation is a subsample of the set of junior high graduates used for estimation of senior high decision equation and consists of those junior high graduates who three or four years earlier chose to attend senior high school. I do not integrate the two schooling decisions of an individual, so there is no dynamic aspect to the analysis. However, I use information revealed by individual's decision about both senior high and college attendance in estimating returns to additional schooling.

Let s_1, s_2 and s_3 denote $\hat{P}_{ij}$ variables constructed from the estimates of decision equation on senior high attendance and corresponding, respectively, to three available choices of no senior high, non-vocational senior high, and vocational senior high. Similarly, let c_1 and c_2 denote $\hat{P}_{ij}$ variables constructed from the estimates of decision equation on college attendance and corresponding, respectively, to the two available choices of no college and college. By construction, s variables capture the bias due to correlation of errors in wage equations and equation describing decision to attend senior high school. They do not directly capture the bias due to correlation of errors in wage equations and equation describing decision to attend college. That is, to accurately account for self-selection when estimating effect of post-junior high schooling on wages we need to use both s and c variables.

Consider three individuals: individual 1 does not plan to attend senior high, individual 2 wants to attend senior high only, and individual 3 wants to attend college and therefore plans to attend senior high school. That is, both individuals 2 and 3 are likely to attend senior high, but for very different reasons. In this case, values of s variables for individuals 2 and 3 will capture this higher likelihood of senior high attendance, but will not directly reflect differences in college preferences of individuals 2 and 3. In other words, while variables s are sufficient for distinguishing self-selection of individual 1 from individuals 2 and 3, effective differentiation of individual 2 from individual 3 requires knowledge of variables c. Because variables c measure likelihood of college attendance conditional on senior high graduation, usage of both s and c variables in wage equations does not result in biased or inefficient estimates.

1.3.1 Brief Literature Overview

In a seminal work that lies at the basis of selection models literature, Roy (1951) develops the idea that distribution of income in the population depends on the distribution of productivity-relevant skills in the population, as well as on how the effect of a given set of individual skills on productivity varies from occupation to occupation. As a result, individuals select into occupations based not only on how skilled they are relative to the rest of the population, but also on how valuable their specific set of skills is in different occupations. For example, consider two occupations. In occupation A, individual productivity varies a lot with individual skills, while in occupation B all individuals are equally productive. In this case, individuals whose skills lead to high productivity in occupation A are more likely to select occupation A over B than individuals whose skills do not lead to high productivity in occupation A. Consequently, distribution of occupation A earnings for individuals who chose occupation A is different than that for the whole population. Another insight from Roy (1951) is that self-selection bias is more pronounced for occupations where productivity is highly affected by individual skills. In more general terms, self-selection bias is more pronounced for choice options where the outcome is highly affected by individual characteristics unobserved by the econometrician but known to the individual.

In the literature, available choice options are commonly referred to as treatments.[10] Consequently, the difference between expected outcome values for two different choice options, where expectations are taken over the whole population rather than groups that self-select into each treatment, is called the average treatment effect. Average treatment effects are of high interest to policymakers, as they allow quantitative evaluation of policies. For example, large positive average treatment effect of college versus no college treatments reveals potential benefit from programs aimed at increasing college enrollment. Heckman, Tobias,

[10] The convention is to label one of the options as "no treatment" in binary choice models, where such distinction is intuitive. For example, in a model of schooling decision between college and no college, college attendees are the treated and high school graduates with no college are the untreated.

and Vytlacil (2001) derive expressions for less aggregated measures of treatment effects and provide overview of further assumptions needed for estimation of the dispersion of treatment effects in the population.

For empirical estimation of models with self-selection bias, central identification strategy is exclusion restriction. This implies variables available to the econometrician that affect individual's choice, such as individual's decision to obtain more schooling, but do not affect the outcome, which in most cases is earnings. In essence, variables satisfying exclusion restriction are instrumental variables.

Heckman (1979) introduced two-step procedure for estimation of switching regression models, where decision equation is estimated in the first step with probit, and these first step estimates are used to construct measures of selection variables which are then used in OLS estimates of outcome equations in the second step.[11] The method assumes joint normality of error terms in decision and outcome equations.

Early parametric application of self-selection model to returns to schooling is Willis and Rosen (1979), who use probit to estimate binary college versus no college schooling decision. They find positive and statistically significant self-selection bias for both college and high school groups. However, their sample is limited to high-ability U.S. military personnel; correspondingly, both their sample college and high school groups are not necessarily representative of college and high school groups in the population.

Parametric estimation methods for polychotomous models were proposed by Hay (1980), Dubin and McFadden (1984), Lee (1982) and Lee (1983). Subsequent work produced semi-parametric estimation methods which avoid specification of the joint distribution of the error terms in decision and outcome equations. Examples include Heckman (1980), Manski (1985), and Newey, Powell, and Walker (1990) for dichotomous choice models and Ichimura and Lee

[11] Imbens and Angrist (1994) and Angrist and Imbens (1995) propose alternative method of instrumental variables. While the switching regression framework estimates levels of conditional means, the instrumental variables framework estimates the slopes of conditional means (Heckman, Urzua, and Vytlacil, 2006).

(1991), Lee (1995), and Dahl (2002) for polychotomous choice models.

The inherent danger of parametric estimation is potential imposition of false distributional assumptions. After examining existing empirical literature, Heckman (2001) finds little evidence of the practical effect on point estimates of relaxing distributional assumptions in parametric specifications, despite conceptual advantages of semiparametric and nonparametric specifications.

1.4 Data

I use data from Indonesian Family Life Survey, administered by RAND in collaboration with the University of Indonesia. The data were collected in four waves, in 1993, 1997, 2000 and 2007. From 1968 to 1997 Indonesia enjoyed thirty years of rapid growth, with average growth rate of 7.4 percent.[12] The first wave of the survey captures the latter part of this high-growth period. The growth phase ended with East Asian Financial Crisis of 1997-1998, with economy shrinking by 13.1 percent. The second wave of the survey provides pre-crisis data, while the third wave reflects the recovery. Between 2000 and 2007, Indonesian economy grew at the average rate of 5.1 percent. This latest growth spur is captured by the fourth wave of the survey.

The sample covers 13 of 26 provinces that existed in 1993, which include provinces on four out of five largest islands, and is representative of 83 percent of population. Figure A.1 in appendix A shows the map of Indonesia and provinces included in the survey. The sample was designed to provide balanced representation of rural and urban areas. The data were collected on individual, household and community levels, and include retrospective records of individual education, employment and migration. Family structure of the survey allows to link data on parents to data on their offspring.

I limit the sample to those born in 1968 and later. Since primary school attendance

[12] Aggregate growth data in this paragraph are from World Bank Development Indicators.

starts at the age of six, 1968 cohort was the first to attend new schools opened under the INPRES school construction program. I consider only individuals with complete education record who have completed schooling and for whom employment data and data on parents are available.

In this section I perform preliminary data analysis, comparing schooling trends in rural and urban areas and looking for intergenerational changes. I split the sample into rural and urban groups based on individual's birthplace. For intergenerational comparisons, I use data on individuals' parents. Most, but not all, parents belong to pre-reform cohort. For convenience, I refer to parents as older generation and to children as younger generation.

Table 1.2 shows, in percent, the breakup of highest attended and completed education levels by area of birth, generation and gender. In the urban area, there is significant increase in maximum education between parents and children, for both men and women. Among parents, half of men attended at most junior high school, and for a third of men highest level completed is senior high school. For women in parents' generation the education levels are even lower. Almost 40 percent of women completed only primary school. In contrast, more than 50 percent of the children, both male and female, completed at least senior high school. College attendance rate increased by 50 percent for men and tripled for women. Another significant positive trend is reduction in schooling gap between men and women.

These positive developments are even more pronounced in rural area. In the parents' generation, 19 percent of men and 36 percent of women in rural area received no schooling. These numbers drop to one and two percent, respectively, for their children. The most widespread level of schooling increased from primary among parents to senior high among children, an increase by two schooling levels. However, increase in college attendance was more moderate than in urban area. In other words, while there was significant increase in schooling between parents and their children in both urban and rural areas, in urban area this increase affected all levels of schooling, while in rural area in was concentrated on lower

Table 1.2: Highest Attended and Completed Education Level, in Percent
(Within a group, each column sums to one)

Generation	Schooling	Males Attend	Males Complete	Females Attend	Females Complete
Urban area					
Parent	No schooling	4.80	4.80	11.46	11.46
	Primary	23.89	31.95	**29.77**	**38.44**
	Junior high	23.53	20.45	26.97	22.93
	Senior high	**34.30**	**32.85**	24.86	22.06
	College	13.03	9.50	6.74	4.91
	Graduate	0.45	0.45	0.19	0.19
Child	No schooling	0.52	0.52	0.42	0.42
	Primary	5.15	13.36	6.22	10.20
	Junior high	21.45	19.17	18.65	18.71
	Senior high	**53.00**	**50.20**	**50.39**	**48.16**
	College	19.75	16.62	23.90	22.09
	Graduate	0.13	0.13	0.42	0.42
Rural area					
Parent	No schooling	18.68	18.68	35.61	35.67
	Primary	**40.63**	**48.93**	**37.84**	**43.59**
	Junior high	19.28	14.83	15.57	11.42
	Senior high	15.48	13.35	7.91	7.08
	College	5.46	3.80	3.00	2.17
	Graduate	0.47	0.42	0.06	0.06
Child	No schooling	0.90	0.90	1.82	1.82
	Primary	16.27	25.94	19.55	25.11
	Junior high	33.37	29.76	33.43	31.62
	Senior high	**41.20**	**36.47**	**36.35**	**33.40**
	College	8.15	6.82	8.74	7.94
	Graduate	0.11	0.11	0.11	0.11

Area is determined by birthplace. For each group, largest percent is in bold.

schooling levels. As could be expected, schooling levels are higher in urban than in rural area, although this gap is narrower for younger generation.

The numbers in table 1.2 suggest that schooling attainment among men in younger generation in rural area is similar to that among men in older generation in urban area. If we were to think of changes in schooling attainment between generations as Markov-type sequential process, it is easy to envisage schooling attainment among men in older generation in rural area as stage n, schooling attainment among men in both older generation urban area group and younger generation rural area group as stage $n+1$, and schooling attainment among men in younger generation in urban area as stage $n+2$.

Table 1.3 shows transition probabilities for highest attended education level of men given father's highest attended schooling level. Men in both urban and rural areas are consistently more likely to achieve higher schooling than their fathers. Transition probabilities are similar

Table 1.3: Father-Son Transition Probabilities for Highest Attended Education Level
(Each row sums to one)

	Son's education				
Father's education	No schooling	Primary	Junior high	Senior high	College
Urban area					
No schooling	0.00	25.00	**50.00**	21.43	3.57
Primary	0.64	14.26	30.93	**48.72**	5.45
Junior high	0.00	4.52	18.98	**59.04**	17.47
Senior high	1.11	1.39	11.14	**58.22**	28.13
College	0.00	0.69	2.78	34.03	**62.50**
Rural area					
No schooling	3.09	**44.14**	34.88	17.28	0.62
Primary	0.55	24.97	**37.24**	33.52	3.72
Junior high	0.29	5.80	30.14	**54.78**	8.99
Senior high	0.00	5.69	12.81	**58.72**	22.78
College	0.00	1.85	4.63	45.37	**48.15**

For each row, largest percent is in bold.

between urban and rural areas when father has attended at least senior high school, but the patterns differ between the areas when father attended less than senior high school. In the latter case, in rural area sons are most likely to attend one schooling level higher than their fathers. In urban area, on the other hand, sons are most likely to attend at least senior high as long as their father had at least some schooling, an increase by more than one schooling level. It seems that difference in parents' education between urban and rural areas does not fully explain the lag in children's schooling advancement in rural area observed in table 1.2.

Table 1.4 shows, in percent, distribution of students in the younger generation across different school types, separately for public and private schools. Two school types considered are general and vocational programs. The third type is residual and consists of adult

Table 1.4: Distribution of Students by School Type, in Percent

		Urban area		Rural area	
School level	School type	Public	Private	Public	Private
Primary	General	15.37	84.50	8.76	91.21
	Other	0.06	0.06	0.03	0.00
Junior high	General	33.01	61.36	28.18	65.76
	Vocational	2.48	3.01	2.40	2.99
	Other	0.00	0.14	0.13	0.53
Senior high	General	27.11	29.24	24.23	33.61
	Vocational	29.10	14.10	26.60	14.60
	Other	0.09	0.36	0.44	0.52

education and schools for disabled. Of note is high importance of vocational training as a senior high school option. Vocational schools account for over 40 percent of all senior high students. At senior high level, the majority of vocational schools are public, both in urban and rural areas, while for general programs shares of public and private schools are much closer. Given the higher cost of private education, this implies that vocational schooling constitutes a cheaper senior high option.

1.5 Estimation Results

1.5.1 Stage One - Schooling Decision

Costs of and benefits from different available schooling choices can be of intellectual, psychic, and financial nature. They vary with a number of factors, which can be grouped into three categories. The first category includes individual-specific characteristics, such as ability, scholastic aptitude, and individual preference for schooling, also known as 'psychic component'. The latter is hard to measure and rare to observe. Measures of ability and aptitude, while available in some datasets, are commonly taken after the schooling decision is made and schooling is at least partially completed. This introduces potential reverse causality between individual's test scores and schooling choices. Individual-specific factors therefore often end up in the error term of a schooling decision equation.

The second type of factors is family-specific. These affect individual's ability through genetics, individual's preference for schooling though home environment, motivation and instilled values (Mare, 1980), and financial constraints through parents' earnings. Most family-specific factors are easy to observe, such as parents' schooling and income, and they serve as partial proxies for unobservable individual characteristics. The 'home environment' effect, on the other hand, is hard to measure. It can be argued that information on parents' education and employment as well as family demographics provide partial insight into the 'home environment' effect.

The third type of factors is location-specific. These mostly have effect on the financial determinants of schooling choice. Wage levels and employment opportunities in the area affect both opportunity cost of schooling and perceived pecuniary returns to schooling. Availability and quality of schools in the area affect direct costs of schooling through tuition and associated fees and through perceived returns to schooling. Local labor market conditions and school availability are usually observable, although the relative timing of location data measures and individual's schooling decision can be an issue. Information on school quality is less easily obtainable and often ends up in the error term.

For the estimation of schooling decision equations, I use explanatory variables of all three types. I use data on male junior high graduates born in 1968 or later, who had at least one long-term employment after they completed their schooling. I use only data on wage earners, excluding self-employed from the sample.[13]

An ability test was administered in the last two waves of the IFLS survey[14], which tests individual's cognitive ability with multiple-choice visual logic questions. Two out of the total of 12 questions are of higher than average difficulty, and I use the fraction of correctly answered hard questions as measure of ability. The majority of the sample answered the test at the age of 19-23. However, because the questions target innate cognition rather than familiarity with facts or acquirable knowledge, these test results are likely to give accurate representation of individual's cognitive ability at the time of junior high graduation.[15] I use indicators for whether individual's junior high school was vocational, public, and located in a rural area as measures of quality of individual's previous education. I also use indicator for individuals that graduated from junior high after the 1989 education law was implemented.[16] The 1989 education law increased mandatory schooling from primary to junior high. In the sample of junior high school graduates, graduation from junior high before it became mandatory can indicate difference in schooling preferences from the rest of the sample.

To control for family's socio-economic position, I use indicators for father's and mother's maximum education levels, father's total income and indicators for father being self-employed

[13] If both salaried and self-employment work data corresponding to different years are available for an individual, only his wage-earning employment record is included in the sample.

[14] An ability test was also administered in the second wave; however, its questions are not comparable with the test questions used in later waves.

[15] While the test also includes five math questions, I do not use them exactly because algebraic computations that math questions incorporate are taught in school and therefore individual's test performance would differ noticeably before and after senior high school.

[16] Because the law was implemented in phases during early 1990s, the indicator is set to one if individual graduated from junior high school after 1993.

and employed by the state in the year when individual graduated from junior high school, and number of siblings. Parents' education is expected to reflect the effects of family's financial and social positions, and so higher levels of parents' education should have positive effect on continued schooling. State employment is associated with stability and additional benefits and therefore is also expected to have positive effect on continued schooling. Larger number of children in the family reduces the amount of resources that can be spent on each child, including family's financial and time recourses available for child's education.

I use location-specific variables corresponding to the area where individual graduated from junior high school. I use number of senior high schools in the area to access school availability. I use indicators for agriculture and social services being among the top three sources of income in the area and for presence of a factory in the area, and distance to the nearest province or district capital, whichever is closer. Large presence of agriculture and factories would increase opportunity cost of schooling, while large presence of social services sector would signal good employment conditions after school completion. Distance to the nearest administrative center is a way to access the ease of migration. Because migration can take place both in search of a better job and better schools, and because job migration can be beneficial for both junior high and senior high graduates, the net effect of the distance variable on schooling decision is unclear. I also include indicators for the island of Java – location of country's capital and home to 59.8 percent of Indonesia's population[17] – and for outer islands, defined to include Kalimantan, Sulawesi, Lesser Sunda islands, Maluku and New Guinea. The island of Java was historically Indonesia's administrative and cultural center, what led to large concentration of prestigious schools on the island. This, however, can be outbalanced by potentially high competition for school placement and employment due to unusually large population density. Conversely, outer islands constitute traditionally peripheral regions, with low population density[18] as well as few schools.

[17] 1990 Population Census figure, Frankenberg and Karoly (1995).

[18] Frankenberg and Karoly (1995) report population density of less than 20 people per square kilometer in

Junior high school graduate has three options to choose from: pursue no more schooling, attend vocational senior high school, or attend non-vocational senior high school. In the logit estimation of this schooling decision, I use as a reference group individuals that choose no senior high school. Multivariate logit then estimates how explanatory variables affect individual's likelihood of belonging to non-vocational senior high group and to vocational senior high group relative to the base, or no senior high, group. That is, estimates of coefficients for the non-vocational senior high group should be interpreted as affecting individual's likelihood of attending non-vocational senior high school relative to not attending any senior high school, and same for the vocational group. Logit estimates of the senior high schooling decision equation are presented in the first two columns of table 1.5.

The estimates for the most part concur with expectations. Parents' education is an important determinant in individual's decision to attend senior high school. The likelihood of senior high attendance increases with parents' education, and the likelihood of choosing vocational over non-vocational school decreases with parents' education. For non-vocational senior high enrollment, the magnitude of this positive effect increases with the level of parents' education. Individuals whose father works for the public sector are more likely to attend non-vocational senior high school. Interestingly, neither father's income nor number of siblings has significant effect on senior high attendance. As public schools in Indonesia charge low tuition rates, this result suggests that the market of senior high education is not dominated by private schools. Overall, the effect of family variables is much less pronounced for vocational than for non-vocational enrollment.

The estimated effect of personal characteristics is similarly intuitive. Graduates of vocational junior high schools are unlikely to attend non-vocational senior high schools, and graduates from public junior high schools are more likely to enroll in senior high school. The latter result is in accordance with higher quality of public rather than private education in

Kalimantan, as opposed to over 700 people per square kilometer in Yogyakarta, one of Java's provinces.

Table 1.5: Logit Estimates of Schooling Choice Equations

	Senior high choice[a]		College choice[b]
	Non-vocational	Vocational	College
Family			
Father junior high	1.0138***	1.1283***	0.2324
	[0.3678]	[0.3647]	[0.4615]
Father senior high	1.2617***	0.9408**	0.6748
	[0.4164]	[0.4246]	[0.4394]
Father at least college	2.4027**	1.7177	0.7718
	[1.0826]	[1.0955]	[0.5553]
Mother junior high	0.7615**	0.33	0.5198
	[0.3826]	[0.3933]	[0.3983]
Mother at least senior high	1.8640**	1.2821	0.8312*
	[0.7847]	[0.7963]	[0.4467]
Father is government employee	1.1331**	0.8652	0.3069
	[0.5259]	[0.5438]	[0.3887]
Father self-employed	0.4895*	-0.2096	0.5727
	[0.2795]	[0.2730]	[0.3758]
Father's income	-0.1224	0.148	0.8236***
	[0.1582]	[0.1626]	[0.2338]
Number of siblings	0.0577	0.0662	-0.1266
	[0.0616]	[0.0623]	[0.0876]
Personal characteristics			
1989 education law	0.1523	0.5106	
	[0.3187]	[0.3263]	
Cognitive score	0.5901*	0.7115**	1.0534**
	[0.3141]	[0.3170]	[0.4562]
Previous school in rural area	-0.3897	-0.1996	-0.4638
	[0.2920]	[0.2915]	[0.3718]
Vocational school	-1.0930**	0.1154	-1.7102***
	[0.5287]	[0.4632]	[0.3679]
Public school	0.7820***	0.5424**	0.4856

Continued on next page...

