Climate and land in turmoil: Welfare impacts of extreme weather events and palm oil production expansion in Indonesia

Outi Korkeala

Submitted for the degree of Doctor of Philosophy

Department of Economics

University of Sussex

December 2011

UNIVERSITY OF SUSSEX

OUTI KORKEALA, DOCTOR OF PHILOSOPHY

CLIMATE AND LAND IN TURMOIL: WELFARE IMPACTS OF EXTREME WEATHER EVENTS AND PALM OIL PRODUCTION EXPANSION IN INDONESIA

SUMMARY

Climate variability and climate change have become important research topics also in economics. The objective of this thesis is not to forecast the future but to learn from the past by studying how two important climate change-related topics have affected Indonesian households. Delayed monsoon onset, *El Niño*, will become more frequent with climate change whereas palm oil production is a contributor to climate change.

The first essay examines how variability in monsoon onset affects rural households' welfare in terms of household expenditure and farm profits. Using the Indonesia Family Life Survey (IFLS) data I find that households in the middle tercile of the expenditure distribution face the biggest albeit temporary losses from delayed monsoon onset. Half of the expenditure decline is due to increase in household size. Conditional on onset, rainfall intensity has only minor effects.

The second essay uses the IFLS data to study how schooling and child labour are affected by delayed monsoon onset. The probability of continuing from primary to secondary school is reduced when a delayed onset coincides with the transition year. In other respects, monsoon onset does not affect education of rural children. However, riskier distribution of rain postpones school entry for young children. Moreover, delayed onset increases child labour.

Using district-level data on palm oil production and area planted and national household survey (SUSENAS) the third essay studies the impact of oil palm expansion on household expenditure and health. Instrumental variable estimates exploit the historical production and district forest area as an exogenous source of variation. I find that smallholder production has a weak negative impact on household expenditure but this effect is not present among rural households. More, total production increases incidence of asthma in Kalimantan. The findings suggest that palm oil is not a panacea to increase rural welfare.

Acknowledgments

I have been fortunate to complete my PhD at the University of Sussex. The friendly and inspiring environment at the Department of Economics has greatly assisted my work. In addition to the Department I have enjoyed fruitful discussions in Indonesia, Finland, and at the final stage in Oxford, and here I would like to thank those who have given me valuable comments, support and encouragement.

First of all, I would like to express my sincere gratitude for my supervisor Professor Andy McKay. Throughout my studies he has shown great professionalism, patience and support. His comments have always been to the point. No matter how busy he was, he has always found time for long discussions. Not only has he given advice on my research but also assisted in career planning.

I would also like to thank the faculty at the department for interesting seminars and discussions, especially Professor Richard Dickens. Special thanks to my second supervisor Andy Newell and also to Professor Barry Reilly for supervising my master degree's dissertation and for encouraging me to continue for the DPhil.

I would like to thank various persons and institutions who have helped me accessing data: John Strauss, Kathleen Beegle, Bastian Zaini, Dewi Kirono, Krystof Obizdinski, NOAA, CEREGE and Indonesian Meteorological Agency.

In Indonesia Mafalda Duarte provided me the great opportunity to work as a consultant at the World Bank Jakarta Office. It was a great experience to meet people and see the amazing country. While travelling across the country the IFLS datasets became real and tangible and gave me motivation to carry on. David Newhouse taught me a great deal about econometrics and research, and I am grateful to have him as a co-author.

I am privileged to have completed the PhD internship at the WIDER Institute in Helsinki. Again I enjoyed an excellent research environment and useful discussions with the staff. I would like to thank Wim Naudé, Finn Tarp, James Thurlow, and Luc Christiaensen for their comments and support.

I am very grateful for Yrjö Jahnsson Foundation for funding my DPhil studies. The generous funding enabled me to focus solely on my research. I have received financial support for this project also from Finnish Cultural Foundation and the Otto A. Malm Foundation and University of Sussex provided funding for my conference fees.

In Oxford I would like to thank the Department for Economics for providing excellent development seminars and Kate and Martina for good discussions. The Social Science Library provided me good facilities for studying and working.

I am very lucky to have friends who not only are invaluable as friends but also top economic researchers in their fields. Thank you Janne and Teemu for your advice, support and comments. Without you I would have struggled much more and longer. I would also like to thank my other friends in Finland for supporting me on this rewarding but at times lonely process, especially Jenni, Lotta, Katri, Marika, Karoliina, Paula, Miika, Jussi, Elina and the rest of Möllisö. Also friends in UK, both in Brighton and in Oxford, have been invaluable throughout the process, thank you Luisa, Sophia, Elisa, Elsa, Marta, Lena, Sanna, Tobias, Suvi, Elina, Elias, Laura, Leandro, Anna, Dimitris, Diego, Marian, Sandrine, Anya and Hannah. Special thanks to Penny for proof-reading.

We had a lovely group of research students at Sussex. Thanks Estela and Gunilla for sharing the office with me during the first year, it was not only great fun but also a very productive arrangement. Thanks Stefania for helping with Stata coding and econometrics. Thanks Emilie for your guidance on how to put the papers together.

I am also very much obliged to my parents Hanna and Hannu for providing such a supportive environment to grow up. Moreover, education and science have always been top priorities in our family. You have also provided the much needed financial support. In addition my Mum provided her help and support by visiting us in Oxford and babysitting Siiri. I am lucky to have the two best brothers in the world, thank you Mauri and Markku.