Table 1.5 – continued from previous page

	Senior high choice[a]		College choice[b]
	Non-vocational	Vocational	College
	[0.2579]	[0.2540]	[0.3263]
Location characteristics			
Distance (km) to region capital	-0.0035	0.0009	0.0015
	[0.0068]	[0.0071]	[0.0097]
# senior high schools	0.0121	0.2521**	
	[0.1017]	[0.1002]	
Agriculture	-0.9137**	-0.7286*	-0.1126
	[0.3928]	[0.3885]	[0.3737]
Factory	-0.3801	0.0715	0.3241
	[0.2675]	[0.2789]	[0.3540]
Social services	-0.2206	-0.1924	-0.4016
	[0.2846]	[0.2915]	[0.4315]
Java island	-0.6546**	-0.3314	0.5444
	[0.2991]	[0.2997]	[0.3836]
Outer island	-0.4441	-1.0307**	0.4268
	[0.4077]	[0.4584]	[0.4905]
Intercept	1.7481	-3.141	-13.0844***
	[2.1874]	[2.2534]	[3.1676]
-2 x ln(Likelihood)	1169.9785		297.1848

[a]Reference group: no senior high.
Number of observations: 657
Number in no senior high group: 138
Number in non-vocational senior high group: 300
Number in vocational senior high group: 219

[b]Reference group: no college
Number of observations: 396
Number in no college group: 300
Number in college group: 96

***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.
Standard errors are in parentheses.

Indonesia. Interestingly, cognitive test score has only marginal (10 percent significance) positive effect on non-vocational senior high enrollment and stronger positive effect on vocational school enrollment. Part of the reason might be the simple nature of the test, which while capturing innate cognitive ability does not reflect other individual traits that can greatly influence the odds of school admission such as how hardworking the individual is.

Large presence of agriculture in the area has significant negative effect on senior high enrollment, indicating the importance of the opportunity cost of schooling. Number of senior high schools in the area has no effect on non-vocational enrollment and positive effect on vocational enrollment. This result is intuitive, as larger availability of schools facilitates more effective sorting of individuals into different school types.

The last column of table 1.5 shows logit estimates of college decision. It was estimated with a subsample of senior high graduates.[19] I use the same explanatory variables as for estimation of decision equation on senior high attendance, excluding indicator for 1989 law and number of senior high schools in the area. Variables on father's employment refer to the year of senior high, rather than junior high, graduation. Similarly, indicators of previous school characteristics describe senior high school, and location variables describe area where individual graduated from senior high. Senior high school graduate has two options to choose from: attend college or not attend college. The reference group in this estimation is no college.

Neither family's non-financial background nor employment conditions affect individual's college decision. Lack of effect of parents' education on individual's college decision, which is so pronounced at the earlier stage of senior high decision, is unsurprising in light of data trends discussed in section 1.4. Table 1.2 illustrated that sons achieve schooling of at least one level higher than their fathers. As a result, by the time of senior high graduation, the majority of individuals already have at least the same level of schooling as their fathers. What does have significant positive effect on individual's decision to attend college is individual's

[19] Out of 519 individuals in the sample who attended senior high, 59 did not graduate, and so are not included in the logit estimation of college decision.

ability and family's finances. It is interesting that father's income is highly significant in college decision while it is completely insignificant in senior high attendance decision. It seems that private schools are much more present in tertiary than in secondary education. Attendance of vocational senior high virtually bars individual from college. This latter result, together with the similarly strong negative effect of previous vocational schooling on non-vocational senior high enrollment illustrates the continued impact on the individual of a decision made as early as primary school graduation.

1.5.2 Stage Two - Wage Equations

I use the two logit estimates to construct measures of self-selection, one corresponding to decision to attend senior high school and the other to college attendance decision. Individuals in the estimation sample have one of three levels of schooling: junior high graduate, senior high graduate, or college attendee. Senior high group includes graduates from vocational and non-vocational programs. I estimate three wage equations, one for each schooling level. While one would expect variation between returns to vocational and non-vocational senior high schooling, it is unlikely that there exist structural differences in wage equations for vocational and non-vocational senior high graduates in the way they do, for example, for senior high graduates and college attendees.

In notation of section 1.3, s_1, s_2 and s_3 denote measures of selection bias constructed from the estimates of decision equation on senior high attendance, and c_1 and c_2 denote measures of selection bias constructed from the estimates of decision equation on college attendance. Define s as the combination of s_1, s_2 and s_3 according to individual's actual decision about senior high attendance:

$$s = \begin{cases} s_1 \text{ if did not attend senior high} \\ s_2 \text{ if attended non-vocational senior high} \\ s_3 \text{ if attended vocational senior high} \end{cases}.$$

Define $\bar{s}_3$ as the product of s_3, measure of bias from self-selection into vocational senior high school, and a dummy variable for attendance of vocational senior high school:

$$\bar{s}_3 = \begin{cases} s_3 \text{ if attended vocational senior high school} \\ 0 \quad \text{otherwise} \end{cases}.$$

The wage equations I estimate are

$$y = X\beta_1 + \delta_1 s_1 + e$$

for the subsample of junior high graduates who didn't attend senior high,

$$y = X\beta_2 + \delta_{21} s + \delta_{22}\bar{s}_3 + \mu_1 c_1 + e$$

for the subsample of senior high graduates who didn't attend college, and

$$y = X\beta_3 + \delta_{31} s + \delta_{32}\bar{s}_3 + \mu_2 c_2 + e$$

for the subsample of senior high graduates who attended college. Coefficients μ capture the effects of self-selection during college choice, coefficients δ_{21} and δ_{31} capture the effects of self-selection into non-vocational senior high schools, and combinations $\delta_{21}+\delta_{22}$ and $\delta_{31}+\delta_{32}$ capture the effects of self-selection into vocational senior high schools.

The dependent variable is natural logarithm of hourly wage (in 2000 rupiah), constructed from monthly wage and hours worked in an average week as reported by the individual. The data provide year when working became individual's primary activity for the first time. Employment data are collected at several points in individual's life; however, there exists variation across individuals in years of working experience for which wage data are available. For each individual, I use wage observation for the year closest to five years since entering labor force. Returns to schooling are most manifest five to 10 years after schooling investment (Willis and Rosen (1979), Trost and Lee (1984)).[20]

I include years after school completion among explanatory variables. I also use indicator for usage on Indonesian language at home. In this country of over 700 languages (Lewis, 2009), Indonesian is the only official language, used in formal commerce and administration, and fluency is likely to be an asset. To account for variations in types of employment, I use indicators for public employment and firm size measured by number of employees. To the extent that firm's size is an indicator of its profitability, it is expected to have positive effect on wages. One of the attractions of state employment is wage stability. To some extent, both firm size and state employment are indicative of the quality of individual's employer. Because I estimate separate wage equations for different schooling levels, these controls do not attenuate estimation of the differences in returns to different education levels. I also use indicators for individual's job being located in an urban area, on Java island, and on outer islands. Equation for senior high graduates with no college includes indicator for vocational senior high program, to capture the difference in returns between vocational and non-vocational schools.

Estimation results are presented in table 1.6. In these second stage estimations, I correct covariance matrices for the fact that measures of self-selection were constructed using estimates from two first stage logit estimations. I also show OLS estimates of wage equations, which do not account for selection bias. Results of the likelihood ratio tests for equivalence

[20] As Trost and Lee (1984) indicate, this relationship was first observed in Mincer (1974).

Table 1.6: Estimates of Wage Equations by Schooling
Dependent Variable: Ln Hourly Wage (2000 Rupiah)

	No senior high		Senior high no college		College	
	OLS	Unbiased	OLS	Unbiased	OLS	Unbiased
Indonesian at home	-0.1311	-0.1361	0.2636***	0.2515***	0.1424	0.0000
	[0.1602]	[0.1620]	[0.0815]	[0.0827]	[0.1707]	[0.1619]
Years after school	0.0196	0.0197	0.0591***	0.0587***	0.0368	0.0280
	[0.0334]	[0.0337]	[0.0173]	[0.0176]	[0.0335]	[0.0311]
State employee	0.0875	0.0853	0.4433***	0.4283***	0.0789	-0.0907
	[0.5210]	[0.5258]	[0.1423]	[0.1446]	[0.1850]	[0.1738]
5-19 workers	-0.0666	-0.0637	0.2865***	0.3019***	-0.1697	-0.0913
	[0.1668]	[0.1682]	[0.1070]	[0.1090]	[0.2711]	[0.2479]
20-99 workers	0.0875	0.0888	0.3278***	0.3387***	0.2179	0.2892
	[0.1931]	[0.1944]	[0.1152]	[0.1170]	[0.2782]	[0.2554]
100 or more workers	0.6256**	0.6330**	0.5580***	0.5863***	0.3371	0.5124*
	[0.2772]	[0.2797]	[0.1192]	[0.1211]	[0.2848]	[0.2651]
Urban area	0.1577	0.1528	0.1225	0.1211	0.3152*	0.1404
	[0.1494]	[0.1514]	[0.0896]	[0.0908]	[0.1856]	[0.1773]
Java island	-0.4252**	-0.4060**	-0.1949**	-0.2325**	-0.0157	-0.0626
	[0.1666]	[0.1787]	[0.0927]	[0.1018]	[0.2113]	[0.2027]
Outer island	-0.0058	0.0024	-0.2280	-0.1152	-0.0275	-0.0819
	[0.2491]	[0.2544]	[0.1431]	[0.1593]	[0.2625]	[0.2532]
Vocational			-0.0276	0.8304**		
			[0.0811]	[0.3363]		

Continued on next page. . .

Table 1.6 – continued from previous page

	No senior high		Senior high no college		College	
	OLS	Unbiased	OLS	Unbiased	OLS	Unbiased
Intercept	7.1596***	7.0822***	6.9475***	6.4641***	7.5950***	8.5162***
	[0.2213]	[0.3239]	[0.1437]	[0.2854]	[0.3154]	[0.3559]
Self-selection (ϕ/F) into:						
No senior high		0.0596				
		[0.1820]				
Senior high				0.4697**		-0.5765*
				[0.2235]		[0.3000]
Vocational senior high (slope)				-0.8719***		0.2459
				[0.3259]		[0.2198]
Vocational senior high (computed)				-0.4021*		-0.3306
No college				0.2054		
				[0.2133]		
College						-0.3819*
						[0.1981]
Number of observations	102	102	271	271	93	93
Adjusted R^2	0.0612	0.0520	0.2029	0.2174	0.0678	0.2426
LR test for equivalence of OLS and unbiased models:						
χ^2		0.1240		8.1272		22.7362
degrees of freedom		1	0	3	0	3
P-value		0.7248		0.0435		

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

Standard errors in parentheses are adjusted for first-stage estimation.

of simple OLS estimation and two-stage estimation accounting for selection bias are presented at the bottom of table 1.6.

Explanatory variables have expected effect on wages in the equation for senior high graduates with no college (two columns in the middle of table 1.6). Individuals with knowledge of Indonesian language earn higher wages. Wages also increase with number of years since school completion, state employment and firm size. The latter effect is most pronounced for larger firms with at least 100 employees, both in terms of its magnitude and significance, as it is also significant for no senior high group (first two columns of table 1.6) and marginally significant at 10 percent level for college group (last two columns of table 1.6).

There is evidence of selection bias in the senior high group. Self-selection is positive and significant for individuals attending non-vocational senior high schools. This means that non-vocational senior high graduates who didn't go to college on average earn more than would a person selected at random from the whole sample of junior high school graduates were he to obtain the same education. We also see that the difference in the effect of self-selection into vocational and non-vocational senior high schools is statistically significant. Note that there is no evidence of significant effect of self-selection into no college. Combined, these results suggest that non-vocational senior high graduates who didn't go to college are better off than an average person would have been in their position not because they made the right decision to not attend college, but because they earlier made the right decision to attend non-vocational senior high school.

Statistical significance of self-selection in the senior high group is confirmed by the likelihood ratio test. The test rejects the null hypothesis of equivalence of the estimates obtained with controlling for self-selection and with OLS model without such controls. While for most of the explanatory variables coefficient estimates are not much different from the simple OLS, the difference is obvious for the indicator for vocational senior high program. The unbiased estimate is large in magnitude and statistically significant, while the simple OLS estimate is close to zero and indeed is not statistically different from zero. The unbiased estimate indicates that for senior high graduates who don't attend college, there is a wage premium for vocational training. This result agrees with earlier findings of positive return to technical versus general high school education (Trost and Lee (1984), Freeman (1974)). Note that this premium on vocational training is unconditional, so that it applies to any randomly selected individual from the sample. However, simple OLS estimation without controlling for selection bias would not have identified this extra return on vocational schooling.

Table 1.6 shows that there is no significant self-selection for individuals attending vocational senior high schools. This implies that individuals who chose to attend vocational senior high school made that decision not because they are better suited for vocational occupations than an average person in the sample but because of the unconditional wage premium on vocational training for senior high graduates.

Selection bias is not significant for no senior high group. Correspondingly, likelihood ratio test does not reject the equivalence of OLS and selection-correcting models for no senior group.

For the college group, selection bias is only significant at 10 percent level. Interestingly, the effect is negative both for bias from self-selection into non-vocational senior high attendance and for bias from self-selection into college attendance. This implies that a person drawn at random from the whole sample, were he to attend college, would in fact earn more than individuals who chose to attend college. This effect would be even stronger if college was preceded by attendance of non-vocational senior high school. Statistical significance of self-selection bias for college group is supported by significance of the likelihood ratio test.

One possible explanation of this negative self-selection result is that my sample is limited to wage earners. If more successful college attendees are self-employed, then the subsample of college attendees I use in the estimation is subject to yet another selection bias.

Table 1.7 shows differences in hourly wages between schooling levels. Panel A shows differences constructed from estimates of wage equations, evaluated for an individual with sample average values of explanatory variables. This average individual does not speak

Table 1.7: Wage Premiums by Schooling

Panel A: Constructed from estimates of wage equations

		Difference with (% change)		
	Wage level[a]	No senior high	Non-vocational	Vocational
No senior high	0.993			
Non-vocational	0.973	-1.98		
Vocational	2.266	128.24	132.85	
College	5.545	458.45	469.73	144.67

Panel B: Actual data

		Difference with (% change)		
	Wage level[a]	No senior high	Non-vocational	Vocational
No senior high	1.634			
Non-vocational	2.638	61.46		
Vocational	2.291	40.17	-13.19	
College	4.305	163.43	63.16	87.94

[a] Wage level is in thousands of 2000 Rupiah.

Indonesian language at home, has completed his schooling 4.38 years ago and is employed in a private firm with five to 19 employees in an urban area on Java island. These constructed estimates show what wage would earn an individual selected at random from the whole sample, had he obtained a given level of schooling. Differences in these constructed values across schooling levels measure returns to additional schooling. For comparison, panel B shows actual sample data.

First column of table 1.7 shows wage levels in thousands of 2000 Rupiah. The rest of the columns show wage differences, in percent, between schooling levels. An average individual would earn 4.5 times more by attending college rather than stopping at non-vocational senior high level or junior high level, and 1.4 times more rather than stopping at vocational senior high level. Were he not to attend college, he would earn 1.3 times more with a vocational senior high degree than with a non-vocational senior high degree. In other words, there are large returns to college and to vocational senior high school. There is no return on non-vocational senior high attendance on its own.

1.6 Conclusion

This paper analyses schooling decisions at two stages of individual's life. The first is decision about senior high attendance, made at the point of completion of mandatory junior high schooling. I differentiate between choice of vocational and non-vocational senior high school. The second is decision to attend college. I then estimate returns to senior high and college education, relative to junior high education.

College education increases individual wage by up to 4.5 times. I also find strong positive premium on vocational, relative to non-vocational, senior high schooling. Sample population responds positively to the wage premium for vocational senior high schooling by choosing to enroll. However, the same does not hold for college attendance.

Father's income, while having no significant effect on individual's decision about senior high attendance, has a strong positive effect on college attendance. Public schools in Indonesia charge much lower tuition than private schools. Government's expansion of national schooling system is focused on primary and secondary schooling. It appears that, as a result, more expensive private schools prevail in post-secondary education market. In other words, while returns to both vocational senior high and college education are positive, difference in public/private composition of schools in the markets for secondary and post-secondary education gives rise to financial constraints to college attendance but not to vocational senior high attendance.

Another evidence of the negative effect of incomplete financial markets on individual's schooling choice is significance of the opportunity cost of senior high education. This result demonstrates that while there are positive returns to extra schooling, individuals prefer current pecuniary gains over these future returns, suggesting lack of ability to smooth consumption over time.

Low response to positive returns on college education might be exacerbated by the wage premium for vocational senior high schooling. The data provide no evidence of significant self-selection bias for vocational senior high graduates. In other words, individuals enroll in vocational senior high programs in response to the unconditional wage premium for vocational relative to non-vocational senior high programs, not because of unobserved personal preferences for vocational schooling. Obtaining admission to college is much more challenging for a graduate of vocational, rather than non-vocational, senior high. As a result, individuals wanting to take advantage of wage premium on vocational training in essence make decision to not attend college at least three years before they actually graduate from senior high school.

Incorporation of individual's schooling decisions at various points in individual's life into estimation of returns to schooling has several benefits. By construction, it allows inclusion of larger part of the sample into estimation, by not limiting it to, say, senior high school graduates only. While this might be of small consequence in case of countries with large high school graduation rates, the resulting change in the sample size is substantial for developing countries like Indonesia, where schooling continuation rates are very low, particularly so for post-secondary education.

There are several advantages to using family panel data like IFLS. One is ability to link individual and parents' data not only for each person but also by years. For example, in the logit estimations of schooling decision equations, I use data on father's employment, including income, in both senior high and college attendance equations. However, I am able to use measures of father's employment data for year closest to when individual was making each schooling decision, and so incorporate any changes in family's income between individual's junior high and senior high graduations. I am similarly able to use location data corresponding to individual's residence at the time of each schooling decision, rather than using the same location data in both cases.

Chapter 2

Impact of Climate Change on Rice Production in Thailand

John Felkner (National Opinion Research Center)

Kamilya Tazhibayeva

Robert Townsend (MIT)

2.1 Introduction

Our goal is to evaluate crop yield impacts from likely climate changes for Southeast Asia. To do so we integrate soil science crop modeling, weather simulators, and global climate change models with an economic model of multi-stage rice production. The economic model is estimated with detailed monthly data on inputs, operations, and environmental data over a five year period. We then forecast impacts on yields under two different future climate change scenarios, one assuming high future global anthropogenic[1] pollution emissions, and the other assuming low.

We compare results of the integrated economic model with those of a biophysical model,

[1] Of, relating to, or resulting from the influence of human beings on nature.

inputting into both the stochastic realizations of a weather generator, calibrated against the present, no climate change benchmark and against the two, mild and severe, climate change scenarios. The more realistic forecasts from the socio-economic model include important farmer behavioral mitigation strategies, whereas forecasts from the biophysical model reflect solely direct effects of climatic changes on plant growth. We discuss both aggregate and average impacts as well as heterogeneity in response across farmers.

This chapter is organized into seven sections. Section 2.2 outlines the economic model. Section 2.3 describes the data. Section 2.4 presents estimates of the economic model and examines sensitivity of yields and estimates to socio-economic variables. Section 2.5 discusses the modeling of climate change for Southeast Asia. Section 2.6 outlines the integration of economic, crop growth, weather and climate models. Section 2.7 presents our results. Section 2.8 concludes the paper.

2.2 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. Crop production is defined by the process of a crop's biological growth. This biological growth consists of distinct, chronologically sequential phases. A crop's need for and responsiveness to a given physical input varies across different growth phases. Depending on the progress of crop growth, the farmer may want to adjust the amounts and types of physical 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. For rainfall shocks, the history up to a given stage is predictive of the future. Other idiosyncratic shocks are not observed. Thus, if using an aggregate single-stage production function, there would be a bias in estimation; that is, production shocks influencing inputs in previous stages are not seen by the econometrician and end up in the overall error term. The farmer also responds to realized production shocks in so far as these alter farmer's expectation of production shocks in future stages, updating his information set.

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 starting stage are realized, and with updated expectations based on history at that point in time.

Let i index the three production stages and let L_i and K_i denote, correspondingly, labor and capital and other inputs in stage i.[2] Let y_i be output of stage i, with y_0 describing initial conditions of production such as plot characteristics. Let e_i be production shock realized during stage i. Then output in stage i is $y_i = f_i\left(y_{i-1}, L_i, K_i\right)\exp\left(e_i\right)$, for $i = 1, 2, 3$, where f_i

[2] To account for several operations performed simultaneously during stage i, L_i and K_i can be thought of as vectors of length J_i, where J_i is the number of operations performed in stage i.

is stage i - specific Cobb-Douglas production function.[3] This three-stage production process is illustrated in figure 2.1. Substituting in recursively for intermediate outputs, we obtain a composite production function which describes final harvest as a function of initial plot conditions, inputs, and realized production shocks: $y_3 = f\left(y_0, \{L_i, K_i, \exp(e_i)\}_{i=1}^{3}\right)$.

This approach incorporates the two distinct manifestations of sequential nature of crop production. One is a forward effect, where production shocks and input decisions from earlier stages affect crop-plot conditions and therefore input decisions at later stages. The other is a backward effect, where input decisions at earlier stages are influenced in turn by their expected effect on inputs in subsequent stages.

The order of evens in each stage i is as follows. First, input decisions L_i and K_i are made based on the history of production shocks $\{e_1, ..., e_{i-1}\}$ and intermediate outputs $\{y_0, y_1, ..., y_{i-1}\}$ realized in previous stages. That is, input decisions L_i and K_i are made before stage i shocks are realized. Next, production takes place and inputs L_i and K_i are used at the same time as production shocks for the current stage, e_i, are realized. At the end of the stage, output for the current stage, y_i, is observed.

At each stage, farmer chooses inputs to maximize expected profits[4]. Let p denote the price of final output, w_i denote wage rate for labor used in stage i, and r_i denote price of non-labor input used in stage i. Assume the farmer knows all current and future input prices for a given growing season, as well as final output price[5]. At the beginning of stage i, farmer

[3] Values of inputs, outputs and production shocks are plot-specific. Plot indexing is omitted for simplicity of presentation. We use the Cobb-Douglas function but will explore other specifications in future work.

[4] 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 and effective social networks that enable consumption smoothing and thus effectively approximate Arrow-Debreu institutions.

[5] 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, exports measured in tons). To this extent, p, the price of final rice harvest, is determined on the world market. Thai villages are 'open economies' in the sense that there exist large flows of production factors, including labor, between both villages in the same province and between provinces. These flows criss-cross between rural and urban areas and go in both directions. Circular migration is prominent. As a result, factor prices are set on a large regional scale, and rice cultivating households take these prices as given. Perfect foresight of price is however

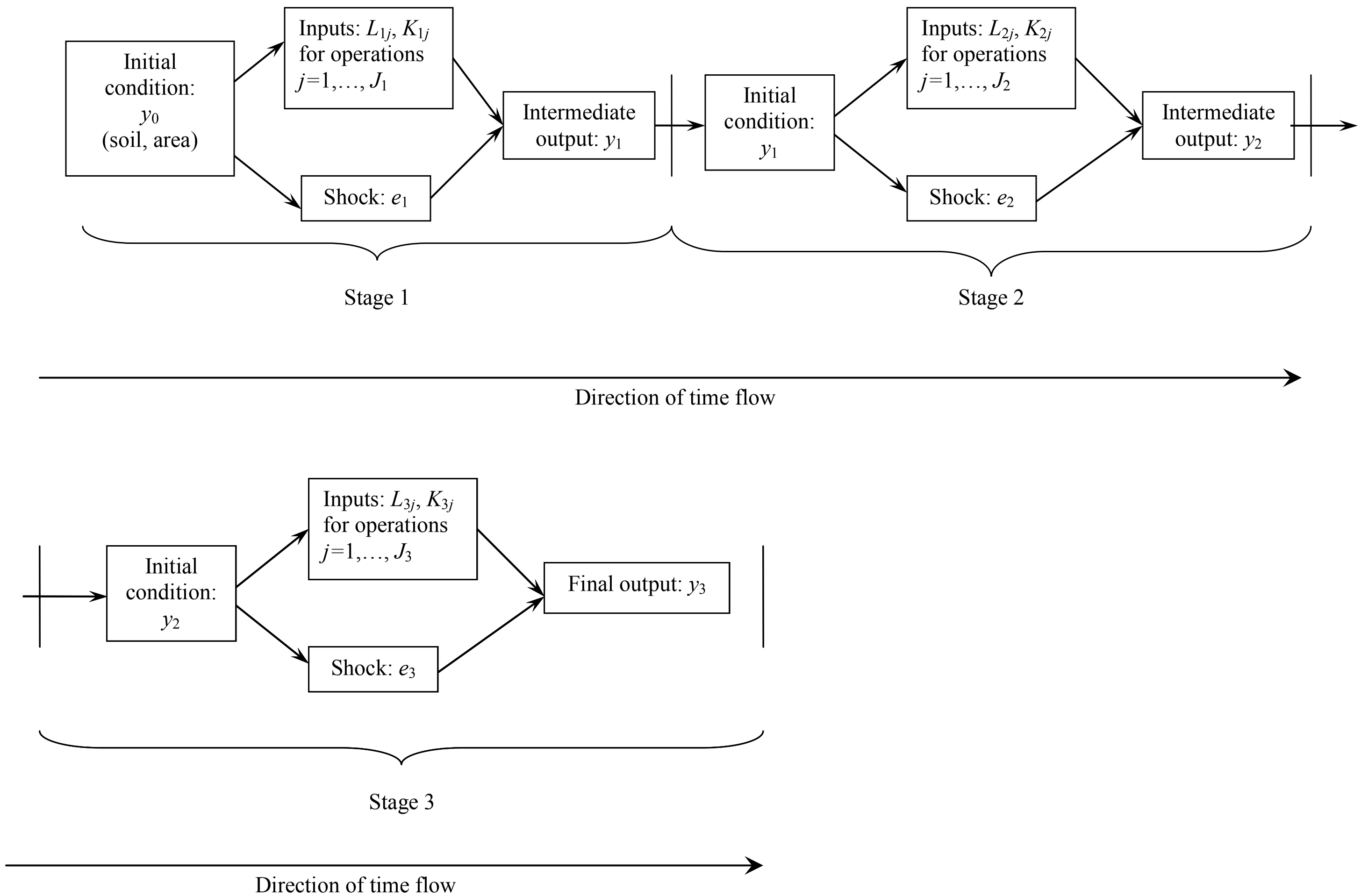

Figure 2.1: Three Stages of Production Process

solves:

$$\underset{L_i,K_i}{Max} E_i\left[\pi\right] = pE_i[y_3] - w_iL_i - r_iK_i - \sum_{j=i+1}^{3}\left(w_jE_i\left[L_j\right] + r_jE_i\left[K_j\right]\right),$$

where expectation $E_i\left[\right]$ is based on the information set available to the farmer at the beginning of stage i, before stage i production shocks e_i are realized[6]. At this point in the production process, the farmer does not yet know all information that determines actual amounts of inputs that will be used in future stages - namely, he does not yet know the size of production shocks that will be realized during stage i and subsequent stages. The farmer chooses optimal levels of stage i inputs based on expected values of input levels in future stages, where the expectation is computed over the information set available to the farmer at the beginning of stage i, before stage i shocks are realized.

Deriving the first-order conditions, we get:

$$\begin{aligned} \text{wrt } L_i \quad : \quad & p\frac{\partial E_i[y_3]}{\partial L_i} = w_i + \\ & + \sum_{j=i+1}^{3}\left(w_j\overbrace{\frac{\partial E_i[L_j]}{\partial y_{j-1}}\frac{\partial E_i\left[y_{j-1}\right]}{\partial y_i}\frac{\partial y_i}{\partial L_i}}^{\text{Backward sequential effect}} + r_j\frac{\partial E_i[K_j]}{\partial y_{j-1}}\frac{\partial E_i\left[y_{j-1}\right]}{\partial y_i}\frac{\partial y_i}{\partial L_i}\right), \end{aligned} \tag{2.1}$$

$$\begin{aligned} \text{wrt } K_i \quad : \quad & p\frac{\partial E_i[y_3]}{\partial K_i} = r_i + \\ & + \sum_{j=i+1}^{3}\left(w_j\frac{\partial E_i[L_j]}{\partial y_{j-1}}\frac{\partial E_i\left[y_{j-1}\right]}{\partial y_i}\frac{\partial y_i}{\partial K_i} + r_j\frac{\partial E_i[K_j]}{\partial y_{j-1}}\frac{\partial E_i\left[y_{j-1}\right]}{\partial y_i}\frac{\partial y_i}{\partial K_i}\right). \end{aligned} \tag{2.2}$$

The marginal cost of each input in stage i has two components. One is increases in current expenses on the input, measured by factor price. Another is change in future expenses on inputs in subsequent stages $j > i$ that will be caused by adjustment of optimal levels of

only an approximation made for analytic tractability.

[6]Note that expectations are indexed by the stage in which they are made, not by the stage corresponding to the most recent shock observable by the farmer when she or he makes the expectation.

stage j inputs with respect to change in actual levels of stage i inputs. Thus the marginal product of all intermediate inputs reflects sequential nature of multistage production process and captures both their immediate direct effect on the crop growth as well as future indirect effects through levels of future inputs.

Now decompose production shock in stage i, e_i, into a rain component, η_i, and a non-rain component, ε_i :

$$e_i = \eta_i + \varepsilon_i.$$

Realized rainfall is observed by the econometrician. The non-rain component ε_i is observed only by the farmer but not by the econometrician. The non-rain component consists of shocks such as pest infestation and plant illness, which are hard to predict. Assume therefore that $E[\varepsilon_i] = 0$. Then farmer's expectation of unrealized production shocks is equal to his rainfall expectation[7]:

$$E_i[e_j] = E_i[\eta_j] \ \forall\, j \geq i.$$

Let vector x_i denote all inputs used in stage i: $x_i' = [L_i'\ K_i']$. In each stage i, crop-plot condition or "output" y_i is a function of output from the previous stage, y_{i-1}, current inputs, x_i, and the realization of production shocks in this stage, $e_i = \eta_i + \varepsilon_i$:

$$y_i = f_i\left(y_{i-1}, x_i, \eta_i, \varepsilon_i\right). \tag{2.3}$$

Choice of inputs for stage i, x_i, depends on the output from the previous stage, y_{i-1}, and expectations of not yet realized production shocks, $E_i\left[\{e_j\}_{j=i}^3\right] = E_i\left[\{\eta_j\}_{j=i}^3\right]$:

$$x_i = x_i\left(y_{i-1}, E_i\left[\{\eta_j\}_{j=i}^3\right]\right).^8 \tag{2.4}$$

[7]Recall that stage i expectations $E_i[]$ are formed and input decisions are made based on the information set available to the farmer at the beginning of stage i, before stage i production shocks e_i are realized.

[8]Input decisions also depend on the amounts of other inputs used in the same stage, input prices in current and future stages, as well as price of the final output. We assume that these variables are implicitly included in the current equation and omit specifying them explicitly for clarity of presentation.

From equation (2.3), output in a given stage j depends on realizations of production shocks up to the end of that stage, $\{\eta_k, \varepsilon_k\}_{k=1}^{j}$. Therefore, inputs x_i in each stage i depend on realizations of production shocks in all previous stages, $\{\eta_k, \varepsilon_k\}_{k=1}^{i-1}$. To demonstrate this, start with stage 1. Using equations (2.3) and (2.4), inputs and intermediate output in stage 1 are given, respectively, by

$$x_1 = x_1\left(y_0, E_1\left[\left\{\eta_j\right\}_{j=1}^{3}\right]\right) \tag{2.5}$$

and

$$y_1 = f_1\left(y_0, x_1\left(y_0, E_1\left[\left\{\eta_j\right\}_{j=1}^{3}\right]\right), \eta_1, \varepsilon_1\right). \tag{2.6}$$

Similarly, using equations (2.3) and (2.4), stage 2 inputs are

$$x_2 = x_2\left(y_1, E_2\left[\left\{\eta_j\right\}_{j=2}^{3}\right]\right), \tag{2.7}$$

and stage 2 intermediate output is

$$y_2 = f_2\left(y_1, x_2\left(y_1, E_1\left[\left\{\eta_j\right\}_{j=2}^{3}\right]\right), \eta_2, \varepsilon_2\right). \tag{2.8}$$

Substituting in expression for y_1 from equation (2.6) above into x_2 equation (2.7), we get

$$x_2 = x_2\left(y_0, \overbrace{x_1\left(y_0, E_1\left[\left\{\eta_j\right\}_{j=1}^{3}\right]\right)}^{x_1}, \eta_1, \varepsilon_1, E_2\left[\left\{\eta_j\right\}_{j=2}^{3}\right]\right),$$

or

$$x_2 = \tilde{x}_2\left(y_0, x_1, \eta_1, E_2\left[\left\{\eta_j\right\}_{j=2}^{3}\right], \varepsilon_1\right). \tag{2.9}$$

Substituting in expression for y_1 from equation (2.6) above into y_2 equation (2.8), we get

$$\begin{aligned} y_2 &= f_2 \Bigg(\underbrace{y_0, \overbrace{x_1 \left(y_0, E_1 \left[\{\eta_j\}_{j=1}^{3} \right] \right)}^{x_1}, \eta_1, \varepsilon_1,}_{y_1} \\ &\quad \overbrace{x_2 \left(y_0, x_1 \left(y_0, E_1 \left[\{\eta_j\}_{j=1}^{3} \right] \right), \eta_1, \varepsilon_1, E_2 \left[\{\eta_j\}_{j=2}^{3} \right] \right)}^{x_2}, \eta_2, \varepsilon_2 \Bigg), \end{aligned}$$

or

$$y_2 = \tilde{f}_2 \left(y_0, x_1, x_2, \eta_1, \eta_2, \varepsilon_1, \varepsilon_2 \right).$$

In the same manner, using equations (2.3) and (2.4), stage 3 inputs and output are, respectively,

$$x_3 = x_3 \left(y_2, E_3 \left[\eta_3 \right] \right) \tag{2.10}$$

and

$$y_3 = f_3 \left(y_2, x_3 \left(y_2, E_3 \left[\eta_3 \right] \right), \eta_3, \varepsilon_3 \right). \tag{2.11}$$

Substituting in recursively for intermediate outputs, we get

$$x_3 = \tilde{x}_3 \left(y_0, x_1, x_2, \eta_1, \eta_2, E_3 \left[\eta_3 \right], \varepsilon_1, \varepsilon_2 \right), \tag{2.12}$$

and

$$y_3 = \tilde{f}_3 \left(y_0, x_1, x_2, x_3, \eta_1, \eta_2, \eta_3, \varepsilon_1, \varepsilon_2, \varepsilon_3 \right). \tag{2.13}$$

Equation (2.13) is the composite production function for final output, y_3, expressed in terms of time 0 initial conditions and inputs and production shock realizations from all three stages. The composite production function (equation 2.13), stage 1 input demands (equation 2.5), stage 2 input demands (equation 2.9), and stage 3 input demands (equation

2.12) constitute the model for crop production.

Composite production function (equation 2.13) includes realizations of production shocks ε_1, ε_2, and ε_3. As equations (2.9) and (2.12) illustrate, inputs x_i on the right-hand side of the composite production function equation also depend on these unobserved production shocks: we see that x_2 depends on ε_1 and that x_3 depends on both $\varepsilon_1$and ε_2. In the estimation, unobserved realized production shocks ε_i form the error term. That is, error term in the composite production function (2.13) is composed of ε_1, ε_2, and ε_3; error term in stage 2 input demands equation (2.9) is a function of ε_1; and error term in stage 3 input demands equation (2.12) is composed of ε_1 and ε_2. Therefore, explanatory variables in the composite production function are endogenous with respect to the error term.

This endogeneity of inputs introduces bias in OLS estimates of the composite production function (equation 2.13). Let z denote the vector of explanatory variables in the equation for the composite production function. Vector z consists of all inputs used in all three production stages, as well as all realized rainfall shocks: $z' = [x_1' \ x_2' \ x_3' \ \eta_1 \ \eta_2 \ \eta_3]$. Let ε denote the error term, composed of the unobserved production shocks realized throughout the growth cycle: $\varepsilon = \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2 + \alpha_3\varepsilon_3$, with $\alpha_i \neq 0$ for $i = 1, 2, 3$. Consider a linearized form of the composite production function equation (2.13), and let vector β denote the corresponding coefficients: $\bar{y}_3 = \beta\bar{z} + \varepsilon$. In our case of Cobb-Douglas production functions, $\bar{y}_3 = \ln(y_3)$ and $\bar{z} = \ln(z)$. Since inputs x_i depend on ε_i, $E[\bar{z}'\varepsilon] \neq 0$, and OLS estimation of β will produce biased results:

$$E\left[\hat{\beta}_{OLS}\right] = \beta + (\bar{z}'\bar{z})^{-1} E[\bar{z}'\varepsilon] \neq 0,$$

where $\hat{\beta}_{OLS}$ denotes the OLS estimate of true coefficient vector β.

One may use instruments to correct the bias in coefficient estimates. However, endogeneity bias is only one of the consequences of the fact that unobserved production shocks enter the error terms of both the composite production function (equation 2.13) and input demands (equations 2.9 and 2.12). Another consequence is the correlation of error terms across equations for input demands and the composite production function. This makes single equation estimation of the composite production function less efficient compared to the system estimation of all equations in the model, as the former approach does not account for the correlation of error terms across equations.

As equations (2.7) and (2.10) show, input demands depend on both realizations and expectations of production shocks. In other words, this framework incorporates the effects of changes in both realized and expected production shocks on production practices and final output. This makes it well-suited for studying the effect of climate change on farmers' cultivation practices.

We next describe our data and how we measure expectations of production shocks and intermediate output levels.

2.3 Data

Our data come from the Townsend Thai Project[9] (see Paulson, Townsend, and Karaivanov (2006)). We focus on rice farmers in four villages in Sisaket province, located in predominantly rural and poor northeastern region of the country. Figure 2.2 shows location of Sisaket province in Thailand. 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). Data are collected monthly at a household-plot level, with many households cultivating several plots in a given year. We use an unbalanced five-year panel for 1998-2002 on 137 households, with a total of 826 crop-plot observations over five years. Table 2.1 shows village-level averages of number of years and plots per year in the data. On average, we have data for about three and a half growth cycles per household, with two crop-plots per cycle.

The data include information on usage and cost of labor, equipment, and other non-labor inputs used in separate production operations. We also have sets of measures of plot soil

[9] Detailed description of the project can be found at Thailand Database Research Archive (2010).

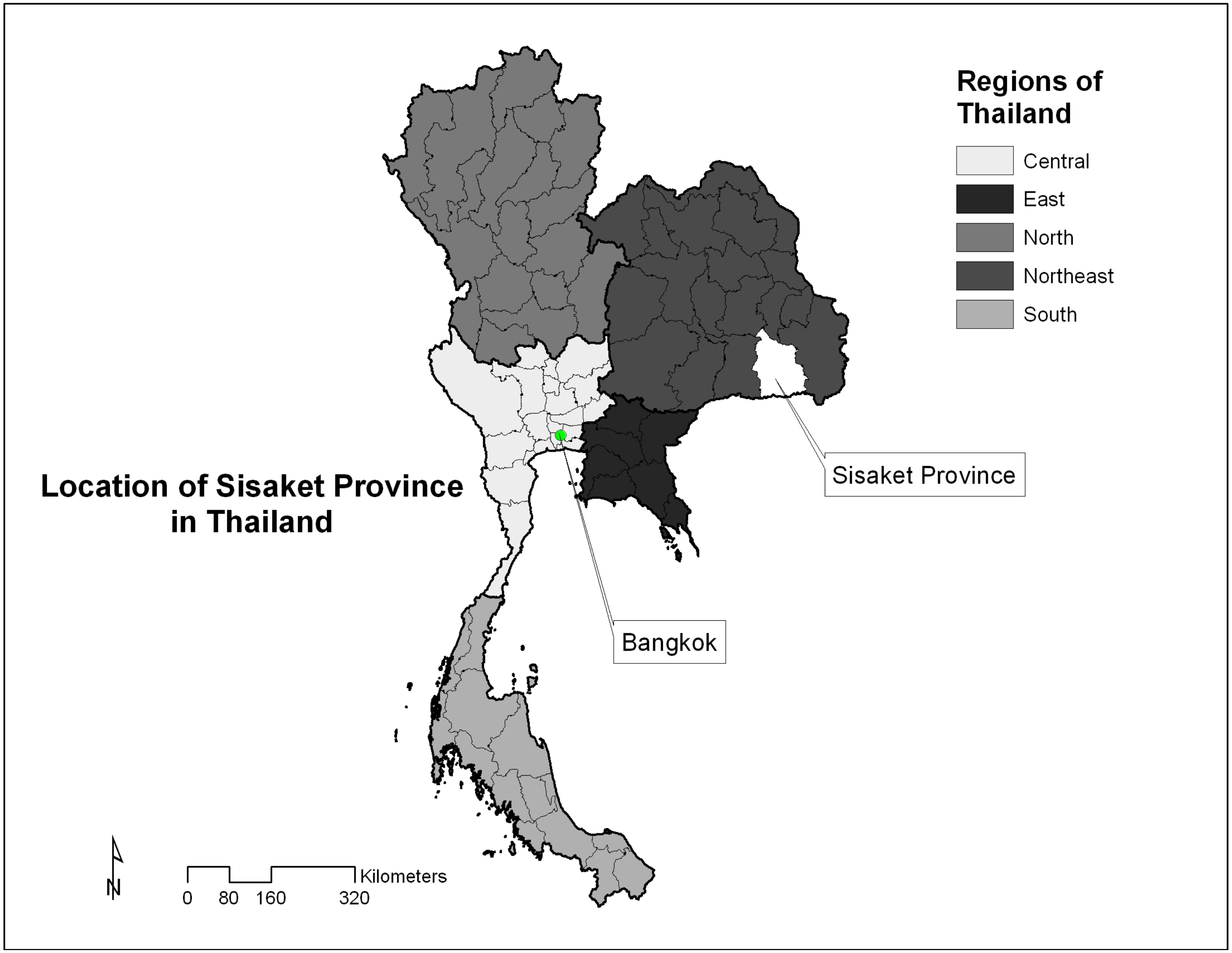

Figure 2.2: Location of Sisaket Province in Thailand

Table 2.1: Number of Crop-Plots and Growth Cycles per Household

	Mean	St. Dev.	Min	Max
Number of crop-plots				
Province	1.84	1.13	1	7
Village 1	2.26	1.31	1	6
Village 2	1.63	0.95	1	6
Village 3	1.70	1.08	1	7
Village 4	1.76	1.03	1	5
Number of growth cycles				
Province	3.61	0.90	1	5
Village 1	3.79	0.95	1	5
Village 2	3.45	0.72	1	4
Village 3	3.63	1.00	1	5
Village 4	3.61	0.90	1	5

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. Farmers also provide their estimate of final grain harvest from this plot given the information they have at the time of the interview.