And to my beloved family, Tuomas and Siiri, thank you for being there for me. Tuomas, I am indebted to you for your loving support and encouragement and for the careful proof-reading. And my daughter Siiri, when I look at you I remember how important it is to have dreams. Now, having completed this thesis, one of my big dreams has become true.

Contents

List of Figures

List of Tables

List of Abbreviations

2SLS	Two-Stage Least Squares
BMG	Badan Meteorologi dan Geofisika
BPS	Central Bureau of Statistics
CEREGE	Centre Européen de Recherche et d’Enseignement des Géosciences de l’Environnement
CIFOR	Center for International Forestry Research
ENSO	El Niño Southern Oscillation
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
FFB	Fresh Fruit Bunches
GDP	Gross Domestic Product
GIS	Geographic Information System
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IV	Instrumental Variable
IFLS	Indonesia Family Life Survey
KKPA	Koperasi Kredit Primer untuk Anggota
LPM	Linear Probability Model
MT	Metric tonne (1000 kg)
NES	Nucleus Estate Smallholder System

NOAA	National Oceanic and Atmospheric Administration
OLS	Ordinary Least Squares
PC	per capita
PCE	Per Capita Expenditure
PIR	Perkebunan Inti Rakyat
PODES	Potensi Desa, Village Potential Statistics
RP	Indonesian Rupiah
SOND	September, October, November and December
SSTA	Sea Surface Temperature Anomaly
SUSENAS	National Socioeconomic Survey (Survey Sosial Ekonomi Nasional)
USDA	United States Department of Agriculture
WIDER	World Institute for Development Economics Research

Chapter 1 Introduction

This thesis consists of three self-contained essays analysing household welfare with respect to two important environment-related topics in Indonesia, timing of monsoon onset and the expansion of palm oil production. Delays in monsoon onset will become more frequent with climate change; on the other hand clearing tropical forest for oil palm plantations is an important contributor to CO_2 emissions and thus, climate change. In the first two essays I study how the timing of monsoon onset affects household welfare and in the third and final essay I study the impact of the expansion of palm oil production on household welfare. This thesis contributes to the large body of literature on welfare impacts of weather shocks (see, for example, Rose, 2001; Dercon, 2004; Kazianga and Udry, 2006; Carter et al., 2007; Maccini and Young, 2009). Indonesia is an interesting choice of study given that it is the world's largest producer of palm oil and also an important producer of rice whose harvest is frequently threatened by delays in monsoon onset. Moreover, the expansion of palm oil production is accused for forest and environmental degradation. However, how Indonesian households are affected by the palm oil production expansion and delays in monsoon onset, remain relatively unknown.

The timing of monsoon onset, i.e. the start of the monsoon rains, and particularly the delay in the monsoon onset, has proved to be an important factor determining rice production in Indonesia (Naylor et al., 2001 and Naylor et al., 2007a). Long delays in monsoon onset, which is also referred to *El Niño* in Indonesia, postpone rice planting and adversely affects area planted, often driving up domestic and international rice prices (Falcon et al., 2004). Further, the current consensus predicts that Asian monsoon will intensify in the future along with global warming, implying more and longer droughts in Indonesia (Overpeck and Cole, 2007). Despite the well-documented relationship between monsoon onset and rice production, the socio-economic impacts of delayed monsoon onset have not been studied thus far.[1] Using the Indonesia Family Life Survey, IFLS, I focus on

[1] Following the publication of the previous version of this study as a working paper (Korkeala et al., 2009), a similar study using a subset of the data was released (Skoufias, et al., 2011). Our paper differs by focusing on

two important monetary outcomes in a rural setting, household per capita expenditure and per capita farm profits. In addition, in the second essay, I introduce individual level outcomes in terms of education and child labour. Education is a widely used measure of socio-economic status and both education and child labour have long-lasting effects in earnings potential and overall welfare later in life. More, the IFLS data has been extensively used to study education outcomes but the effect of weather shock and risk on education is less explored in previous studies.

Indonesia is home for the third largest share of the global forest cover after Brazil and Democratic Republic of Congo. One of the major threats to forest cover, and therefore also to global warming, is the expansion of palm oil production. According to the FAO (2005) over 56% of the Indonesian oil palm plantation expansion between the years 1990-2005 realized at the expense of natural forest cover. Indonesia, where total land area devoted to oil palm increased over 2100 per cent since the early 1980s, is now the largest producer of palm oil in the world. In addition to deforestation, palm oil producers are accused of forest fires, soaring food prices and environmental degradation due to toxic waste being released by refineries (see, for example, Naylor et al., 2007b; Sheil et al., 2009; McCarthy and Zen, 2010; Rist et al., 2010). On the other hand, proponents of palm oil claim that the expansion will lead to an increase in rural welfare through employment opportunities and strong spillover effects. Again, there is little systematic empirical evidence on how households are affected by the expansion of palm oil production. The main welfare outcome used is household per capita expenditure. In addition, I study the effect of palm oil production on individual health, measured as the probability of an individual reporting symptoms of asthma or difficulties in breathing. Again, per capita expenditure is the standard measure used in welfare analysis while the measure of the health outcome was selected to evaluate the non-monetary impacts of the expansion. The choice of asthma is grounded on previous studies suggesting that both forest fires and toxic waste correlate with asthma (National Research Council, 1991 and Osterman and Bauer, 2001).