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. More-over, 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 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. To determine which operations, besides the obvious preparatory operations, take place in stage 1, we see which operations are performed simultaneously with preparatory operations. We define two operations as performed simultaneously if both were performed since the last interview. Note that because households in our data are interviewed at 30 days intervals, in practical terms simultaneous operations are performed within the same 30 day period. Similarly, to determine which operations take place in stage 3, we see which operations are performed simultaneously with terminal operations.

Time period between stages 1 and 3 constitutes stage 2. Correspondingly, operations performed between stages 1 and 3 fall into stage 2. Our data indicate that application of chemical fertilizer is generally performed both in stage 1 during planting and in stage 2 during plant development. Other plant care operations are performed in stage 2 only. Because fertilizing takes place in both stages 1 and 2, each physical input used in fertilizing - amounts of chemical fertilizer and manure and hours spent on fertilizing - constitutes two different production inputs, one corresponding to stage 1 application and the other corresponding to stage 2 application.

Table 2.2 shows variation in stage duration and timing across years. As noted earlier, we determine the timing of stages individually for each plot in each cycle. The first two columns show the mean and standard deviation of stage length, in number of months. The last four columns show stage timing in terms of calendar months. For each stage, column four shows the earliest starting month in the sample, column five shows the average starting month, and columns six and seven show, respectively, the average and the latest ending month. Stage 1 is on average just over one month long and typically occurs in July. Stage 2 is on average one to one and half months long and typically occurs from August to September. Stage 3 is on average about two and a half to three months long and typically occurs from October to November. Both stage duration and timing vary from year to year.

Table 2.3 lists cultivation operations performed and inputs used in each stage. There are several types of inputs into rice production. They can be divided into four groups: land, labor, equipment and, finally, non-labor and non-equipment production factors such as seeds

Table 2.2: Duration and Timing of Stages

	Length (months)		Calendar Month: Starting Month		Calendar Month: End Month	
	Mean	St. Dev.	Min	Mean	Mean	Max
1998						
Stage 1	1.00	0.00	8	8	8	9
Stage 2	1.18	0.39	9	9	9	10
Stage 3	2.33	0.65	10	10	12	12
1999						
Stage 1	1.18	0.41	5	6	7	8
Stage 2	1.69	0.74	6	8	8	10
Stage 3	3.12	0.82	8	9	11	1
2000						
Stage 1	1.16	0.38	5	7	7	8
Stage 2	1.55	0.62	6	8	9	10
Stage 3	2.49	0.73	8	10	11	12
2001						
Stage 1	1.17	0.38	5	7	7	8
Stage 2	1.38	0.60	6	8	9	10
Stage 3	2.60	0.76	9	10	11	12
2002						
Stage 1	1.21	0.41	5	7	7	9
Stage 2	1.55	0.77	6	8	9	10
Stage 3	2.29	0.62	8	10	11	12

and fertilizer. For one of stage 1 operations, soil preparation and plowing, we also know if different types of activities, such as multiple plowing, were performed. Table 2.3 lists these activities. We only know which types of activities were performed on a given plot, not the amount of time spent on them. We use indicator variables to control for these activities.

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 to take the possibility of adverse rainfall shocks into account

Table 2.3: Cultivation Operations and Input Usage by Stages

Stage 1

Operations	Non-labor inputs	Equipment	Activities
Soil Preparation and Plowing		Walking tractor Water buffalo	Vegetation clearing First plowing Multiple plowing
Planting or transplanting	Seeds Seedlings		
Fertilizing	Chemical fertilizer Manure		

Stage 2

Operations	Inputs	Equipment	Types of activities
Weeding or thinning			
Fertilizing	Chemical fertilizer Manure		

Stage 3

Operations	Inputs	Equipment	Types of activities
Harvesting		Harvesting machine	
Collection for threshing		Walking tractor	
Threshing		Shelling machine Threshing machine	
Transport to storage		Walking tractor Small four wheel tractor Truck	

when making input decisions. We use historic village-level daily rainfall to construct a measure of expected future rainfall at the beginning of each production stage. Although rainfall is an aggregate shock, expected rainfall varies across plots due to variation in stage timing. Soil type and slope also impact soil moisture, the key latent variable.

We construct expectations of monthly rainfall based on 12 monthly lags as well as year and month dummies. For each plot we then compute aggregate rainfall expectation for each of the three production stages. Let i and j index calendar months, with j indicating the month when rainfall expectation is made and i indicating the month whose rain is estimated. For example, if expectation of July rainfall is made in June, $i = 7$ and $j = 6$. For each village rainfall series, we estimate the following equation:

$$
\begin{aligned}
rain_{i,j} \quad &= \quad \sum_{k=i-j}^{12} \beta_k rain_k + \sum_{k=1973}^{2002} \gamma_k \mathrm{I}\left(year = k\right) + \\
&\quad + \sum_{k=4}^{10} \delta_k \mathrm{I}\left(calendar\ month = k\right) + error_{i,j},
\end{aligned}
\tag{2.14}
$$

where a_k means kth monthly lag of variable a and $\mathrm{I}(A)$ is an indicator function for event A. In our data, index j, which indicates how much in advance the expectation is made, ranges from 1 to 8. Larger values of j correspond to farmer's rain expectations for stages 2 and 3 made at the beginning of stage 1. For each village, we estimate equation (2.14) for $j = 1, ..., 8$, and then use these estimates to construct crop-plot specific rainfall expectations. This approach takes into account difference in stage timing across plots, as well as the fact that farmers have to make input decisions at the start of each stage. For example, if for a given plot stage 1 went from June through August, farmer had to make input decisions for stage 1 before June rainfall was realized. We therefore construct expectation of June rainfall based on 12 monthly lags starting with May, expectation of July rainfall based on 11 monthly lags starting with May, and expectation of August rainfall based on 10 monthly lags starting with May.

Figure 2.3 graphs monthly rainfall in each of the five years, by village. Rainfall was the lowest in 1998. Rainfall was also low in 2002, and had an atypical late peak in September. Year 1999 had the earliest onset of rainfall. Year 2000 was the most abundant and also had an early onset. Year 2001 had average rainfall, both in terms of quantities and timing. Figure 2.4 is a box plot version of table 2.2 information on variation in stage timing across years. In 1999, which had the earliest onset of rainfall, planting started much earlier than in other years. In 2002, when rainfall was low and had a very late peak in September, median timing of stages 2 and 3 occurred one month later than in the average rainfall year 2001. Table 2.4 shows average yield per acre in each of the five years, by village. In 1998, which had the lowest rainfall, yields are significantly lower than in other years. In 2000, when rain was most abundant, yields were the highest or second highest in three villages. This illustrates the effect rainfall has on yields and on farmers' production decisions.

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. Figure 2.5 shows location of plots in the four surveyed villages relative to the rest of the province and figure 2.6 shows a close-up on plot locations. The four villages are located near each other. As a result, distributions of plots for villages overlap, so 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.

For a number of households, the survey collects data on soil quality for one of the plots.

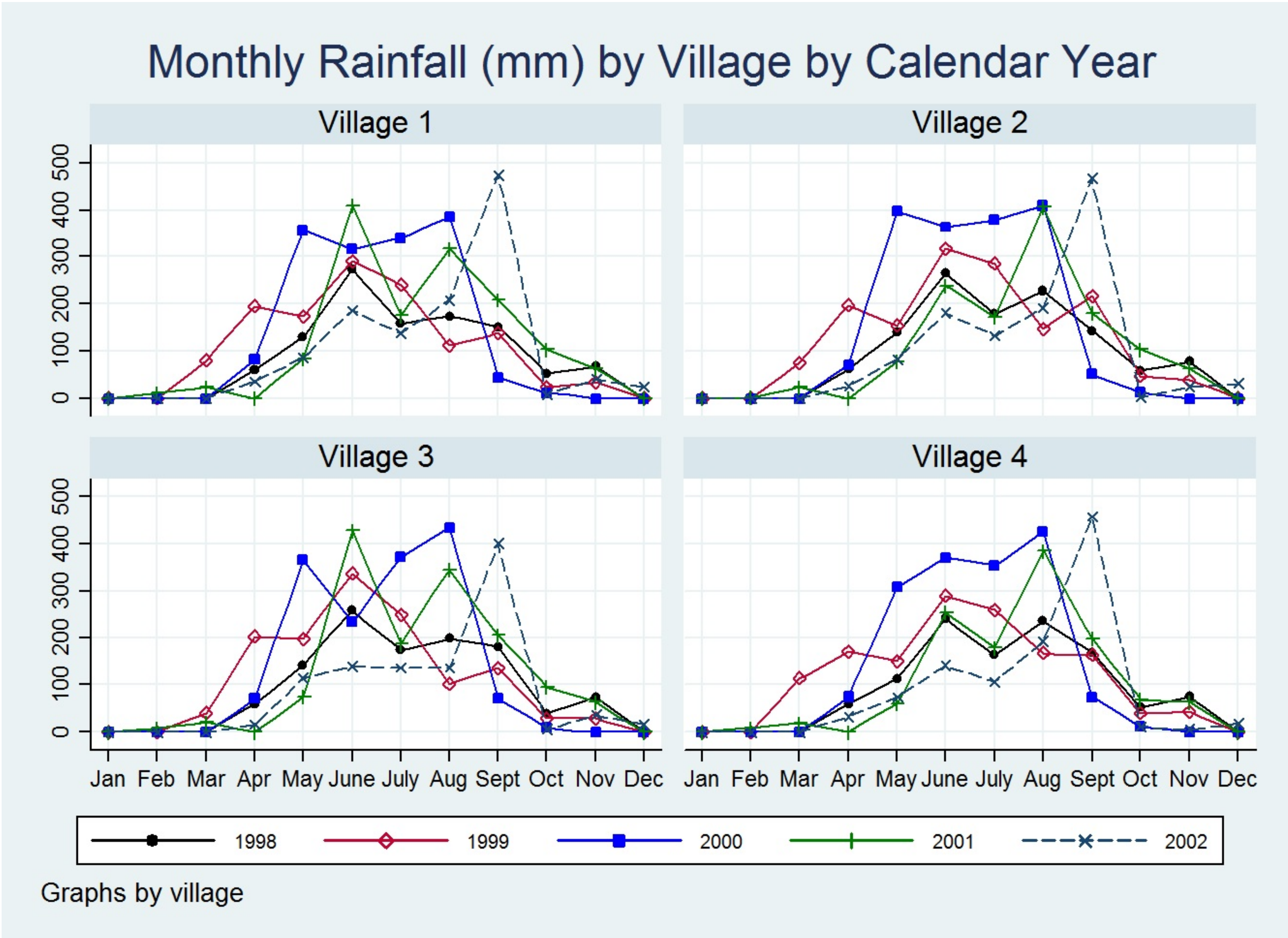

Figure 2.3: Monthly Rainfall by Village by Calendar Year

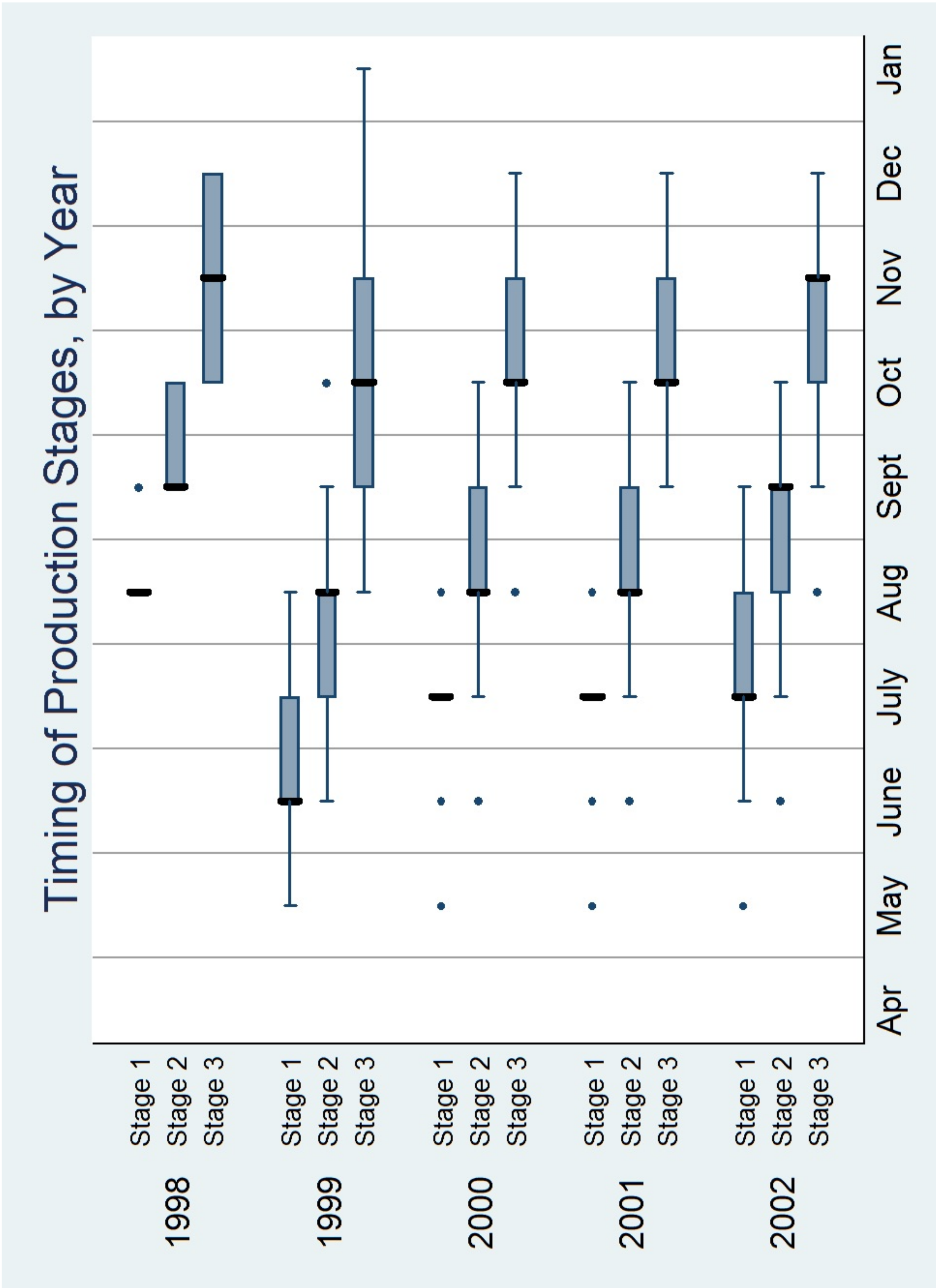

Figure 2.4: Timing of Production Stages, by Year

Table 2.4: Average Yield, by Growth Cycle

	Yield, kg/acre				Number of Obs	
	Mean	St. Dev.	Min	Max	Crop-Plots	Households
Province					826	137
1998	481	362	32	1,779	33	28
1999	760	386	38	2,070	201	115
2000	809	352	61	2,100	206	110
2001	710	300	41	2,372	190	98
2002	713	249	198	1,797	196	98
Village 1					273	35
1998	466	351	113	1,112	9	6
1999	558	278	109	1,269	52	29
2000	719	330	268	1,730	65	30
2001	606	189	135	1,186	72	29
2002	639	199	324	1,362	75	27
Village 2					227	43
1998	602	426	32	1,779	15	13
1999	762	369	38	2,070	76	41
2000	769	322	61	1,504	52	33
2001	850	476	41	2,372	37	25
2002	782	257	303	1,797	47	27
Village 3					199	37
1998	281	123	93	436	5	5
1999	992	390	412	1,977	40	27
2000	965	395	185	2,100	64	31
2001	790	245	297	1,433	49	28
2002	890	268	198	1,557	41	26
Village 4					127	22
1998	309	135	185	500	4	4
1999	792	412	185	1,799	33	18
2000	728	211	434	1,369	25	16
2001	662	220	344	1,318	32	16
2002	565	133	222	884	33	18

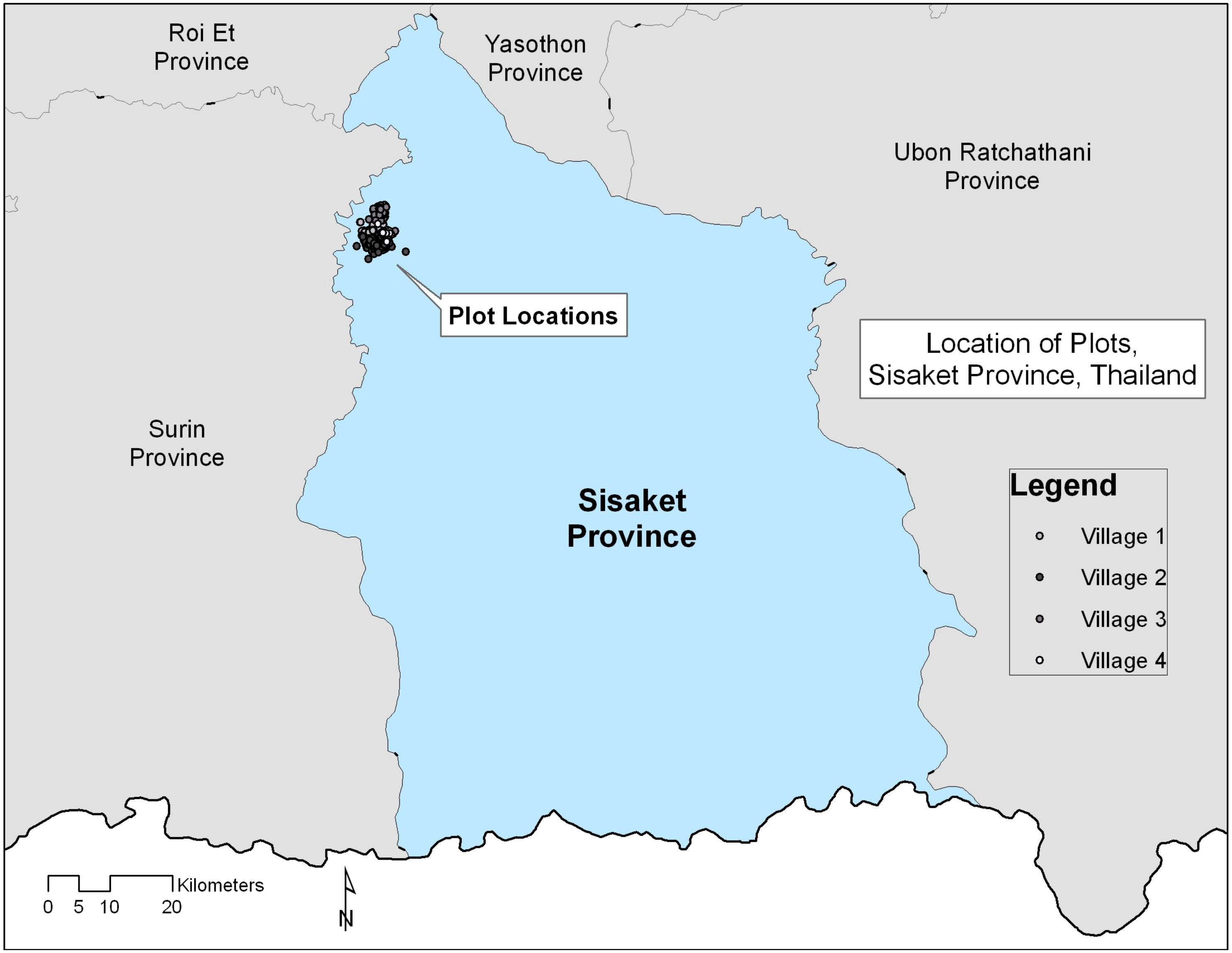

Figure 2.5: Location of Plots in Sample Villages

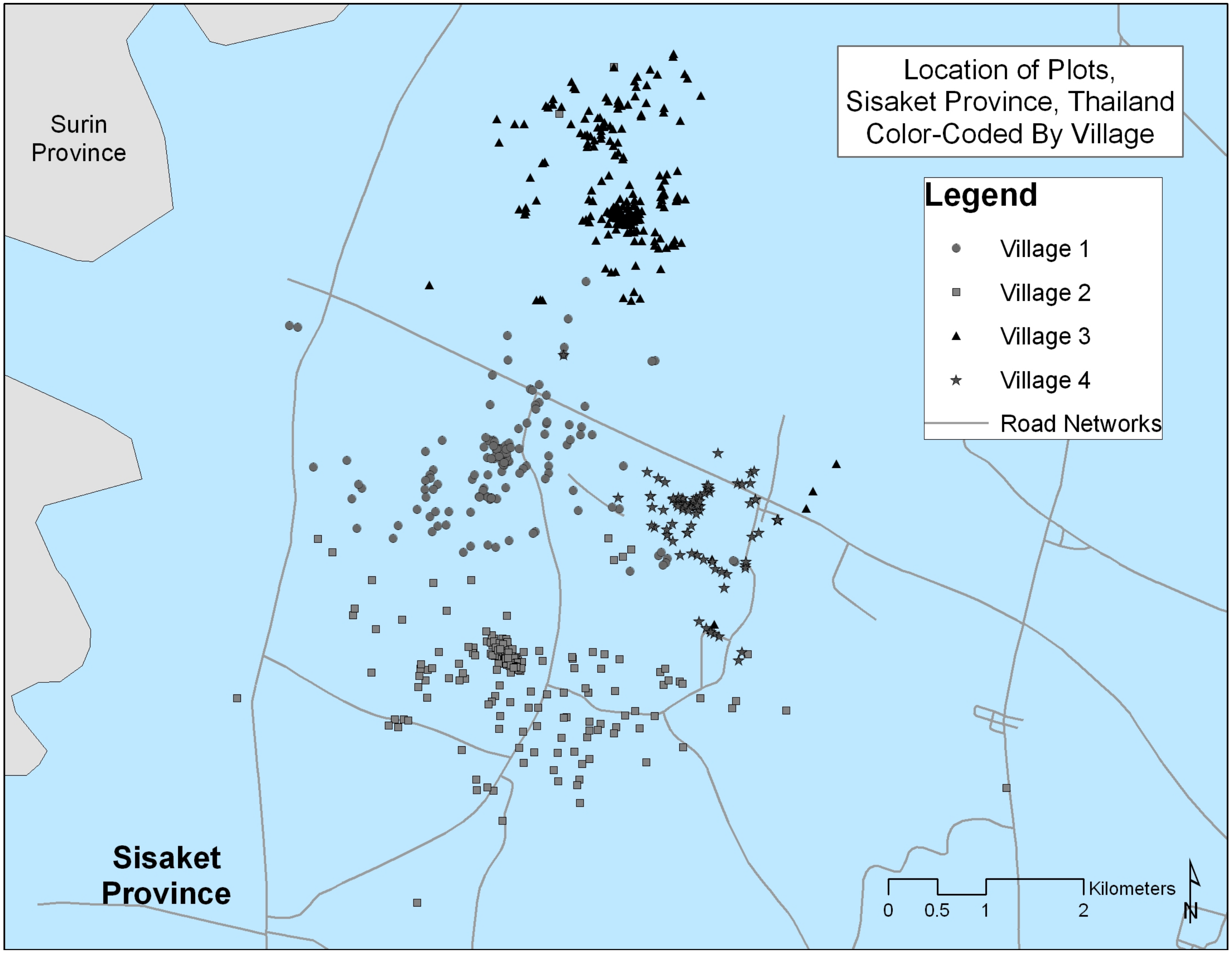

Figure 2.6: Zoom In on Plot Locations in Sample Villages

We also know location coordinates of each plot for every household. For plots with no soil data, we link each plot to the geographically closest plot for which we have soil variables, and use them as measures of soil quality. 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 quality variables are summarized in table 2.5. Levels of soil pH in our sample of plots range from very acidic to relatively neutral, so higher values of pH level correspond to better soil conditions. Soil variables describe initial conditions of rice production, corresponding to y_0 in terms of section 2.2 notation.

Labor variables include labor input, measured in total hours. Typically there are several groups of laborers working on a given plot. They include household members who work on their own plot; villagers outside of the household, both relatives and non-relatives, who work for free, for labor exchange, or for pay; and workers hired through a broker, usually in teams. In other words, there are several ways for a household to adjust its labor input at a given point of rice production in response to realization of intermediate output from the previous stage or production shocks. We construct measures of labor input and the corresponding wage rates separately for each operation in every production stage.

We start with the wage rates of different groups of workers for each operation in every stage. For workers that work for pay, both contracted directly by the household and hired through a broker, we know the total cost of their services and thus can compute the corresponding wage rates[10]. To compute wage rates for labor input by household members and individuals working for labor exchange, we use wage rates of household members which they earn outside of rice production. We use these wages of household members as mea-

[10] We treat as missing observations with 0 wage rate. To identify outliers, we construct average wage rate across all operations in each month. We then impose the 99th percentile of this average wage as the upper bound on operation-specific wages. Finally, we replace missing wage observations with province average, by operation. We perform these calculations separately for labor hired directly by the household and labor hired through a broker.

Table 2.5: Measures of Soil Quality

Variable name	Characteristic	Mean	St. Dev.	Percentile 5th	Percentile 95th
Soil pH, from 0 (acidic) to 14 (alkaline or basic), 7 is neutral	Affects solubility of nutrients - best level for healthy. plant growth is about 6.3-6.8. When pH falls below 5.5, most major plant nutrients become unsoluble.	5.763	0.878	4.9	7.5
Cation exchange capacity (capacity to hold cation nutrients)	Determined by the amount of clay and/or humus in the soil, which improves the nutrient and water-holding capacity of the soil	2.331	1.414	1	4.91
Organic matter, %	Helps the soil hold water and supplies nutrients.	0.570	0.438	0.18	1.43
Field capacity, %	The maximum amount of water soil can hold	10.055	2.188	7.59	13.7

sures of value of their time. For each household member participating in rice production, we know the number of hours she or he spent on a given cultivation operation. We then use shares of these household member-specific labor inputs in total labor input by all household members in a given operation in a given stage as weights on non-rice production wages of these household members to compute the average wage rate for labor input by household members[11]. We assume that labor input by group working for labor exchange has to be repaid by household members and so wage rate constructed for labor input by household members is representative of wage rate for non-household labor working for labor exchange. Now that we know wage rates for every group of workers for each operation in every stage, we compute operation- and stage-specific wage rate as the average of three wage rates: for household labor, labor hired directly by the household, and labor hired through a broker.

Equipment variables correspond to machines and physical technology used in production. These include different types of tractors, machines used for spraying, harvesting, and so on. We treat equipment inputs as predetermined because most households use equipment which they already own at the start of production cycle. Other inputs include seeds, seedlings, and different types of plant fertilizers. For non-equipment non-labor inputs, as well as for the price of the final output, we use the village average as the measure of actual price[12]. This reflects the assumption that all households are price takers in inputs and output markets.

To construct a measure of intermediate "outputs", we use DSSAT - a powerful computer crop growth model.[13] The DSSAT system takes in amounts and timing of application of non-labor and non-equipment production factors such as seeds and chemical fertilizer,

[11] For individual wages outside of agriculture, we treat as missing observations with 0 wage rate. We impose the upper bound on individual wage, computed as follows. For each village in a given year, we compute the 95th percentile of individual wage distribution, and the standard deviation for observations below the 95th percentile. The upper bound is the sum of the 95th percentile and this standard deviation. After operation-specific wage for household labor was computed, we replaced missing observations with village average.

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

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

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 great advantage of DSSAT is that it allows us to capture 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 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 root weight as indicators of intermediate output from stage one, and measures of leaf weight, root weight and stem weight as indicators 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.

We now turn to the estimation of the model.

2.4 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 as instruments stage- and operation-specific input prices, as well as observed rainfall and farmer-specific rainfall expectations.[14]

Composite production function is equation (2.13) from section 2.2:

$$y_3 = \tilde{f}_3\left(y_0, x_1, x_2, x_3, \eta_1, \eta_2, \eta_3, \varepsilon_1, \varepsilon_2, \varepsilon_3\right).$$

It expresses final yield as a function of initial production conditions, or plot characteristics, y_0, all inputs used throughout the growth cycle, x_1, x_2, x_3, and all rainfall realizations, η_1, η_2, η_3. As described in the data section above, plot characteristics y_0 include measures of soil quality and area used for cultivation. Rainfall realizations (as well as input amounts) are aggregated by stage according to crop-plot specific stage timings.

In our data, there are multiple inputs in each production stage, so each x_i is a vector. Overall, there are nine different labor demand equations, one for each production operation, and six different non-labor input demand equations (see table 2.3). Three different

[14] To repeat, while we perform very detailed analysis to incorporate heterogeneous timing of stages and thus inputs application across plots and cycles, we do not endogenize timing decisions, but treat them, rather, as predetermined. This is a margin of adjustment that may matter and mitigate climate change effects.

cultivation operations were performed in stage 1: soil preparation and plowing, planting of seeds and transplanting of seedlings, and fertilizing. Soil preparation and plowing operation requires only the labor input. The planting and transplanting operation requires both labor and seeds and seedlings inputs. The fertilizing operation requires both labor and chemical fertilizer and manure inputs. Two different cultivation operations were performed in stage 2: weeding and thinning, and fertilizing. Weeding and thinning requires only labor input. Fertilizing requires both labor and chemical fertilizer and manure inputs. Four different operations were performed in stage 3: harvesting, collection for threshing, threshing, and transport to storage. All stage 3 operations use only labor input. At each stage, inputs are determined simultaneously and also depend on intermediate output, or crop state, from the previous stage, farmer's expectations of production shocks, and real input prices. In addition, equations for labor inputs that use agricultural equipment include measures of equipment usage. For example, equation for labor used in plowing includes measures of equipment used for plowing, such as tractor.

Stage 1 input demands are described by equation (2.5):

$$x_1 = x_1 \left(y_0, E_1 \left[\{\eta_j\}_{j=1}^{3} \right] \right).$$

They depend on plot characteristics, y_0, and expectations of rainfall in each of the three stages, where the expectations are formed at the start of stage 1.

Stage 2 input demands are described by equation (2.9):

$$x_2 = \tilde{x}_2 \left(y_0, x_1, \eta_1, E_2 \left[\{\eta_j\}_{j=2}^{3} \right], \varepsilon_1 \right).$$

Let L_i denote labor inputs used in stage i, and let K_i denote non-labor non-equipment inputs used in stage i. Then vector x_i of stage i inputs is equal to $x_i' = [L_i' \; K_i']$ and we can rewrite equation (2.9) as

$$x_2 = \tilde{x}_2 \left(\underbrace{\overbrace{y_0, K_1, \eta_1}^{\text{DSSAT}_1}, L_1, \varepsilon_1}_{y_1}, E_2 \left[\left\{ \eta_j \right\}_{j=2}^{3} \right] \right).$$

As described in the data section above, we use DSSAT as a measure of intermediate outputs y_1 and y_2 which are not available in the survey data but which are observed by the farmers. DSSAT incorporates measures of soil quality, weather realizations, and usage of non-labor inputs into its simulation procedure. In other words, DSSAT indicators of plant development, measured at the end of stage 1, incorporate the effects of soil quality and stage 1 rainfall and non-labor inputs on plant growth. However, because DSSAT does not incorporate labor inputs or climate-unrelated production shocks, effects of stage 1 labor input L_1 and non-rain production shocks ε_1 on plant growth are not accounted for by DSSAT. Therefore, we include stage 1 labor inputs together with DSSAT measure of plant development into equations for stage 2 input demands. Unobserved shocks ε_1 are part of the error term. Stage 2 input demand equations also include expectations of rainfall in stages 2 and 3, with the expectations formed at the start of stage 2.

Stage 3 input demands are described by equation (2.12):

$$x_3 = \tilde{x}_3 \left(y_0, x_1, x_2, \eta_1, \eta_2, E_3 \left[\eta_3 \right], \varepsilon_1, \varepsilon_2 \right),$$

which can be rewritten as

$$x_3 = \tilde{x}_3 \left(\underbrace{\overbrace{y_0, K_1, \eta_1, K_2, \eta_2}^{\text{DSSAT}_2}, L_1, \varepsilon_1, L_2, \varepsilon_2}_{y_2}, E_3 \left[\eta_3 \right] \right).$$

By the same logic, we include stage 1 and stage 2 labor inputs together with DSSAT indicators of plant development, measured at the end of stage 2, into equations for stage 3 input demands. Unobserved shocks ε_1 and ε_2 are part of the error term. Stage 3 input demand equations also include expectation of stage 3 rainfall, formed at the start of stage 3.

First-order conditions (2.1) and (2.2) in section 2.2 demonstrate that input's marginal cost depends not only on its own price, but also on changes in input expenditures in future stages caused by adjustments of input usage in future stages in response to input usage in current stage. All input demands therefore include input's own real price and real prices of inputs in all future stages.

We assume the following dynamics in demands for inputs used in the same stage. First, in our data there are two types of endogenous inputs: labor and non-labor non-equipment inputs such as fertilizer. We assume that in each stage decisions about inputs of the same type are made simultaneously. For example, in stage 2, there are two labor inputs: one for the weeding operation and the other for fertilizing operation. Decision about how many hours are spent on weeding is affected by the amount of hours spent of fertilizing, and vice versa. Accordingly, amounts of other labor inputs used in the same stage are included in each labor input demand equation, and amounts of other non-labor inputs used in the same stage are included in each non-labor input demand equation.

Second, for operations that involve both labor and non-labor inputs, such as planting or fertilizing, we assume that decisions about labor and non-labor inputs used in the same operation are made simultaneously, while rainfall expectations affect decision about non-labor inputs. For example, in planting operation, rainfall expectations affect farmer's demand for seedlings directly. Demand for seedlings and demand for labor to be used in planting operation are decided simultaneously. In this way, rainfall expectations affect demand for planting labor indirectly through their effect on the demand for seedlings.

As a consequence of these two assumptions, in a given stage labor input used in operation A and non-labor inputs used in operation B affect one another indirectly through the labor input used in operation B. For example, in stage 2, chemical fertilizer does not enter the equation for labor used in weeding, and labor used in weeding does not enter the equation for chemical fertilizer, but equations for both include labor used in fertilizing, and both in turn enter equation for labor used in fertilizing.

In any given year one household can be cultivating rice on several plots, which can introduce a common factor into timing and input decisions for these plots. We include household indicators as controls for the household fixed effect. We use year dummies to account for general province-wide time trend and village indicators to account for village fixed effects. For year dummies, year 2001 is chosen as comparison group because it had average amounts and timing of rainfall. For villages dummies, village 2 is randomly chosen as comparison group. We include month indicators to keep track of differences in stage timing across farmers and years. For example, stage 1 in our sample can start as early as May and last as late as September. We include dummies for each month between May and September into all stage 1 input demands. This accounts for the effects of earlier/later timing of stages on input usage apart from those due to rainfall. Similarly, we include month dummies for the start of production process into equation for the composite production function. For each group of month indicators, the earliest month is chosen as comparison group. To separate effects of production scale and quantities of inputs used we measure all production factors per acre of plot area under cultivation. We include area under cultivation as a (predetermined) explanatory variable in every equation in the system to capture the scale effect.

We estimate the system of simultaneous equations with three stage least squares (3SLS). As discusses above in section 2.2, estimation of the composite production function has to take into account endogeneity of production inputs and correlation of error terms across equations for input demands and the composite production function. OLS estimation of the composite production function equation (2.13) produces biased results. Endogeneity can be corrected by using instruments. However, single-equation instrumental variable (IV) estimation of the composite production function does not take into account across-equation correlation of error terms when computing standard errors of the estimates. System instrumental variables approach addresses both concerns. 3SLS, also known as full information instrumental variables estimator (FIVE), is a special case of system GMM, with the weighting matrix dependent on the estimate of the variance-covariance matrix of the system error terms. This error variance-covariance matrix for 3SLS is constructed using estimates of the two stage least squares (2SLS). 2SLS, also known as limited information instrumental variables estimator (LIVE), is another special case of system GMM. 2SLS uses variance-covariance matrix of the instruments as the weighting matrix, and its coefficient estimates are equivalent to equation-by-equation IV. 3SLS is asymptotically more efficient than 2SLS. The potential downside to 3SLS is the fact that misspecification of any one equation in the system can lead to biased estimates for the whole system, not only for the misspecified equation. For 2SLS, on the other hand, misspecification of one equation contaminates the estimates only for that equation and does not affect the rest of the system.

Table 2.6 shows the estimates of the composite production function using 3SLS, 2SLS, and OLS. The second column lists explanatory variables by group (soil variables, labor inputs, and so on), and the first column indicates production stages in which each input was used. Columns three and four show, respectively, coefficient estimates and their standard errors obtained from the 3SLS. As expected, higher usage of seeds, seedlings, and fertilizers contributes to higher yields, as indicated by positive and significant coefficients on non-labor inputs. Insignificance of soil quality variables combined with significant and positive coefficients on chemical fertilizer suggests that differences in soil quality are moderated by fertilizer applications. Surprisingly, coefficient on stage 1 rainfall is not significant, while coefficient on stage 2 rainfall is actually negative. Coefficients on year dummies are more intuitive. Recall from figure 2.3 that year 1998 had the lowest rainfall, while rainfall was both most abundant and had an early onset in 2000. Correspondingly, yields are significantly lower in 1998 and significantly higher in 2000.

The case of fertilizing illustrates the importance of timing characteristic of production factors. The coefficients on the amount of chemical fertilizer used in both stage 1 and stage

Table 2.6: the Composite Production Function
(equation for the final yield)

	3SLS (benchmark)		2SLS		2SLS = 3SLS	OLS		OLS = 3SLS
Explanatory variable	Coefficient	St. error	Coefficient	St. error	P-value	Coefficient	St. error	P-value
Land and Soil								
Used land, acre	-0.01532	0.02958	-0.0142	0.03878	0.982	-0.00308	0.01028	0.696
Soil pH	-0.01177	0.04263	-0.02913	0.05609	0.805	-0.00778	0.03441	0.942
Organic matter, %	-0.0149	0.09775	0.01984	0.12784	0.829	0.02094	0.08792	0.785
Field capacity, %	0.02472	0.01897	0.01632	0.02475	0.788	0.01547	0.01773	0.722
Cation exchange capacity	0.02598	0.04177	0.03055	0.05477	0.947	0.00977	0.03585	0.768
Labor Inputs by Operation, ln(hours/acre)								
Stage 1								
Soil preparation and plowing	-0.10370**	0.04652	-0.06074	0.06171	0.578	-0.00103	0.00642	0.029
(Trans)planting	0.02495	0.14747	0.06366	0.19737	0.875	0.04881*	0.0284	0.874
Fertilizing	-0.18674***	0.037	-0.09704**	0.04912	0.145	-0.00872	0.00848	0.000
Stage 2								
Weeding or thinning	-0.02386***	0.00661	-0.01542*	0.0088	0.443	-1.51E-03	0.00152	0.001
Fertilizing	-0.03270*	0.01692	-0.00916	0.0224	0.402	0.00127	0.00468	0.053
Stage 3								
Harvesting	0.08279	0.1273	0.12141	0.17118	0.856	0.18948***	0.03586	0.420
Collection for threshing	0.04711	0.16307	0.06217	0.22237	0.956	0.07764***	0.02862	0.854
Threshing	-0.22295***	0.08624	-0.10088	0.11368	0.392	0.06068***	0.02332	0.002
Transport to storage	0.05201	0.09868	0.02796	0.12798	0.882	0.05683***	0.01666	0.962

Continued on next page...

Table 2.6 – continued from previous page

	3SLS (benchmark)		2SLS		2SLS = 3SLS	OLS		OLS = 3SLS
Explanatory variable	Coefficient	St. error	Coefficient	St. error	P-value	Coefficient	St. error	P-value
Non-Labor Inputs, ln(kg/acre)								
Stage 1								
Seeds	0.35205*	0.18192	0.17617	0.24847	0.568	0.0562	0.03704	0.111
Nursery rice (seedlings)[a]	0.30926*	0.17169	0.15529	0.23496	0.597	0.05023	0.03448	0.139
Chemical fertilizer	0.18137***	0.03671	0.08987*	0.04899	0.135	0.00797	0.00759	0.000
Manure	0.03689	0.12134	0.0524	0.16229	0.939	0.03293**	0.01649	0.974
Stage 2								
Chemical fertilizer	0.04157**	0.01759	0.0162	0.02341	0.386	-0.00051	0.00329	0.019
Manure	0.05624*	0.03278	0.03765	0.04399	0.735	0.00397	0.00612	0.117
Equipment Usage, ln(units/acre)								
Stage 1								
Walking tractor	0.00367	0.00342	0.00268	0.00473	0.865	0.00124	0.00222	0.551
Water buffalo	0.01764***	0.00682	0.01426	0.00934	0.770	0.00551	0.00504	0.153
Stage 3								
Harvesting machine	0.01729	0.01614	0.02089	0.02214	0.895	0.02694***	0.00863	0.598
Walking tractor (collection)	0.00268	0.00282	0.00283	0.00391	0.975	0.00107	0.00247	0.668
Shelling machine	0.05604**	0.02298	0.05390*	0.03047	0.955	0.05557***	0.01089	0.985
Threshing machine	-0.00747*	0.00439	-0.00142	0.00591	0.411	0.00662***	0.0021	0.004
Walking tractor (transport)	0.01156***	0.00243	0.01068***	0.00329	0.830	0.00693***	0.00197	0.139
Small four-wheel tractor	-0.75146***	0.19853	-0.58164**	0.26709	0.610	0	0	0.000
Truck	0.00962***	0.00372	0.00819	0.00508	0.820	0.00503*	0.00261	0.313

Continued on next page...