---

the distributional effect of delayed onset. In addition, we utilize data from outside Java, and include three rather than one wave of panel data, allowing us to identify the impact of rainfall shocks with greater precision.

All essays use household-level data combined with either rainfall data or palm oil data. The first essay utilises first three rounds of the IFLS data, and the 1993-2004 rounds of the annual national socio-economic survey, SUSENAS data, together with daily rainfall data. The second essay uses first three rounds of the IFLS data together with daily rainfall data. And finally, the third essay utilises the 2004-2008 rounds of the SUSENAS data, together with district-level data on palm oil plantations and production as well as the 2003 Village Survey, PODES data. Therefore, each chapter has its own data section and a section on empirical methodology. All essays use reduced form specifications, i.e. the variable of interest, whether monsoon onset or palm oil, enters the estimation equation directly. This method enables me clearly to inform the policy formers about the effects of delayed onset and the expansion of palm oil production. In the following I present the contribution of this thesis separately for each essay, accompanied by a short description of the empirical methodology and the results obtained.

The first essay of this thesis draws on the joint work with Mafalda Duarte and David Newhouse (see Korkeala et al., 2009). A revised version of that article was submitted to an academic journal together with David Newhouse and this paper is currently under review.

In the first essay I study the impact of monsoon onset on per capita farm profits and per capita expenditure in rural Indonesia. An important contribution of the essay is to study the heterogeneity among households in order to identify which households are the most vulnerable to delayed monsoon onset. The panel dimension of the data allows me to divide households into expenditure terciles according to their mean per capita expenditure across the surveys.[2] Economic theory does not provide unambiguous guidance on this matter. Poor households may be most vulnerable to climatic shocks due to limited access to formal insurance mechanisms, finance, and irrigated crop land. On the other hand poor households may adapt low risk low return strategies protecting them from weather shocks (Rosenzweig and Binswanger, 1993). Another significant contribution of the essay is to investigate the

[2] However, assumption on homogenous impacts within a group might be too strong. If, for example, two households start in the top of the expenditure quintile and a shock hits and then one household is severely affected, that household is likely to be in a lower tercile than the unaffected household. In order to address this problem I also present an alternative method to group household into expenditure terciles. I regress the average household per capita expenditure on predetermined characteristics and then predict the per capita expenditure. Finally, I use the predicted values for creating expenditure terciles.

importance of the intensity of rainfall relative to the timing of rainfall in terms of household welfare.

I estimate a reduced form model where monsoon onset, defined as the deviation of the mean date expressed in standard deviations, enter the estimation equation directly. A household fixed effect model is employed and the use of linear splines allows a non-linear effect of monsoon onset on per capita farm profits and per capita expenditure. Using the IFLS data the estimation results suggest that households in the middle tercile of the expenditure distribution are the most vulnerable to delayed monsoon onset: for these households reduction in per capita expenditure reflects the reductions in per capita farm profits. Moreover, rainfall intensity has a limited impact on per capita expenditure after the timing of the rainfall is controlled for. However, SUSENAS data suggest that variation in monsoon onset has a negligible impact on household per capita expenditure. I discuss the relative merits of the two data sets in the current application and conclude that the IFLS data is the preferred data in this context.

In the second essay I investigate the impact of monsoon onset on school attendance and child labour in rural Indonesia. This chapter expands the welfare analysis presented in the first essay by introducing individual level outcomes. In addition to monsoon onset I study the impact of weather risk, measured as the coefficient of variation of monsoon onset, on the probability of children entering school. This essay contributes to the literature by identifying the effects of a weather shock, and hence, agricultural production shock on education and child labour in rural Indonesia. The existing studies on child labour in Indonesia are largely descriptive (see, for example, Manning 2000; Priyambada et al., 2005). In addition, previous studies on education outcomes in Indonesia (see, for example Fitzsimons, 2007) have not considered the two important corner stones of schooling; transfer to secondary school and the probability of entering school. I employ a pooled probit model and as a robustness check I present an instrumental variable model in the linear probability framework which addresses the endogeneity of the household per capita expenditure variable by instrumenting it with the value of household's land holdings. I find that delayed monsoon onset increases child labour. With respect to education I find that delayed monsoon onset coinciding with the transition year from primary to secondary school reduces the probability of attending school in the following years. In addition, parents delay the enrolment of their younger children in riskier environments.

In the third essay I evaluate welfare impacts of the expansion of palm oil production in Indonesia. Despite being one of the most important topics in Indonesia, in terms of environmental policy, climate change and rural development, there is strikingly little systematic empirical research on how households are affected by the expansion of palm oil production. This essay provides the first study on the socio-economic impacts of palm oil production expansion using large samples of survey data. Palm oil production requires timely access to mill and good infrastructure, suggesting that individual farmers might be at a disadvantageous position compared to large companies. Critics claim that large companies have benefitted most from the expansion (World Bank, 2010). Smallholders account for around 40% of the total national production. I utilise district-level data on smallholder oil palm plantations and smallholder palm oil production together with SUSENAS household data. The proposed empirical approach allows me to study the overall welfare effect (household consumption, individual health) but not to separate the effects on actual producers. To address the possible endogeneity of the district-level palm oil measure I employ an instrumental variable regression where the predicted values of palm oil are used as an instrument for actual values. As an exogenous source of variation in the oil palm plantations and palm oil production, I use the historical values of plantations and production as well as the historical measure of district-level forest area. I discuss validity and the relative merits of the chosen instruments more in detail in the chapter. In addition to smallholder production, I investigate the impact of total palm oil area and production (including both large plantations and smallholders) in selected provinces in Kalimantan.