Table 2.6 – continued from previous page

	3SLS (benchmark)		2SLS		2SLS = 3SLS	OLS		OLS = 3SLS
Explanatory variable	Coefficient	St. error	Coefficient	St. error	P-value	Coefficient	St. error	P-value
Actual Total Rainfall, mm								
Stage 1	0.00023	0.00048	0.00017	0.00065	0.941	2.10E-04	0.00023	0.970
Stage 2	-0.00056*	0.00033	-4.50E-04	0.00045	0.844	-1.70E-04	0.00022	0.326
Stage 3	-0.00056	0.00034	-3.70E-04	0.00046	0.740	8.00E-05	0.00025	0.130
Month Indicator, Start of Stage 1								
June	-0.1945	0.26334	-0.05914	0.3679	0.765	1.72E-01	0.11962	0.205
July	-0.18887	0.37179	-0.10884	0.51949	0.900	0.15372	0.14916	0.393
August	-0.16143	0.44094	-0.13137	0.61294	0.968	1.34E-01	0.17904	0.535
September	0.11113	0.46236	0.05936	0.639	0.948	7.19E-02	0.25987	0.941
Year Indicator								
Year 1998-1999	-0.89730***	0.22177	-0.65456**	0.28452	0.501	-0.40509***	0.12577	0.054
Year 1999-2000	0.05235	0.1383	-0.00959	0.18514	0.789	-0.0142	0.07737	0.675
Year 2000-2001	1.13328**	0.46096	0.48209	0.61164	0.395	0.07899	0.10286	0.026
Year 2002-2003	-0.09896	0.08609	0.00254	0.10942	0.466	0.09864*	0.05325	0.051
Village Indicator								
Village 1	-1.35938**	0.69265	-1.08351	0.85599	0.802	-1.09790*	0.57176	0.771
Village 3	-1.02714	0.73076	-0.70072	0.91095	0.780	-0.48673	0.58616	0.564

[a]Seedlings are measured in ln(sets of nursery rice/acre).

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

2 are positive and significant, however, the magnitude of this positive effect is much larger in stage 1 (0.18 versus 0.04). This indicates that the effect of fertilizing on final yields depends not only on the total quantity of fertilizer used, but also on the timing of its application. In stage 1, fertilizer is applied during transplanting of seedlings. Like rainfall, at this stage fertilizing facilitates successful acclimatization of seedlings to soil and so affects crop's survival. During stage 2, plants are already in the middle of their growth, and fertilizer is used to augment soil quality. This in turn has effect on how strong the plants are, but is less likely to affect their survival.

Columns five and six show, respectively, coefficient estimates and their standard errors obtained from the 2SLS. We test for equality of 3SLS and 2SLS coefficients, and the corresponding p-values are reported in column seven. We do not reject equality of coefficient estimates by 3SLS and 2SLS. This shows that our 3SLS estimates are not biased by system misspecification. 3SLS is expected to be more efficient than 2SLS, and most coefficient estimates have larger standard errors under 2SLS.

The last three columns of table 2.6 show estimation results for the OLS. Columns eight and nine show, respectively, coefficient estimates and their standard errors obtained from the OLS. Column ten shows p-values for the test of equality of 3SLS and OLS coefficients. We reject the equality of 3SLS and OLS estimates for most of statistically significant 3SLS coefficient estimates.

Table 2.7 shows 3SLS coefficient estimates of stage 1 labor input equations. The second column lists explanatory variables by group (soil variables, labor inputs, and so on), and the first column indicates production stages to which each explanatory variable corresponds. Column three contains coefficient estimates of the equation for labor used in soil preparation and plowing, column four contains estimates of the equation for labor used in planting and transplanting, and column five shows estimates of the equation for labor used in fertilizing. Table 2.8 shows 3SLS coefficient estimates of stage 1 non-labor input demands

Table 2.7: Stage 1 Labor Input Demand Equations

Explanatory variable	Soil Preparation & Plowing	Planting & Transplanting	Fertilizing
Land and Soil			
Used Land, acre	0.19304***	-0.0179	0.0697
Soil pH	-0.1269	0.0439	0.0547
Organic matter, %	-0.83406*	0.0254	0.0724
Field capacity, %	0.0179	0.0008	-0.0395
Cation exchange capacity	0.3023	-0.0346	0.0094
Stage 1 Labor Inputs by Operation, ln(hours/acre)			
Soil preparation and plowing		0.0155	-0.09491***
(Trans)planting	0.0806		-0.15857*
Fertilizing	0.08054**	0.01979***	
Stage 1 Operation Type Indicator			
Vegetation clearing	0.5772		
First plowing	2.07524***		
More than one plowing	1.94317***		
Stage 1 Non-Labor Inputs, ln(kg/acre)			
Seeds		0.68456***	
Nursery rice (seedlings)[a]		0.72400***	
Chemical fertilizer			0.96183***
Manure			2.25274***
Stage 1 Equipment Usage, ln(units/acre)			
Walking tractor	0.03953***		
Water buffalo	0.05693***		
Expected Total Rainfall, mm			
Stage 1	-0.0032		
Stage 2	0.0010		
Stage 3	0.0016		
Stage 1 Month Indicator			
June	0.5672	-0.0888	0.1912
July	0.5963	-0.0944	0.0587
August	0.1871	-0.1501	0.66923**

Continued on next page...

Table 2.7 – continued from previous page

Explanatory variable	Soil Preparation & Plowing	Planting & Transplanting	Fertilizing
September	2.65918*	0.47001**	0.4207
Year Indicator			
Year 1998-1999	-1.8354	-0.0247	-2.24322**
Year 1999-2000	1.91723*	-0.0997	-0.6138
Year 2000-2001	3.25883**	-0.2272	-0.5375
Year 2002-2003	1.2127	-0.53102***	-0.7368
Village Indicator			
Village 1	-3.7003	-0.95831*	0.6036
Village 3	0.6967	0.1505	1.6760
Village 4	-4.6237	-0.3858	-1.7416
Log Real Wages by Operation			
Stage 1			
Soil preparation and plowing	-0.2181		
(Trans)planting		-0.0833	
Fertilizing			-0.0436
Stage 2			
Weeding and thinning	1.03212**	-0.20728*	0.4118
Fertilizing	-1.06583***	0.0523	-0.1986
Stage 3			
Harvesting	-1.11292*	0.2073	-0.0476
Collection for threshing	0.2369	-0.1005	0.2443
Threshing	0.7097	0.0829	0.0948
Transport to storage	-0.0867	-0.0719	-0.1346
Stage 2 Log Real Input Prices			
Chemical fertilizer	7.51016**	-0.7343	-2.7827
Manure	-0.37490*	-0.0365	-0.34825*

[a]Seedlings are measured in ln(sets of nursery rice/acre).

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

equations. The third column shows estimates of the equation for chemical fertilizer, column four shows estimates of the equation for manure, column five shows estimates for seeds equation, and column six shows estimates for the seedlings equation.

Table 2.8: Stage 1 Non-Labor Input Demand Equations

Explanatory variable	Chemical Fertilizer	Manure	Seeds	Seedlings
Land and Soil				
Used Land, acre	-0.0413	-0.0059	-0.02583**	-0.02290**
Soil pH	-0.0408	-0.0114	-0.05846*	-0.06111*
Organic matter, %	-0.1029	-0.0430	-0.0437	-0.0481
Field capacity, %	0.0470	0.0181	-0.0164	-0.0142
Cation exchange capacity	-0.0100	-0.0133	0.07413**	0.07323*
Stage 1 Labor Inputs by Operation, ln(hours/acre)				
Soil preparation and plowing				
(Trans)planting			0.48853***	0.61030***
Fertilizing	1.03980***	0.17216***		
Stage 1 Operation Type Indicator				
Vegetation clearing				
First plowing				
More than one plowing				
Stage 1 Non-Labor Inputs, ln(kg/acre)				
Seeds	0.1613	0.27910**		-1.03302***
Nursery rice (seedlings)[a]	0.1766	0.24205**	-0.95576***	
Chemical fertilizer		-0.15529***	-0.0024	-0.0045
Manure	-1.87542***		0.0717	-0.0114
Expected Total Rainfall, mm				
Stage 1	-0.0015	0.0018	-0.0001	0.0001
Stage 2	-0.0003		0.00059**	0.00058**
Stage 3	-0.0001	-0.0002	0.00051**	0.00048*
Stage 1 Month Indicator				
June	0.2397	-0.62943**	0.0505	-0.0122
July	0.1825	-0.4641	0.1044	0.0640
August	-0.5061	-0.59624*	0.2326	0.1908
September	-0.0638	-0.4988	-0.0265	-0.1389
Year Indicator				
Year 1998-1999	1.94683**	0.5399	-0.53756**	-0.45913*

Continued on next page...

Table 2.8 – continued from previous page

Explanatory variable	Chemical Fertilizer	Manure	Seeds	Seedlings
Year 1999-2000	0.2364	0.3765	-0.0948	-0.0192
Year 2000-2001	0.2970	0.2756	-0.1223	-0.0466
Year 2002-2003	0.5204	0.2202	0.26116*	0.34753**
Village Indicator				
Village 1	-1.0381	-0.4257	-16.99780***	0.2215
Village 3	0.5513	-0.2652	-17.60061***	0.3049
Village 4	0.5540	-0.0620	-17.34298***	1.18866**
Log Real Wages by Operation				
Stage 2				
Weeding and thinning	-0.3223	-0.1194	0.18592**	0.19602**
Fertilizing	0.0891	0.0902	-0.13001**	-0.12093*
Stage 3				
Harvesting	-0.0110	0.0998	-0.28619**	-0.28375**
Collection for threshing	-0.1485	-0.1057	0.0250	0.0242
Threshing	-0.1237	0.0199	0.0260	0.0162
Transport to storage	0.1562	0.0230	0.0096	0.0199
Log Real Input Prices				
Stage 1				
Chemical fertilizer	0.4310			
Manure		-0.0469		
Seeds			-0.0006	
Seedlings				-0.0075
Stage 2				
Chemical fertilizer	1.8370	0.9386	-0.1348	0.0850
Manure	0.2744	0.0817	-0.0138	-0.0004

[a]Seedlings are measured in ln(sets of nursery rice/acre).

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

Positive and significant coefficients on equipment in the equation for labor used in soil preparation and plowing (column 3 in table 2.7) show that available equipment is not labor-saving but, on the contrary, labor-intensive. Note the significant and negative coefficients on wage for labor used in harvesting in stage 3 in equations for labor used for plowing (column 3 in table 2.7), seeds (column 5 in table 2.8), and seedlings (column six in table 2.8). Recall the first-order condition (2.1) for inputs which indicated that each input's marginal cost reflects this input's effect on future expenditures through its effect on plant growth and therefore on optimal input usage in future stages. The significant and negative coefficients on wage for harvesting labor in equations for stage 1 inputs supports this notion - the higher are anticipated costs of harvesting, the less is planted in the first place.

Note the negative and significant coefficient on cultivated area in equations for seeds and seedlings (columns five and six, respectively, in table 2.8). Recall that inputs are measured in units per acre. Negative coefficient on area indicates that owners of larger plots are more likely to plant more sparsely. In other words, planting density is not in constant proportion to area. As expected, higher anticipated rainfall in stages 2 and 3 has positive and significant effect on the amounts of seeds and seedlings planted. Similarly, the coefficient on the 1998 dummy is negative and significant, indicating that less planting was done in the year with low rainfall. Note also the positive and significant coefficient on the 1998 dummy in the equation for chemical fertilizer (column three in table 2.8), which suggests that plants require more fertilizing when rainfall is low. Intuitively, seeds and seedlings enter each other's equations with negative and significant coefficients, which indicates their substitutability. The same result is observed for chemical fertilizer and manure (columns three and four in table 2.8) - chemical fertilizer has a negative and significant coefficient in the equation for manure, and vice versa, suggesting that the two are substitutes.

Tables 2.9 and 2.10 show coefficient estimates of stage 2 and stage 3 input demand equations, respectively. These tables have the same structure as table 7. Significant coefficients on DSSAT variables in stage 3 equations in table 2.10 show that DSSAT is an effective measure of stage 2 intermediate output. However, DSSAT is not significant in stage 2 input equations in table 2.9. In both stage 2 and stage 3 equations, labor input in previous stages is a significant measure of intermediate output. Equations for stage 2 chemical fertilizer and

Table 2.9: Stage 2 Input Demand Equations

Explanatory variable	Labor Inputs: Weeding & Thinning	Labor Inputs: Fertilizing	Non-Labor Inputs: Chemical Fertilizer	Non-Labor Inputs: Manure
Land and Soil				
Used Land, acre	0.1495	-0.18853**	0.36157***	0.14223**
DSSAT Intermediate Output, End of Stage 1, kg/ha				
Leaf weight	-0.0019	0.0032	0.0011	0.0025
Root weight	0.0005	-0.0170	-0.0015	-0.0090
Labor Inputs by Operation, ln(hours/acre)				
Stage 1				
Soil preparation and plowing	-0.4486	0.24910*	-0.53422**	-0.1739
(Trans)planting	2.14967**	-1.18751***	1.99461***	0.80078***
Fertilizing	0.35486**	-0.13837***	0.0915	0.0197
Stage 2				
Weeding or thinning		-0.05862**		
Fertilizing	-0.1526		1.06834***	0.39191***
Stage 2 Non-Labor Inputs, ln(kg/acre)				
Chemical fertilizer		0.62809***	0.0000	-0.41312***
Manure		0.73406***	-1.32418***	
Expected Total Rainfall, mm				
Stage 2	0.0122		0.00784**	0.00618***
Stage 3	-0.02890***		-0.0012	0.0001
Stage 2 Month Indicator				
July	1.3349	-2.30139***	1.9286	-0.0829
August	1.2032	-0.1198	-0.9523	-1.25404**
September	-4.68374*	-0.77516**	-0.4561	-0.6474
October	0.2369	0.0405	0.1414	-0.2989
Year Indicator				
Year 1998-1999	-15.61862***	0.2763	-4.25332*	-2.15624**
Year 1999-2000	-4.52860*	2.68319***	-3.2583	-1.72450***
Year 2000-2001	-11.77127***	16.14273***	-25.43526***	-10.69628***
Year 2002-2003	-6.32324***	0.4297	-0.2343	0.2555

Continued on next page...

Table 2.9 – continued from previous page

Explanatory variable	Labor Inputs: Weeding & Thinning	Labor Inputs: Fertilizing	Non-Labor Inputs: Chemical Fertilizer	Non-Labor Inputs: Manure
Village Indicator				
Village 1	-2.1920	16.84331***	0.0556	-30.37271***
Village 3	-2.6841	19.18121***	-3.3271	-26.40264***
Village 4	16.5323	24.13207***	30.17766***	-29.28461***
Log Real Wages by Operation				
Stage 2				
Weeding and thinning	-3.74809**			
Fertilizing		0.0044		
Stage 3				
Harvesting	-5.99696**	1.75826*	-3.74102**	-2.25162***
Collection for threshing	-1.4489	-1.0412	1.6739	0.6557
Threshing	-3.0107	1.1686	-1.5905	-0.6788
Transport to storage	1.1147	-0.7226	1.6784	1.4340
Stage 2 Log Real Input Prices				
Chemical fertilizer			0.8591	
Manure				0.69895***

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

manure usage (columns five and six, respectively, of table 2.9) support earlier results for stage 1. The two inputs enter each other's equations with negative and significant coefficients, indicating their substitutability. Negative and significant coefficient on wage for harvesting labor illustrates the sequential aspect of input choices. Interestingly, in both equations the coefficient on cultivated area is positive and significant, suggesting higher per acre usage of fertilizers on larger plots. This agrees with the earlier result for cultivated area in equations for seeds and seedlings that plants on larger plots are more likely to experience optimal growth conditions. The coefficients on equipment in stage 3 labor input equations in table 2.10 indicate that the more advanced equipment such as harvesting and threshing machines is labor-saving, while more traditional equipment is labor-intensive.

Table 2.10: Stage 3 Input Demand Equations

	Labor Inputs			
Explanatory variable	Harvesting	Collection for Threshing	Threshing	Transport to Storage
Land and Soil				
Used Land, acre	-0.0204	-0.03480**	-0.0231	0.0282
DSSAT Intermediate Output, End of Stage 2, kg/ha				
Leaf weight	-0.00070**	0.00107***	-0.0001	-0.00252***
Root weight	0.00454***	-0.00710***	0.0029	0.01446***
Stem weight	-0.00021***	0.00039***	-0.00033**	-0.00068***
Labor Inputs by Operation, ln(hours/acre)				
Stage 1				
Soil preparation and plowing	0.0171	0.0251	0.0106	-0.12190**
(Trans)planting	0.0274	-0.0181	0.0376	0.0190
Fertilizing	0.0091	-0.0085	-0.0138	0.02806*
Stage 2				
Weeding or thinning	-0.0011	0.02207***	0.0061	-0.04603***
Fertilizing	0.0075	-0.0105	-0.03427*	0.10666***
Stage 3				
Harvesting		0.55953***	0.0234	-0.77113***
Collection for threshing	0.66870***		-0.2734	1.52360***
Threshing	-0.0405	-0.0069		0.23825**
Transport to storage	-0.19171***	0.31247***	0.79606***	
Stage 3 Equipment Usage, ln(units/acre)				
Harvesting machine	-0.07241***			
Walking tractor (collection)		-0.0017		
Shelling machine			0.06098**	
Threshing machine			-0.03264***	
Walking tractor (transport)				0.01121***
Small four-wheel tractor				-0.0191
Truck				0.01247**
Expected Total Rainfall, mm				
Stage 3	0.0004	0.0004	-0.00144*	0.0002

Continued on next page...

Table 2.10 – continued from previous page

Explanatory variable	Labor Inputs Harvesting	Collection for Threshing	Threshing	Transport to Storage
Stage 3 Month Indicator				
September	-0.19077*	0.1132	0.2564	-0.48272*
October	-0.0803	0.1250	0.0857	-0.32403*
November	-0.0273	-0.0488	-0.18259*	0.0469
December	-0.08303*	0.17881***	0.14651*	-0.29590***
January	-0.40769***	0.48482**	-0.1479	-0.75069*
Year Indicator				
Year 1998-1999	0.1975	-0.0622	-0.2803	0.0596
Year 1999-2000	0.17869*	-0.0797	-0.32154**	0.2589
Year 2000-2001	0.25846***	-0.0419	-0.1568	-0.0386
Village Indicator				
Village 1	0.89550*	-1.65214**	-0.6756	0.6865
Village 3	0.5678	0.2189	-0.2803	0.7042
Village 4	-1.54231***	1.0028	-0.1863	-2.1168
Stage 3 Log Real Wages by Operation				
Harvesting	0.0277			
Collection for threshing		-0.0105		
Threshing			-0.48374***	
Transport to storage				-0.1199

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

To summarize, multistage approach to modeling rice cultivation gives us insights into the cultivation process which are unavailable from the single-stage approach. Our results demonstrate the importance of rainfall expectations in input decisions, as well as significant yearly variation of yields depending on the overall abundance or scarcity of rainfall in a given year. The results also support the multistage approach by demonstrating the importance of the timing of input application in addition to amounts of inputs used, as well as the sequential nature of input choices where input decisions at earlier stages are influenced in turn by their expected effect on input usage in subsequent stages. Another important result is insignificance of soil quality conditional of fertilizer usage, which suggests that farmers are not constrained by poor soil quality.

We next examine significance of DSSAT as a measure of intermediate output. Intermediate outputs from stages 1 and 2 are explanatory variables in input equations for stages 2 and 3, respectively. Table 2.10 demonstrates that DSSAT seems an effective measure of intermediate output from stage 2. However, DSSAT measures of stage 1 intermediate output are not significant in equations for stage 2 inputs, as shown in table 2.9. In our sample of four villages, farmers in two villages do not use manure, while farmers in the other two villages do use it. As a measure of intermediate output, DSSAT measures of stage 1 intermediate output are statistically significant within each of these two groups of villages, but their importance averages out when data are pooled over all four villages. This result is presented in table 2.11, which shows DSSAT coefficients in stage 2 input equations for each of the two village groups. Table 2.11 illustrates that DSSAT is statistically significant as a measure of intermediate output. However, its effect varies depending on practices of rice cultivation.

Table 2.11: Significance of DSSAT Measures of Stage 1 Intermediate Output in Stage 2 Input Equations

	Stage 2 input equations			
DSSAT End of stage 1 value	Weeding / Thinning	Fertilizing	Chemical Fertilizer	Manure
Villages 1 and 4				
Leaf weight (kg/ha)	-0.1509***	-0.0711***	0.1047***	
Root weight (kg/ha)	0.3120***	0.1104*	-0.1656**	
Villages 2 and 3				
Leaf weight (kg/ha)	0.0552***	0.0235***	-0.0315***	0.0047
Root weight (kg/ha)	-0.2053***	-0.1212***	0.1519***	-0.0249

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

2.4.1 Effects of Socio-Economic Variables

We next examine whether our results are sensitive to the household's economic position. We consider two indicators of economic affluence. One is level of household's per capita consumption relative to official poverty line for the province. The other is whether household's per capita income is above or below the province median in a given year.

Other natural options are per capita income relative to poverty line, and per capita consumption relative to the median. Table 2.12 compares these four poverty indicators, as well as two additional indicators based on village-level median. For each pair of indicators compared, the table gives percent of all households which are classified as, say, poor by one indicator and non-poor by the other, for all possible classifications combinations. As expected, the poverty line is a more severe indicator than the median. Village and province median approaches generate very similar results. However, both poverty line and median approaches yield different results depending on whether they are based on per capita income or consumption, with variation particularly pronounced for the median approach. For this reason, we use two indicators based on different per capita measures. We use the consumption-based poverty line indicator to isolate most pronounced differences in economic affluence. Income-based median indicator identifies more general level of relative economic well-being and has the advantage of putting equal number of households in each category.