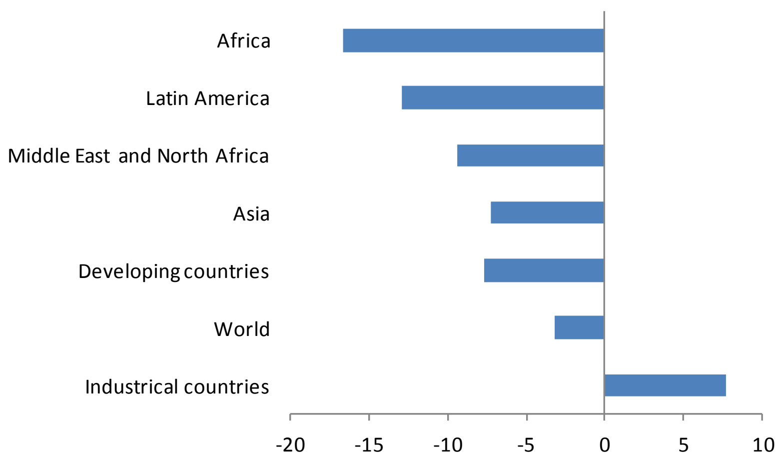
The results suggest that smallholder production has a weak negative impact on household expenditure, but the effect is not present when restricting the sample to rural households only. More, total area and total production increase incidence of asthma in selected provinces in Kalimantan. I argue that palm oil is not a panacea to increase rural welfare and that there is no evidence of positive spillover effects.

Chapter 2 Evaluating the impact of monsoon onset on household welfare in rural Indonesia

2.1 Introduction

Over 60 per cent of the world's poor reside in rural areas, and a significant share of their income is derived, either directly or indirectly, from agriculture. These households are potentially at risk from the projected increases in climate variability and extreme weather events resulting from global warming (see figure 2.1).[1] Households in developing counties near the equator, many of whom have already reached the optimal temperature for agricultural activities, are most vulnerable to further temperature increases (Cline, 2007).

Figure 2.1. Change in the agricultural output potential (2080 as % of 2000 potential).



Note: effects are with carbon fertilisation. Source: Cline (2007).

[1] Increased temperatures have caused an estimated combined loss of wheat, maize, and barley equal to roughly $5 billion per year between 1981 and 2002 (Lobell and Field, 2007). A temperature rise of 2.0 °C and an 8 per cent increase in precipitation, in the absence of carbon fertilization, could lead to a 12 per cent reduction in agricultural revenue in Brazil and a 20 per cent reduction in India (Sanghi and Mendelsohn, 2008).

This chapter examines the effects of volatility in the timing of the rainfall on households' economic welfare in rural Indonesia. Indonesia is an important case study to investigate the impacts of climate variability, given the country's size and dependence on agriculture.[2] While previous studies have examined the impact of delayed monsoon onset on national rice production, I am not aware of other studies examining the implications of delayed onset on the welfare of Indonesian households.[3] Two main indicators for household welfare are used: household per capita expenditure and per capita farm profits. The former is the basis for poverty measurement in Indonesia, while the latter is an important source of income in rural areas.

Households' response to shocks has important implications for both their short and long-term income generating prospects. Well-developed financial and insurance markets, where they exist, help insulate household consumption from shocks. In developing countries, however, formal financial and insurance markets are typically limited. As a result, most *ex-ante* and *ex-post* strategies for coping with shocks are costly, and shocks often reduce consumption and asset holdings.[4]

A key contribution of this study is to examine the distributional impact of climate variability, by identifying how the impact of variation in monsoon onset depends on the economic well-being of the household. Wealthier households are likely to be well-protected from rainfall shocks, as they tend to have greater access to informal insurance networks through relatives or other sources, and irrigated farmland that offers partial protection against droughts. Effects on poor households are difficult to infer from theory, however. Poorer households may be particularly vulnerable to delayed onset, if they cannot access formal or informal insurance and irrigation. But poor households may also be less vulnerable to climactic shocks, to the extent that they protect themselves, at the expense of higher expected income, by adopting low risk and low return strategies, for example by

[2] One estimate suggests that Indonesian agricultural output could be reduced by 5.6% (with carbon fertilizer) or 17.6% (without carbon fertilizer) by 2080 attributable to climate change (Cline, 2007).

[3] Following the publication of the former version of this study as a working paper (Korkeala et al., 2009), a similar study using a subset of the data was released (Skoufias, et al., 2011). Our paper differs by focusing on the distributional effect of delayed onset. In addition, we utilize data from outside Java, and include three rather than one wave of panel data, allowing us to identify the impact of rainfall shocks with greater precision.

[4] For example, in response to the 1998 financial crisis in Indonesia, households sold non-productive assets such as jewelry (Frankenberg et al., 2003).

choosing low risk and return crops (Rosenzweig and Binswanger, 1993; Zimmerman and Carter, 2003). The use of longitudinal data in this study allows for the separate estimation of the effect of delayed onset on groups of households defined on the basis of their time-invariant household characteristics, including average per capita expenditure of the household.

The second objective of the study is to examine the importance of the timing of rainfall, relative to its intensity, in explaining changes in the household welfare. While most studies on the effects of rainfall examine variation in intensity rather than onset, the planting season in Southeast Asia is structured such that the timing of the monsoon may be more important than rainfall intensity in determining the success of the harvest.[5] In addition, I examine the role of delayed onset on rice prices, which may be an important mechanism through which household welfare is affected.

Using the IFLS data the results show that delayed monsoon onset has minor effects on all rural households, but substantial effects on the per capita farm profits and expenditure of middle-class households (defined as the second expenditure tercile). One standard deviation, or 24 day, delay in monsoon onset reduces farm profits per capita for middle-class farmers by 13,500 rupiah per month. This reduction represents 44% of per capita profits, and 11% of total profits of the middle-class farmers. Falls in farm income are reflected in a drop in expenditure, as delayed monsoon onset reduces per capita expenditure of the middle-class households by 15.1 percent. Approximately half of the reduction in per capita expenditure is realized through an increase in household size. These negative effects are short-lived, however. Conditional on the prior year's onset date, there is no evidence suggesting that delayed onset two years ago reduces expenditure. This finding is robust to the use of rainfall intensity instead of the timing of monsoon onset. After conditioning on onset, however, variation in intensity has substantially smaller effects on household welfare, suggesting that the date of onset is the key factor. Finally, delayed onset increases the local price of rice, which helps mitigate its negative effects for net rice sellers. The findings are also robust to different functional forms as well as non-parametric estimation. However, using the SUSENAS data, delayed monsoon onset has only a negligible effect on

[5] Delayed onset in Indonesia is associated with reduced aggregate rice production, even if average rainfall levels followed a delayed onset (Naylor et al., 2001; Naylor et al., 2007a).

household expenditure. The plausible explanations for the different results are discussed in section 2.5.2.

The remainder of the chapter is structured as follows. Section 2.2 provides a short introduction to rice farming in Indonesia and the previous literature. Section 2.3 introduces the data and the estimation strategy. Section 2.4 discusses the estimation results for the IFLS data and section 2.5 for the SUSENAS data. Finally, section 2.6 concludes, discusses policy implications and outlines areas for future research.

2.2 Background

2.2.1 Climate variability and rice farming in Indonesia

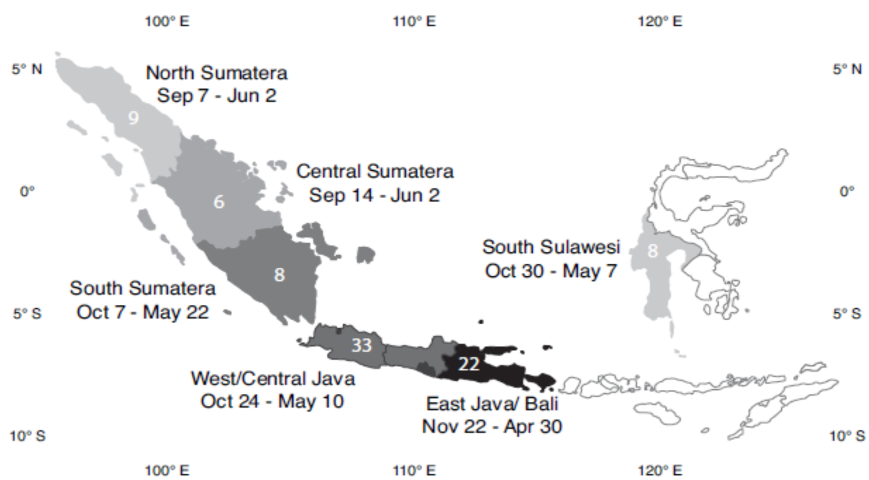
In Indonesia, rainfall patterns are the most important source of climactic uncertainty, since the country's proximity to the equator limits variation in temperature. Rainfall patterns vary greatly, both across years and districts within a year, and long delays in monsoon onset occur periodically. In the 20 years preceding 2004, monsoon onset started 30 days later than usual in 18 per cent of the years in West Java, and 10 per cent of the years in East Java/Bali (Naylor et al., 2007a).

Rainfall patterns are important in rural Indonesia because the agricultural sector continues to employ the majority of poor workers. While agriculture's contribution to GDP fell from 47 per cent in 1969 to around 13 per cent in 2006, the sector currently accounts for 60 per cent of rural employment, declining only slightly from 70 per cent in 1990. Two thirds of the households in the bottom two expenditure quintiles work in agriculture (Kishore et al., 2000; World Bank, 2008).

Only a minority of farm households have access to irrigated farmland, and rain-fed agriculture continues to play an important role in agricultural production. The distribution of irrigated rice land (*sawah*) is skewed towards larger landholders: three quarters of agricultural households own less than 0.5 hectares each, and together they control only 38% of *sawah* in Indonesia (McCulloch, 2008).

Typically, the bulk of the rice crop is planted at the beginning of the rainy season between October and December, though this varies by region (see figure 2.2). Planting begins when there is sufficient moisture to prepare the land for cultivation and facilitate early rooting. The main planting period ends before the peak of the monsoon, because excessive water hampers rooting. During the 3-4 months grow-out period from planting to harvest, rice requires 600-1200 mm of water depending on the agro-ecosystem and the timing of the rainfall or irrigation. A smaller, dry season planting takes then place in April and May after the wet season crop is harvested (De Datta, 1981 in Naylor et al., 2001). Figure 2.2 also shows the share of rice that is produced in each of the provinces.