One important consideration for rice cultivation is that poorer households may have plots with inferior soils, and, in addition, are not able to afford adequate level of fertilization. This would have adverse effect on yields. To further examine this possibility, we look at whether plot area, soil quality indicators, per acre usage of chemical fertilizer, and yields, vary by poverty. We use soil pH level and cation exchange capacity measures of soil quality. For each of these variables, we test the equality of means between poor and non-poor subsamples. We also examine whether model prediction errors and DSSAT prediction errors vary significantly

Table 2.12: Comparison of Four Poverty Indicators
(numbers in the table are percent of total sample)

Panel A: By Per Capita Income

		Province median		Village median			
		Above	Below	Above	Below		
Province poverty line	Non-poor	11.26	0.00	42.98	6.42	Above	Province median
	Poor	38.14	50.61	5.33	45.28	Below	

Panel B: By Per Capita Consumption

		Province median		Village median			
		Above	Below	Above	Below		
Province poverty line	Non-poor	16.10	0.00	43.95	5.45	Above	Province median
	Poor	33.29	50.61	3.51	47.09	Below	

Panel C: By Different Per Capita Measure

		Province Poverty Line Consumption		Province Median Consumption			
		Non-poor	Poor	Above	Below		
Income	Non-poor	4.48	6.78	23.85	25.54	Above	Income
	Poor	11.62	77.12	25.54	25.06	Below	

by poverty[15]. Another prediction error we consider is based on farmers' expectations of their yields at earlier stages of growth cycle. These results are presented in table 2.13.

Panel A of table 2.13 uses the consumption-based poverty line indicator. Poor households have statistically significantly smaller plots with inferior soils. They also use significantly more chemical fertilizer. Interestingly, actual yields are significantly higher for poor households. These results support the earlier observation that intensive use of fertilizer may be indicative of and successfully mitigates low soil quality. Yields variation between poor and non-poor groups is picked up by the model, which overpredicts more for non-poor than poor households, and this difference is statistically significant. Neither DSSAT nor farmers' prediction errors differ significantly between the two groups.

Consumption-based poverty line indicator identifies most pronounced cases of rich households. This makes the comparison group very large, and close to sample average. The richer households have larger plots with better soil. However, the rest of the population is able to compensate for lower-quality soil through fertilizer usage, so much so that they end up having higher yields. Indeed, the model's noticeable overprediction for rich households relative to the rest of the sample indicates failure of richer households on their potential.

Panel B of table 2.13 uses the income-based median indicator. Less affluent households have soil with lower cation exchange capacity. However, there is no significant difference in plot area, soil pH, or usage of chemical fertilizer. Overall lack of significant variations in panel B is expected given the balanced 50/50 nature of median-based poverty indicator. Model yield predictions are not significantly different for poor households. Actual yields are again significantly higher for less affluent households, though the magnitude of the difference is less pronounced than in the above result for poverty line indicator. Together, these results on yields for median and poverty line indicators suggest that financial constraints do not have an effect on yields. Again, neither DSSAT nor farmers' prediction errors differ significantly

[15] Prediction error is computed as actual yield minus prediction.

Table 2.13: Land Characteristics, Actual Yield, and Prediction Errors by Poverty

Panel A: Per capita consumption above/below poverty line

	Mean value		Equality of means test		
	Non-poor	Poor	t statistic	Lower P-value	Upper P-value
Plot area, acre	**3.01**	**2.70**	**1.366**	0.914	**0.086**
Soil pH	**5.94**	**5.73**	**2.562**	0.995	**0.005**
Soil cation exchange capacity	**2.68**	**2.27**	**3.081**	0.999	**0.001**
Chemical fertilizer usage, kg/acre	**33.62**	**58.15**	**-4.879**	**0.000**	1.000
Yield, kg/acre	**705.47**	**745.12**	**-1.254**	**0.105**	0.895
Model prediction error, kg/acre	**-58.23**	**-28.99**	**-0.684**	**0.247**	0.753
Model SE[a]prediction error, kg/acre	29.18	-1.41	1.079	0.860	0.140
DSSAT prediction error, kg/acre	11.62	-287.92	3.539	1.000	0.000
Farmer prediction error, kg/acre	-52.47	-40.04	-0.301	0.382	0.618

Panel B: Per capita income above/below median

	Mean value		Equality of means test		
	Above median	Below median	t statistic	Lower P-value	Upper P-value
Plot area, acre	2.70	2.81	-0.657	0.256	0.744
Soil pH	5.74	5.79	-0.732	0.232	0.768
Soil cation exchange capacity	**2.43**	**2.24**	**1.946**	0.974	**0.026**
Chemical fertilizer usage, kg/acre	55.25	53.16	0.557	0.711	0.289
Yield, kg/acre	**733.49**	**743.86**	**-0.446**	0.328	**0.672**
Model prediction error, kg/acre	-7.76	-59.02	1.633	0.949	0.051
Model SE[a]prediction error, kg/acre	16.49	-9.14	1.231	0.891	0.109
DSSAT prediction error, kg/acre	-239.26	-240.10	0.013	0.505	0.495
Farmer prediction error, kg/acre	-43.81	-40.81	-0.094	0.462	0.538

Bold indicates statistical significance at 10% level or better.

[a] "Model SE" denotes estimation results for the model with socio-economic variables included into input equations, as described below.

between the two groups.

To check for a hidden household effect, we look at persistence in performance from one year to another. A household's farming performance is measured by yield per acre. For each household, we compute its average per acre yield over all plots that household has cultivated in a given year. We then look at the distribution of these yields over all households in the province for a given year, separating yields into quintiles, and construct forward transition probabilities across the five years. These results are presented in table 2.14, with the largest

Table 2.14: Persistence in Performance, Forward Transition Probabilities
(each row sums to one)

		End Point, One Year Later				
Starting Point		< 20	20 - 40	40 - 60	60 - 80	> 80
	< 20	0.00	0.25	0.25	**0.50**	0.00
	20 - 40	**0.33**	0.00	0.00	**0.33**	**0.33**
1998	40 - 60	0.33	0.00	0.00	0.00	**0.67**
	60 - 80	0.00	**0.38**	0.25	0.13	0.25
	> 80	0.00	**0.33**	**0.33**	0.00	**0.33**
	< 20	**0.54**	0.31	0.08	0.08	0.00
	20 - 40	0.08	**0.44**	0.28	0.16	0.04
1999	40 - 60	0.16	0.16	0.26	**0.32**	0.11
	60 - 80	0.15	0.25	0.20	**0.30**	0.10
	> 80	0.05	0.16	0.05	**0.37**	**0.37**
	< 20	0.15	0.23	**0.38**	0.15	0.08
	20 - 40	**0.45**	0.25	0.10	0.05	0.15
2000	40 - 60	0.29	**0.36**	0.21	0.07	0.07
	60 - 80	0.05	0.25	0.25	0.10	**0.35**
	> 80	0.00	0.14	0.14	0.29	**0.43**
	< 20	**0.60**	0.00	0.07	0.27	0.07
	20 - 40	0.13	**0.27**	**0.27**	0.20	0.13
2001	40 - 60	0.19	0.14	**0.29**	0.24	0.14
	60 - 80	0.00	0.09	0.18	**0.36**	**0.36**
	> 80	0.00	0.17	0.28	0.11	**0.44**

Bold indicates the largest transition probability for a given quintile.

transition probabilities for each group emphasized in bold.

Overall, there is no pronounced persistence in performance. With the exception of 2001 - 2002 comparison, no other pair of years displays the diagonal pattern for largest transition probabilities which would be indicative of persistence. We see that "best" - above 80th percentile - farmers tend to remain best. However, "worst" performance - below 20th percentile - does not persist. It seems that being the "best" is a skill, while being the "worst" may be partly due to adverse shocks.

We now examine the model's robustness to the assumption of complete markets. To do this, we include socio-economic variables into each input demand equation and re-estimate our system of simultaneous equations. We use four types of socio-economic variables: borrowing, education, occupation, and demographic composition. Borrowing variables include indicators for household borrowing from a relative, and a non-relative or an institution[16]. These borrowing options are not exclusive. Education is measured as average years of schooling of household members aged 15 and older. Occupation is measured as percent of household members aged 11 to 65 with primary employment in non-agricultural sector. Demographic variables include number of people in the household, percent of males aged 11 to 65 in the household, and age composition of household members.

Variables in the demographics and occupation categories measure potential availability of household members for work on the plots. We include these socio-economic variables into equations for labor inputs. Variables in the borrowing and education categories provide different measures of potential financial constraints. We include these socio-economic variables into equations for non-labor inputs. We use stage-specific measures of socio-economic variables and, as with production variables, take into account plot-specific timing of stages. In other words, the equation for each input used in stage i includes measures of socio-economic variables characterizing the household during stage i.

Table 2.15 shows coefficient estimates for socio-economic variables in non-labor input

[16] Indicator for borrowing from a non-relative or an institution combines borrowing from a money-lender, BAAC, PCG, agricultural cooperative, village fund, and / or other institution.

equations. The first column lists explanatory socio-economic variables. Columns two though five show coefficient estimates for each of stage 1 non-labor input equations. Columns six and seven show coefficient estimates for each of stage 2 non-labor input equations. Neither education nor borrowing variables enter significantly into any of the non-labor equations. This suggests that the model is robust with respect to the assumption of no credit constraints.

Table 2.16 shows coefficient estimates for socio-economic variables in labor input equations. The first column lists explanatory socio-economic variables. Columns two, three and four show coefficient estimates for each of stage 1 labor input equations. Columns five and six show coefficient estimates for each of stage 2 labor input equations, and columns seven through ten show coefficient estimates for stage 3 labor input equations. We see that none of the socio-economic variables are significant in equations for labor used in plowing and fertilizing in stage 1. For the rest of labor inputs, at least one of the socio-economic variables is significant. We would expect that larger households, households with mostly adult males, and households with mostly young and/or middle-aged adults potentially have more household labor available to work on the plot. At the same time, households with members occupied mostly in non-agriculture, and households with mostly infant and/or elderly members potentially have less household labor available for work on the plot. The signs of significant coefficient estimates are broadly consistent with these expectations. Of interest is the alternation of significant variables. For labor used in stage 1 planting and stage 2 weeding operations, what seems to matter is the number of available household members but not their age composition. On the opposite, for labor used in stage 3 operations, the number of available household members does not seem to matter but their age composition is important. This suggests that stage 3 cultivation operations make use of age-specific labor skills but are not affected by household size.

In summary, the model appears robust to the assumption of no credit constraints. Separation of household's labor supply and demand is supported in earlier stages of production.

Table 2.15: Socio-Economic Variables in Non-Labor Input Equations

Explanatory variable	Stage 1 Chemical Fertilizer	Stage 1 Manure	Stage 1 Seeds	Stage 1 Seedlings	Stage 2 Chemical Fertilizer	Stage 2 Manure
Borrowing Source						
Relative	-0.0600	-0.0370	-0.1257	-0.1248	-0.3475	0.2342
Non-relative or an institution	0.0217	0.0430	-0.0654	-0.0608	0.0762	0.0154
Average Household Education (years)						
Education of hhd members aged 15 and older	0.0111	-0.0013	-0.0266	-0.0288	0.2144	0.1683

***, **, and * denote, respectively, significance at 1%, 5%, and 10% level.
Bold coefficients indicate significance at 10% level or better.
Bold variables indicate significance at 10% level or better in at least one input demand equation.

Table 2.16: Socio-Economic Variables in Labor Input Equations

	Stage 1			Stage 2		Stage 3			
Explanatory var.	Plowing	Planting	Fertilizing	Weeding	Fertilizing	Harvesting	Collection	Threshing	Transport
Primary Occupation of Adult (11 to 65) Household Members									
% non-agr[a]	-0.0074	-0.0003	0.0002	0.0287	-0.03424**	-0.00548**	0.00680**	0.00642*	-0.01655***
Household Demographics									
Hhd size[b]	0.1600	0.03999**	-0.0049	2.03484***	0.0021	-0.0049	-0.0249	-0.0024	0.0133
% adults[c]	-0.0005	0.00311**	-0.0002	-0.08778**	-0.0078	-0.0007	-0.0021	-0.0011	0.0047
% of hhd members aged:									
< 4	-0.0280	0.0040	0.0023	-0.0409	0.0006	0.0012	0.00959*	0.0030	-0.01774*
4 to 14	0.0017	-0.0002	0.0006	0.0498	-0.0100	0.0008	0.0071	0.0021	-0.0087
15 to 30	0.0141	0.0001	0.0001	-0.1061	0.0015	-0.0021	0.01519***	0.01300**	-0.01894**
31 to 50	0.0144	-0.0004	0.0000	0.0081	0.0124	-0.0037	0.01043**	0.0084	-0.0122
aged 51 to 65	0.0132	-0.0006	0.0000	-0.0582	0.0249	-0.00540*	0.01400***	0.0070	-0.01814**

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

[a] % of household members with non-agricultural occupation.
[b] Number of people in the household.
[c] % of adult (age 11 to 65) males in the household.

In the final production stage, household's labor usage is sensitive to the age composition of its members.

The next section introduces our approach to modeling climate change, and after that we describe how we integrate economic model with DSSAT and climate change models.

2.5 Climate Change Scenarios and IPCC SRES

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"[17] 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[18]. 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)[19], both

[17] "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).

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

[19] 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

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.

2.5.1 Predicted Climate Changes and Agricultural Impacts for Southeast Asia

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

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

differences in purchasing power. However, for this micro-study, we accept these scenarios as given.

2.6 Climate Change Impact Modeling: Integration of Crop, Weather, Climate and Economic Models

The integrated approach began by running DSSAT to simulate rice growth for the 826 individual crop-plots in northern Sisaket province. The DSSAT predictions were positively and significantly correlated with yield variation across the plots for those years, with a statistically significant correlation coefficient of 0.09. These initial simulations allowed calibration of DSSAT and confirmed its usefulness in capturing crop growth.

Next, the economic model was estimated, using the original plot data. Actual rain data were used to construct farmer's rain expectations. DSSAT predictions from the first step were used to construct measures of intermediate output from stages one and two.

Next, 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.

These neutral and alternative climate scenario realizations were then used as inputs to DSSAT, and rice yields were simulated for a stratified sample of 100 plots. These 100 plots were drawn at random from our larger sample of 826 plots, with equal share drawn from each of five years of actual data. This produced 300 yield realizations for each of 100 plots, with 100 realizations for each of neutral, high emissions and low emissions climate scenarios. Under each climate scenario, variation across 100 yield realizations for each plot are due to variation in the stochastic weather realizations. Throughout these simulations, all non-weather inputs were kept the same for each plot, at values of the actual data from 1998-2002. In other words, no adjustment was made to inputs and timing from one weather realization to another. Thus, for a given plot, variation in assumed climate and weather realizations was the only source of difference in yields in the 300 DSSAT yield simulations for that plot.

The final step was to use the three generated weather scenarios together with corresponding DSSAT crop simulations as inputs into the estimated economic model to predict yields. For each plot, individual rain expectations were constructed for each of 100 weather realizations under each climate scenario. Similarly, measures of intermediate stage one and stage two outputs were constructed from weather realization-specific DSSAT simulations for a given plot. Prediction then proceeded in four steps. In the first step, input levels for stage one were predicted. These predictions incorporate rain expectations for stage one and thus reflect variation in input usage due to differences in weather realizations. In the second step, predictions of levels of first stage inputs were used together with DSSAT measures of stage one intermediate output and stage two rain expectations to predict levels of stage two inputs. These predictions reflect variation in input usage due to both differences in weather realizations and adjustments made by the farmer in the first stage. In the same manner we predicted levels of stage three inputs. We then used predictions of all inputs together with rain realizations as inputs into composite production function and predicted final yields.

These final yields predictions integrate models of climate change, weather variations within each climate scenario, plant's biological development as modeled by DSSAT, and estimation of farmer's production choices as modeled by economic model.

2.7 Results

We first provide a summary of the two alternative climate changes that we consider - the high and low emission scenarios. Table 2.17 uses the 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 2.17 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 of the 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. However 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.

DSSAT predictions are summarized in panel A of table 2.18 and in table 2.19. We

Table 2.17: 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

Table 2.18: Aggregate Yield Changes Across Climate Scenarios

Panel A: DSSAT predictions

	Neutral to high emissions	Neutral to low emissions	Low emissions to high emissions
Yield change	-53.521	-209.154	155.633
Percent change	-3.53	-13.79	11.91
P-value[a]	2.683E-02	1.030E-12	2.390E-09

Panel B: Economic model predictions

	Neutral to high emissions	Neutral to low emissions	Low emissions to high emissions
Yield change	-71.152	-79.230	7.981
Percent change	-10.81	-12.04	1.40
P-value[a]	0.000E+00	0.000E+00	1.170E-13

[a] Corresponds to one-sided test in the direction indicated by sign of yield change in the first row: H_a is decrease in yields for columns one and two and H_a is increase in yields for column three.

first look at table 2.18, which provides aggregate yield comparisons across the three climate scenarios. Row one shows mean yield change, measured in kilograms per acre, and row two expresses this change as percent of aggregate mean yield under initial climate scenario. Row three shows p-value for the test of equality of means under initial and final climate scenarios. Compared to neutral climate, aggregate yields decrease under both high and low emissions scenarios, and these yield decreases are highly statistically significant. Yields are also lower under low emissions than high emissions scenario, despite the fact that low emissions climate is less extreme of the two. This may be due to the damaging effect on the crop of higher rainfall during the final production stage, when grain is mature and harvesting takes place.[20]

Table 2.19 provides plot-level analysis of DSSAT predictions. The first three rows of table 2.19 compare predicted yields, measured in kilograms per acre, when shifting from neutral to high emissions climate. For each plot in our sample of 100 plots, we test the equality of mean yields under neutral and high emissions climates. We then compute the percent of

[20] Alternatively, this may be due to interactive effects under CO2 regimes (similar results were obtained by Parry, Rosenzweig, Iglesias, Livermore, and Fischer (2004)).

plots that have statistically significant change in yields. These numbers are reported in the first row of table 2.19, separately for increases and decreases in yields, for 1, 5 and 10 percent significance levels. The second row of table 2.19 reports the actual size of mean yields change over plots conditional on the change being statistically significant. To give the idea of the scope of yield changes, third row expresses mean yields change of row two in percent. In the same manner, rows four to six compare predicted yields when shifting from neutral to low emissions climate, and rows seven to nine compare yields when shifting from low to high emissions climate.

DSSAT predicts lower yields for a quarter to a third of the plots under both the low and high emission scenarios. For these plots, decrease in yields is severe, ranging from 30 to over 40 percent. Decrease in yields is stronger when shifting to low emissions scenario, under which both more plots are affected and the scale of yield decrease is higher. Note also that, comparing plots with decreased yields under low and high emissions climates, plots affected under low emissions scenario were more productive under neutral climate than plots affected under high emissions scenario.

DSSAT predictions thus suggest that yields decrease more under the milder low emissions scenario. Despite the fact that high emissions climate has less rain during the second half of the year while low emissions climate has moderately more rain throughout the year, farmers fare worse in low emissions climate.

Model predictions are summarized in panel B of table 2.18 and in table 2.20. Panel B of table 2.18 provides aggregate results for model predictions. As is the case with DSSAT predictions, the model predicts lower aggregate yields under both high and low emissions scenarios when compared to neutral scenario. These yield decreases are again statistically significant. Similarly, yields are lower under low emissions then under high emissions scenario, although model predicts much smaller gap between the two.

Table 2.20 provides plot-level analysis of model predictions and is constructed in the same

Table 2.19: DSSAT Predictions of Yield Changes

Climate shift	Variable	1% significance Increase	1% significance Decrease	5% significance Increase	5% significance Decrease	10% significance Increase	10% significance Decrease
Neutral	Percent of sample	4.21	20.00	10.53	26.32	13.68	36.84
to	Yield change	306.308	-272.687	325.976	-220.047	286.748	-260.302
High emissions	Percent change	102.50	-49.25	49.26	-33.98	38.51	-27.97
Neutral	Percent of sample	3.16	29.47	5.26	35.79	12.63	36.84
to	Yield change	119.723	-581.584	221.085	-605.151	206.053	-591.514
Low emissions	Percent change	123.43	-50.27	78.77	-43.57	36.42	-42.76
Low	Percent of sample	5.26	1.05	9.47	5.26	20.00	10.53
to	Yield change	2301.83	-61.71	1322.69	-75.67	690.29	-117.94
High emissions	Percent change	36.83	-100.00	21.45	-8.60	21.30	-7.68

Table 2.20: Economic Model Predictions of Yield Changes

		1% significance		5% significance		10% significance	
Climate shift	Variable	Increase	Decrease	Increase	Decrease	Increase	Decrease
Neutral	Percent of sample	15.85	62.20	15.85	68.29	17.07	69.51
to	Yield change	2.427	-114.872	2.427	-104.719	2.396	-102.973
High emissions	Percent change	0.42	-14.08	0.42	-12.84	0.43	-12.62
Neutral	Percent of sample	79.27	12.20	81.71	12.20	81.71	12.20
to	Yield change	3.840	-675.937	3.898	-675.937	3.898	-675.937
Low emissions	Percent change	0.55	-98.20	0.55	-98.20	0.55	-98.20
Low	Percent of sample	4.82	84.34	4.82	85.54	4.82	85.54
to	Yield change	313.04	-8.32	313.04	-8.29	313.04	-8.29
High emissions	Percent change	0.83	-0.99	0.83	-0.98	0.83	-0.98

manner as table 2.19. Rows one to three compare predicted yields, measured in kilograms per acre, when shifting from neutral to high emissions climate, rows four to six compare predicted yields when shifting from neutral to low emissions climate, and rows seven to nine compare yields when shifting from low to high emissions climate.