Figure 2.2. The timing of the rainy season and the share of rice production out of total production in selected provinces of Indonesia.



Note: Onset date is the date past August 1 when accumulated rainfall equals 20 cm, averaged over reporting rainfall stations in the region for the years 1979–2004; termination date is the date on which 90% of that year’s rainfall has accumulated. The number on each province indicates the share of province’s rice production out of total production.
Source: Naylor et al., (2007a).

The timing of the onset of the monsoon is affected by the El Niño Southern Oscillation (ENSO), which causes anomalies in the sea surface temperature and sea-level pressure. El

Niño events can delay rice planting by up to two months, reducing the area cultivated and delaying the plantings of next year's dry-season crop.[6] In addition to delays in monsoon onset, *El Niño* events are associated with reductions in the length of the rainy season (Cook et al., 2001).

The timing of the monsoon affects the total amount of land planted for many crops in Indonesia, and is particularly important for rice. Variation in a sea-surface temperature index explains 60 per cent of rice planted in Java, and 40 per cent of the variation in rice production (Naylor et al., 2001). From 1983-2004, a 30-day delay in monsoon onset caused rice output to fall, on average, by 580,000 metric tons (11.6%) in East Java/Bali and 540,000 metric tons (6%) in West/Central Java during the main rice harvest season between January and April (Naylor et al., 2007a).

Weather variation reduces production by decreasing the area cultivated rather than by reducing the yield. In Indonesia, each degree increase in the Sea Surface Temperature Anomaly (SSTA) reduces the national cultivation area by an estimated 261,000 hectares (2.3 percent) which reduces production by 1,318 tmt (2.6 percent), production reduction by province is presented in table 2.1.[7] Most of the fall in production comes from Java although the largest percentage effects are in Kalimantan and Sulawesi. Similar patterns between weather shocks and area cultivated are observed also in other countries. For example in India the variability in area cultivated is higher than in yield (Walker and Ryan, 1990).

[6] During El Nino events, the warmer ocean water shifts eastward away from Indonesia causing rain to fall over the central Pacific Ocean.

[7] According to the Indonesian Statistical Office (Badan Pusat Statistik, BPS) the area harvested in Indonesia was 11,499,997 hectares and the rice production was 50,460,782 tons in 2001 (http://www.bps.go.id/tnmn_pgn.php?eng=0, accessed 5 Feb 2010).

Table 2.1**.** Estimated effects of a one degree Celsius increase in August SSTA on rice production (thousand metric tons, tmt), by province, 1983-2002.

Province	Crop-Year Production Effect (Sep-Aug) (tmt)	Percentage of National Effect	Significance of Production Effect (t-statistic)	Ratio of Production Effect to Average Yearly Production 1997/98 – 2001/02
West Java	-380	28.83	-3.01	-0.037
Central Java	-238	18.06	-3.67	-0.026
East Java	-232	17.60	-4.06	-0.026
South Sulawesi	-102	7.74	-2.02	-0.033
North Sumatra	-54	4.10	-1.57	-0.016
West Sumatra	-46	3.49	-2.18	-0.026
East Kalimantan	-41	3.11	-2.60	-0.118
North Sulawesi	-38	2.88	-3.31	-0.104
West Nusa Tenggara	-30	2.28	-2.63	-0.021
Riau	-17	1.29	-2.14	-0.041
Southeast Sulawesi	-10	0.76	-1.68	-0.033
Bali	-3	0.23	-2.82	-0.003
Subtotal	*-1.191*	*90.36*		
Coefficient for all Indonesia	-1.318	100.0		

Source: Falcon et al., (2004).

2.2.2 Household responses to weather shocks

Existing evidence on the extent to which households are insured against shocks, including weather shocks, is mixed. Evidence from Thailand and India, for example, suggests that households are largely able to smooth consumption in response to rainfall shocks. In Thailand, farmers tend to save during favourable rainfall years, and use these savings to protect consumption from income shocks (Paxson, 1992). In ICRISAT villages in India, on the other hand, unemployment and sickness are only weakly associated with household consumption after village-level risk (i.e. weather) is controlled for. Credit markets and gifts, as well as asset sales, appear to smooth much of the fluctuations in income (Townsend, 1994). Coping mechanisms in India also include pulling poor children out of school

(Jacoby and Skoufias, 1997).[8] Households are not, however, able to smooth consumption in response to rainfall shocks in all low-income countries. In Bangladesh children's growth was adversely affected, especially in landless households, after the 1998 floods (Foster 1995). In Ethiopia, common shocks such as rainfall shocks reduce growth in household consumption (Dercon, 2004). In Burkina Faso, there appeared to be little evidence of consumption smoothing during a drought (Kazianga and Udry, 2006).

Several studies indicate that households in Indonesia have been, and continue to be, vulnerable to rainfall shocks. Early-life drought between 1953 and 1974 adversely affected health, education attainment, and adult socio-economic status for women in rural areas (Maccini and Yang, 2009). Self-reported crop loss was associated with reduced education expenditure (Cameron and Worswick, 2001). Low rainfall in specific quarters has, in the past, correlated with substantial and lasting reductions in farmers' income (Newhouse 2005). Finally, one study has considered the impact of delayed onset on household expenditure among rural farmers in Java (Skoufias et al., 2011). That study considers households on rural Java in 2000 and finds a small and statistically insignificant penalty of delayed monsoon onset. A larger 15 per cent reduction is estimated for rice farmers who experience low rainfall following the monsoon. This is based on the results from a single cross-section of households matched to data from 18 rain stations, raising questions about the generalizability of the results. This study builds on this past literature for Indonesia by looking specifically at the impact of delayed onset on farm profits and per capita expenditures for different types of households.

[8] Chapter three examines this issue in Indonesia.

2.3 Data and methodology

2.3.1 Data sources

To obtain empirical evidence on the welfare effects of delayed onset, I use the first three rounds of the Indonesia Family Life Survey (IFLS), which is the main data source in this study. IFLS waves are surveyed in autumn. The fourth and most recent round is excluded because I was unable to obtain more recent rainfall data. The IFLS is a longitudinal household survey that began in 1993 with roughly 7,200 households taken from 320 communities in 13 out of Indonesia's 33 provinces. The 13 provinces, which overlap nearly perfectly with figure 2.2, cover 83 per cent of the population and 85 per cent of national rice production. The subsequent rounds of the survey, which were conducted in 1997, 2000, and 2007 attempted to re-contact all households interviewed in 1993, and household attrition rates were generally below five percent. Split-off households, which were tracked as long as they remained in the 13 provinces, are included in the sample but treated as new households. The exclusion of urban areas limits the total sample to 11,400 observations on 4,000 households, taken from 181 communities of 1993.[9]

The main indicator of household welfare is per capita expenditure. This is derived from a consumption module consisting of 37 food and 19 non-food items, with purchased and self-consumption reported separately. Household per capita expenditure is equal to the sum of all expenditures in the past month, typically reported by the head's spouse, deflated by a local price index and divided by the number of persons in the household.[10,11] There is an inconsistency in the data collection, as the 1993 questionnaire, unlike subsequent waves,

[9] Additional information about the survey is provided in Strauss et al., (2004), Frankenberg and Thomas (2000), and Frankenberg and Karoly (1995).

[10] For IFLS1, per capita expenditure is taken from the expend2.dta file provided in the re-release of IFLS, which imputes non-food expenditures for households that were not asked to report them. In addition, because no price index is provided for the 1993 data, the national price index for the local provincial capital is used to deflate expenditures from 1993 to 1997.

[11] Price series for the rural villages for 1997 and 2000 are from the "Survey Harga Konsumen Pedesaan" (Rural Consumer Price Survey). The respondents of this survey are farmers/farm workers. This series is a province-level series.

asked for personal rather than total household consumption for non-food items.[12] The resulting differences in household consumption do not, however, appear to substantially affect the results presented below.[13]

The rainfall data was taken from the NOAA (National Oceanic and Atmospheric Administration) Global Summary of the Day combined with additional data obtained from the Indonesian Meteorological Agency (*Badan Meteorologi dan Geofisika, BMG*). Imputed values for the missing values in the Global Summary of the Day data were provided by CEREGE, (*Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement)*.[14] Unfortunately, the most recent year for which daily rainfall data was available is 2004.

The rainfall data cover 52 stations, of which 36 match with IFLS households and 49 match with the SUSENAS households. In the IFLS data households were matched to the nearest weather station at the community level, for households interviewed in the original set of 1993 villages. Households that had moved to another village were matched to the rainfall station closest to the geographic centre of their new district.

The start of the monsoon is defined as the day that cumulative rainfall since August 1 exceeds 20 cm, following Naylor et al., (2007a). The underlying rationale is that approximately 20 cm of cumulative rainfall is needed to moisten the ground for rice planting. For each station, the start date of the onset was calculated for each year and then standardized using each station's "leave-out" mean and standard deviation across years. A standardized value of zero would therefore indicate that the nearest station's monsoon onset last year was equal to its historical average. The standard deviation of monsoon onset across the entire sample is 24 days. Alternative definitions of monsoon onset are also available, such as in Moron et al. (2009) that takes into account false starts, that is, dry spells occurring after the threshold. Nevertheless, Moron et al. (2009) argue that their

[12] In 1993, the consumption questionnaire asked the respondents how much they had purchased in the past month, and whether this was equal to the total amount purchased by the household.

[13] I excluded the households in 1993 that stated that the reported non-food expenditure is not for the entire household or that they do not know whether it is. The impact of excluding these households to the point estimates is negligible; see section 2.5.2 for further discussion.

[14] The quality of the rainfall data is discussed in Appendix A.

estimations of mean onset dates for various regions in Indonesia are highly correlated with the definition used herein.

2.3.2 Estimation Strategy

My estimation strategy allows for non-linear effects of delayed onset, given the likelihood that the cost of delayed onset accelerates as the length of the delay increases. This would occur, for example, if farmers faced with small delays forego planting on their lowest-quality soil, or if larger delays inhibit informal insurance networks. My preferred method estimates a spline (a piecewise linear regressions) with one knot, which maintains some flexibility and is relatively straightforward to interpret. I check robustness by estimating flexible non-parametric regressions as well as quadratic functions. The knot of the spline was selected to be -0.6 standard deviations, based on the results of the non-parametric regressions. It is apparent that the number of linear splines could be more than two but this proved unnecessary in the current application.