Model predictions stand in stark contrast with DSSAT predictions. The first thing to note is that the fraction of the sample experiencing statistically significant yield decrease under high emissions climate more than doubles compared to DSSAT. Yields go down for 68 percent of the plots, with the average decrease of about 13 percent. However, under the low emissions climate yields actually increase for over 80 percent of the plots, albeit only by half a percent. For a small number of plots the crop has failed altogether under low emissions climate. Further, we also see that there is no difference in productivity for plots affected under low emissions scenario versus those affected under high emissions scenario.

Thus, according to model predictions, farmers manage to take advantage of the moderate increase in rainfall under low emissions climate. The majority of farmers do not experience large scale changes in yields. At the same time, however, there is a chance of complete crop failure. We next look if this risk is associated with soil quality or farmer's finances.

We look at the connection between yield changes and per capita income in farmer's household. We compute the probability of a household's per capita income being below the median[21] given that the household experienced statistically significant increase (decrease) in yields. We also consider differences in soil quality between plots with and without statistically significant yield changes. We use two measures of soil quality. One is pH, which indicates the relative acidity or alkalinity of soil. Another is cation exchange capacity (CEC), which indicates soil's capacity to hold cation nutrients. CEC is determined by amounts of clay and humus in the soil and is not easily adjusted. For both measures, we compute the difference in

[21] Household's per capita income is compared to the province median per capita income of all households in our larger sample of 147 households in each of five years in the sample. Our results hold when we do comparisons using village-specific median per capita income.

soil quality between plots with and without yield increase (decrease), expressed in percent. We also test for equality of mean pH and CEC values between plots with and without yield increase (decrease) and report the resulting probabilities.

These results are presented in table 2.21. Soil quality is not associated with yield changes no matter which climate change is considered. This is true for both DSSAT and model predictions. This result is intuitive. Soil quality is already taken into account in yield predictions for each climate, and should have no further unexplained effect on farmer's productivity. Household income also does not correlate with yield changes, with one notable exception. We see that the few plots that experience crop failure under low emissions climate according to model predictions have a 70 percent chance of having per capita income below median.

We now compare model and DSSAT predictions for the two climate change scenarios. To repeat, there are several differences between the model and DSSAT that are key to this analysis. DSSAT takes into account only amounts of non-labor non-equipment inputs such as fertilizer and seedlings that are applied to soil. When we simulated DSSAT under alternative climate scenarios, these input values were not adjusted from their actual levels reported under current climate. Thus changes in yields predicted by DSSAT are driven solely by changes in climate in which the crop is grown, with no adjustment to any inputs.

Changes in rainfall across different climate scenarios are taken into account in model prediction through adjustment of farmers' rainfall expectations. This is a big plus for using the economic model as it can accommodate these adjustments. The model also incorporates results of DSSAT simulations though changes in measures of crop's intermediate states. Thus all levels of production inputs used in model prediction incorporate adjustments to climate change through these two channels.

For the shift from neutral to low emissions climate, comparison of DSSAT and model predictions shows that taking into account farmer's response to climate change makes a

Table 2.21: Soil Quality and Household Income in Yield Changes

Variable	DSSAT predictions			Model predictions		
	Neutral to high	Neutral to low	Low to high	Neutral to high	Neutral to low	Low to high
Panel A: Yield changes significant at 1% level						
Increase in yield						
Below median per capita income	50.00	0.00	40.00	53.85	47.69	50.00
pH mean change, in percent	-3.03	3.20	0.92	-4.94	2.98	-10.33
pH mean change, P-value	0.665	0.691	0.884	0.223	0.333	0.136
CEC mean change, in percent	73.56	-4.38	16.95	31.07	-7.71	41.63
CEC mean change, P-value	0.186	0.903	0.620	0.199	0.577	0.363
Decrease in yield						
Below median per capita income	68.42	60.71	100.00[a]	49.02	70.00	47.14
pH mean change, in percent	-4.38	-0.74	n.a.	2.32	-6.25	4.40
pH mean change, P-value	0.207	0.810	n.a.	0.416	0.168	0.181
CEC mean change, in percent	16.14	7.22	n.a.	-9.12	23.20	-11.24
CEC mean change, P-value	0.385	0.634	n.a.	0.470	0.370	0.438
Panel B: Yield changes significant at 5% level						
Increase in yield						
Below median per capita income	50.00	40.00	33.33	53.85	46.27	50.00
pH mean change, in percent	-3.53	-2.05	4.30	-4.94	3.68	-10.33
pH mean change, P-value	0.439	0.745	0.372	0.223	0.243	0.136
CEC mean change, in percent	23.60	23.78	-8.33	31.07	-6.09	41.63
CEC mean change, P-value	0.363	0.510	0.689	0.199	0.668	0.363

Continued on next page. . .

Table 2.21 – continued from previous page

Variable	DSSAT predictions			Model predictions		
	Neutral to high	Neutral to low	Low to high	Neutral to high	Neutral to low	Low to high
Decrease in yield						
Below median per capita income	60.00	55.88	60.00	50.00	70.00	46.48
pH mean change, in percent	-3.93	-0.54	-2.42	3.46	-6.25	4.22
pH mean change, P-value	0.212	0.853	0.701	0.235	0.168	0.205
CEC mean change, in percent	8.62	8.74	23.91	-14.57	23.20	-9.07
CEC mean change, P-value	0.588	0.547	0.508	0.246	0.370	0.540
Panel C: Yield changes significant at 10% level						
Increase in yield						
Below median per capita income	46.15	50.00	52.63	50.00	46.27	50.00
pH mean change, in percent	-4.94	-2.32	8.20	-5.07	3.68	-10.33
pH mean change, P-value	0.223	0.583	0.020	0.197	0.243	0.136
CEC mean change, in percent	20.65	17.03	-11.07	33.12	-6.09	41.63
CEC mean change, P-value	0.360	0.454	0.460	0.163	0.668	0.363
Decrease in yield						
Below median per capita income	60.00	54.29	70.00	49.12	70.00	46.48
pH mean change, in percent	-2.03	-1.00	-3.34	3.65	-6.25	4.22
pH mean change, P-value	0.483	0.730	0.465	0.211	0.168	0.205
CEC mean change, in percent	-5.25	9.85	15.80	-14.08	23.20	-9.07
CEC mean change, P-value	0.691	0.497	0.519	0.266	0.370	0.540

[a]Note from row seven of table 2.19 that when moving from low to high emissions scenario DSSAT predicts decrease in yield which is statistically significant at 1 % level for only one crop-plot.

substantial difference. Without input adjustments, we see statistically significant yield decrease of large magnitude in a third of our sample plots. Once farmer's responses to climate change are incorporated, the majority of plots do not experience yield decrease and even enjoy a slight increase in yields. Farmers are thus able to adjust to climate change from neutral to low emissions scenario.

The role of farmers' adjustment to climate change is also evident in the shift from neutral to high emissions climate. Without input adjustments, we see statistically significant yield decrease of around 30 percent in a quarter of our sample plots. Once farmer's predicted responses to climate change are incorporated, the fraction of the sample experiencing yield decrease more than doubles, but the magnitude of average yield decreases by more than half. Farmers are thus predicted to respond to this more severe climate change with adjustments that prevent large crop failures, at the cost of reducing their yields by about 13 percent. In other words, farmers are unable to fully neutralize the effects of the more severe climate change. However, by adjusting their crop cultivation routine they are able to mitigate the adverse effects of this more extreme climate change scenario.

It should also be noted that under milder climate change from neutral to low emissions scenario, farmers do not find it necessary to adjust their cultivation methods sufficiently to reduce the chance of crop failure. Our results thus suggest that various climate changes pose different challenges to the farmers. One is overall reduction in yields, when crops do not fail but are less productive. Another is crop failure on a large scale. It appears that there is a trade-off in adjustment techniques for these two challenges. Under less severe climate change large crop failure may be a result of bad weather draw, so farmers choose adjustment that maintains their yields but does not guard against crop failure. Under more severe climate change any weather realization can lead to large crop failure, and so farmers switch the adjustment technique to preventing large crop failure at the cost of lower yields.

2.8 Conclusion

In this paper we specify and estimate a three-stage production function for rice cultivation. A traditional single-stage production function approach does not reflect the inherent sequential structure of crop cultivation, in which farmers continually update their input usage in response to realizations of production shocks, rainfall in particular.

We construct crop-plot specific timing of production stages, and use it to form crop-plot specific rainfall realizations and expectations. We also successfully incorporate soil science crop simulation model (DSSAT) into our economic model of rice production function. Our estimation results indicate significance of the multistage approach and demonstrate the importance of rainfall expectations in input demands. We find that yields are not affected by credit constraints. In particular, inferior soil quality is effectively augmented by fertilizer usage and does not have a negative effect on yields.

We next apply our model to measure the effect of climate change on rice yields. To simulate potential climate change scenarios for northeastern Thailand we integrate economic model and DSSAT crop simulations with global climate change models and weather simulators. We consider two climate change scenarios: low emissions, with moderate increases in temperature and rainfall, and high emissions, with higher increase in temperature and less rainfall during the months of rice cultivation.

Our results illustrate the complexity of climate change effects on rice yields at both the aggregate and individual levels, as well as the extent of farmers' ability to counter climate change. Milder climate change does not necessarily mean smaller adverse effect on yields. Overall, farmers are unable to neutralize the adverse effects of the more extreme climate change. However, they are able to cope with milder climate change and even benefit slightly from small increases in rainfall. We find that farmers' ability to adjust to climate change for the most part is not correlated with soil quality of their land or their incomes.

It should be noted that in our analysis we consider only farmers' adjustment through input decision rules, namely, their choices of levels of production inputs. We do not model or incorporate possible changes in timing of input usage. We also do not consider broader adjustments such as changes in the type of crop grown or migration. As a result, our findings may overstate both yield changes and implied welfare effects due to climate change.

Bibliography

(2011): FAOSTAT, Food and Agriculture Organization of the United Nations, http://faostat.fao.org.

(2011): Thailand Database Research Archive, NORC and University of Chicago, http://cier.uchicago.edu/.

(2011): World Development Indicators and Education Statistics, World Bank, http://data.worldbank.org/.

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

Angrist, J. D., and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

Becker, G. S. (1994): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. The University of Chicago Press, Chicago, IL.

Bjork, C. (2005): *Schooling Indonesian Education: Teachers, Schools, and Central Bureaucracy*, East Asia (New York, N.Y.). Routledge, New York, NY.

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.

Dahl, G. B. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367–2420.

Dubin, J. A., and D. L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52(2), 345–62.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *The American Economic Review*, 91(4), 795–813.

Evans, W. N., and R. M. Schwab (1995): "Finishing High School and Starting College: Do Catholic Schools Make a Difference?," *The Quarterly Journal of Economics*, 110(4), 941–74.

Felkner, J., K. Tazhibayeva, and R. M. Townsend (2009): "An Integrated Economic-Crop Model of Multi-Stage Rice Production," working paper, University of Chicago.

Fischer, G., M. Shah, and H. van Velthuizen (2002): *Climate Change and Agricultural Vulnerability*International Institute for Applied Systems Analysis (IIASA).

Frankenberg, E., and L. Karoly (1995): *The 1993 Indonesian Family Life Survey: Overview and Field Report*RAND.

Frankenberg, E., and D. Thomas (2000): *The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2*RAND.

Freeman, R. B. (1974): "Occupational Training in Proprietary Schools and Technical Institutes," *The Review of Economics and Statistics*, 56(3), 310–18.

Ginther, D. K. (2000): "Alternative Estimates of the Effect of Schooling on Earnings," *The Review of Economics and Statistics*, 82(1), 103–116.

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.

Hay, J. W. (1980): "Occupational Choice and Occupational Earnings: Selectivity Bias in a Simultaneous Logit-OLS Model," unpublished Ph.D. dissertation, Yale University.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

——— (1980): "Addendum to ŞSample Selection Bias as a Specification ErrorŤ," *Evaluation Studies Review Annual*, 5, 69–74.

——— (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *The Journal of Political Economy*, 109(4), 673–748.

Heckman, J. J., J. L. Tobias, and E. Vytlacil (2001): "Four Parameters of Interest in the Evaluation of Social Programs," *Southern Economic Journal*, 68(2), 210–223.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, 88(3), 389–432.

Ichimura, H., and L.-F. Lee (1991): "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in *Nonparametric and Semiparametric Methods in Economics and Statistics: Proceedings of the Fifth International Symposium in*

Economic Theory and Econometrics, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen, International symposia in economic theory and econometrics, pp. 389–432. Cambridge University Press.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Johnson, K., W. Gaylord, and G. Chamberland (1993): *Indonesia: a Study of the Educational System of the Republic of Indonesia and a Guide to the Academic Placement of Students in Educational Institutions in the United States*, World education series. AACRAO: NAFSA, Washington, DC.

Lee, L.-F. (1982): "Some Approaches to the Correction of Selectivity Bias," *The Review of Economic Studies*, 49(3), 355–372.

——— (1983): "Generalized Econometric Models with Selectivity," *Econometrica*, 51(2), 507–12.

——— (1995): "Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models," *Journal of Econometrics*, 65(2), 381–428.

Lee, L.-F., G. S. Maddala, and R. P. Trost (1980): "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica*, 48(2), 491–503.

Lewis, M. (ed.) (2009): *Ethnologue: Languages of the World*, World education series. SIL International, Dallas, TX, 16 edn.

Manski, C. F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27(3), 313–333.

Mare, R. D. (1980): "Social Background and School Continuation Decisions," *Journal of the American Statistical Association*, 75(370), 295–305.

Mincer, J. A. (1974): *Schooling, Experience, and Earnings*, no. minc74-1 in NBER Books. National Bureau of Economic Research, distributed by Columbia University Press, New York, NY.

Murphy, K. M., and F. Welch (1990): "Empirical Age-Earnings Profiles," *Journal of Labor Economics*, 8(2), 202–229.

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.

Newey, W. K., J. L. Powell, and J. R. Walker (1990): "Semiparametric Estimation of Selection Models: Some Empirical Results," *The American Economic Review*, 80(2), 324–28.

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.

Paulson, A. L., R. M. Townsend, and A. Karaivanov (2006): "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship," *Journal of Political Economy*, 114(1), 100–144.

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.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati, and F. Witoelar (2004): *The Third Wave of the Indonesia Family Life Survey (IFLS3): Overview and Field Report*RAND.

Strauss, J., F. Witoelar, B. Sikoki, and A. M. Wattie (2009): *The Fourth Wave of the Indonesia Family Life Survey (IFLS4): Overview and Field Report*RAND.

Trost, R. P., and L.-F. Lee (1984): "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity," *The Review of Economics and Statistics*, 66(1), 151–56.

Willis, R. J., and S. Rosen (1979): "Education and Self-Selection," *The Journal of Political Economy*, 87(5), S7–S36.

Appendix A

Map and Additional Statistics for Chapter 1

This appendix contains additional data information on Indonesia.

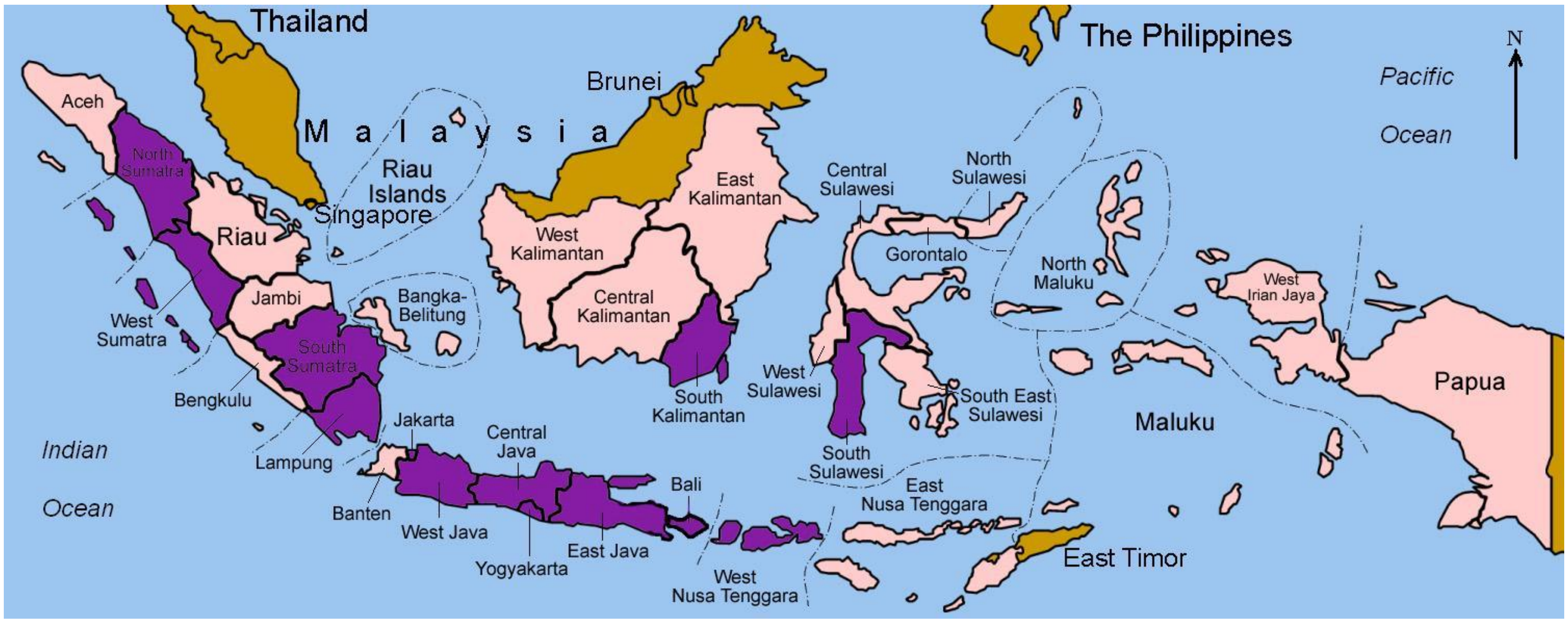

Figure A.1: Map of Indonesia and Sample Provinces (sample provinces are in darkest shade)

Table A.1: Public Tertiary Schools, Means by Province

				Fulltime teaching staff		
	Total	Year	1990	Ratio to	% with degree	
Province	number	founded	enrollment[a]	enrollment	S2[b]	S3[c]
Sumatra island	18					
DI Aceh	2	1962	10.24	0.04	12.52	3.80
North Sumatra	3	1960	9.72	0.08	8.96	2.37
West Sumatra	4	1960	6.43	0.10	9.45	3.39
Riau	2	1966	5.19	0.06	9.33	1.47
Jambi	2	1965	4.53	0.05	14.89	0.43
South Sumatra	2	1962	7.44	0.06	14.61	1.39
Bengkulu	1	1982	2.69	0.18	17.58	1.41
Lampung	2	1967	5.72	0.06	11.35	0.62
Java island	27					
DKI Jakarta	1	1963	9.16	0.09	8.03	6.90
West Java	9	1962	11.33	0.09	15.86	9.07
Central Java	6	1966	10.50	0.09	10.23	1.75
DI Yogyakarta	4	1964	13.49	0.06	15.11	5.49
East Java	7	1961	11.04	0.07	20.42	4.44
Bali island	2					
Bali	2	1966	7.28	0.15	10.47	2.41
Nusa Tenggara island	2					
West Nusa Tenggara	1	1962	6.76	0.08	14.58	0.56
East Nusa Tenggara	1	1962	6.41	0.10	11.23	0.30
Kalimantan island	5		0.00			
West Kalimantan	1	1963	8.41	0.07	19.36	1.28
Central Kalimantan	1	1963	3.93	0.10	5.41	0.25
South Kalimantan	2	1962	1.97	0.39	6.31	1.18
East Kalimantan	1	1962	5.01	0.08	14.55	4.23
Sulawesi island	7					
North Sulawesi	2	1958	9.53	0.13	10.28	2.07
Central Sulawesi	1	1981	6.56	0.09	6.16	0.34
South Sulawesi	3	1962	14.02	0.06	13.80	5.46
South East Sulawesi	1	1981	5.82	0.05	6.92	0.35
Maluku island	1					
Maluku	1	1956	7.52	0.07	13.08	0.80
Irian Jaya island	1					
Irian Jaya	1	1962	3.91	0.09	6.87	0.60

Data source: Johnson, Gaylord and Chamberland (1993).

[a] Enrollment number is in thousands of students.
[b] S2 degree is equivalent to master's degree.
[c] S3 degree is equivalent to doctorate degree.