To estimate the effect of monsoon onset on household per capita expenditure and farm profits, I estimate the following spline models with two lags of onset:

$$\ln C_{it} = \alpha_1 + \beta_1 spline_{1,t-1} + \beta_2 spline_{2,t-1} + \beta_3 spline_{1,t-2} + \beta_4 spline_{2,t-2} + \alpha_i + D_t + \varepsilon_{it} \qquad (2.1)$$

$$F_{it} + \alpha_1 + \beta_5 spline_{1,t-1} + \beta_6 spline_{2,t-1} + \beta_7 spline_{1,t-2} + \beta_8 spline_{2,t-2} + \alpha_i + D_t + \varepsilon_{it}, \qquad (2.2)$$

where below the threshold value, $O_{it-1} \leq -0.6$:

$$spline_{1,t-1} = O_{it-1}$$
$$spline_{2,t-1} = 0.$$

And above the threshold value $O_{it-1} > -0.6$:

$$spline_{1,t-1} = -0.6$$
$$spline_{2,t-1} = O_{it-1} - (-0.6).$$

The onset variables from two years ago are defined analogously. C_{it} represents the per capita expenditure of household i in year t, while F_{it} is monthly per capita farm profits of household i in year t. α_i represents a household i's fixed effect, which captures all time-invariant characteristics of the household, including all household characteristics determined prior to 1993. D_t represents a vector of dummy variables for each survey year. Farm profits per capita are expressed in levels rather than logs, in order to allow negative values for profits. ε_{it} is a stochastic error term, which is robust to heteroscedasticity and clustered on rain station, to allow for unobserved correlation between households that have been matched to the same rain station. In all regressions, households are weighted by their mean sample weight over the course of the three survey waves.[15]

O_{it-1} stands for monsoon onset the previous year, and it indicates the timing of monsoon onset at the nearest weather stations in the previous year. As onset is standardized, a value of O_{it-1} equal to zero indicates that the previous year's monsoon, according to the nearest rain station, arrived at the same time as the historical average. Values of O_{it-1} equal to one and negative one indicate that last year's monsoon arrived one standard deviation (24 days) late or early, respectively. Parameters β_1, β_3, β_5, β_7 represent coefficients for the slope of the left linear segment of onset, while parameters β_2, β_4, β_6, β_8 represent coefficients for the slope of the right linear segment.

Two lags of onset are included because rainfall exhibits negative serial correlation in the sample.[16] Including an additional lag of delayed onset reduces omitted variable bias if the

---

[15] However, consistent results are obtained without weighting.
[16] The correlation between rainfall and lagged rainfall in the data is -0.33.

second lag of monsoon onset influences per capita expenditure and/or farm profits. Finally, examining the effect of variation in onset two years ago can shed light on whether the effects of climactic shocks persist for two years. For these reasons, despite the reduction in the precision of the estimates, my preferred specification includes an additional lag of monsoon onset. However, in the Appendix tables I also present estimation results for specifications with the first lag of monsoon onset only.

The model is re-estimated separately for farm and non-farm households. Per capita farm profits are only estimated for farm households, defined as those owning a farm in 1993. Owning a farm in 1993 is time-invariant, meaning that it is orthogonal to the residual ε_{it} in equations (2.1) and (2.2).

In addition, I re-estimate the model separately for poor, middle-class and rich households. I use two methods to determine household class. The first involves taking the tercile of their average real per capita expenditure, over the course of the three surveys. If the impact of rainfall shock is constant across all households within an income group, stratifying the sample based on these time-invariant characteristics does not directly introduce bias into the estimates.[17]

The assumption that the effect of rainfall shocks is constant is strong, however. Households that suffer a large loss following a rainfall shock are more likely to be in the bottom tercile according to their average per capita expenditure. This could lead the results to overstate the extent to which poor households are vulnerable to rainfall shocks. Thus, I employ an alternative strategy to group households, based on a welfare indicator that is predetermined with respect to rainfall shocks.

This alternative strategy involves estimating the portion of average household expenditure that is predetermined three years before the household enters the survey, prior to the earliest rainfall shock used in the analysis.[18] To do this, I regress the average per

[17] Two other sources of bias may be present: attenuation bias due to measurement error in average per capita expenditure measure, as well as correlation between the impact of the shock and welfare status if households' response to shocks is heterogeneous. These issues are discussed below.

[18] While dynamic panel data models are commonly utilized to control for past consumption (i.e. models with lagged dependent variable), they cannot be employed in this case because there is insufficient data to sacrifice one of the three years, and because the interval between panels is not constant across years.

capita expenditure of the household over the course of the panel on several retrospective variables as follows:

$$\overline{C_i} = \alpha + \beta Z_i + \varepsilon_i, \tag{2.3}$$

where $\overline{C_i}$ represents the per capita expenditure of household, averaged over each year in which the household appears in the panel survey. Z_i is a vector of retrospective predetermined variables for household i, including the age and education of the household head, as well as a set of indicators for his or her district of residence at age 12, a dummy for whether the head worked three years before entering the panel, and for those that did, the number of hours and weeks worked. Each of these variables was determined prior to the first rainfall shock considered in the analysis.[19] Households were then classified into terciles based on their predicted per capita expenditure. Appendix table A3 displays the results of estimating equation (2.3).

This additional robustness, however, comes at a cost. First, a full set of predetermined variables is available for only 92 per cent of the sample, which limits the representativeness of the results. In addition, this procedure raises the prospect that predicted expenditure is an inaccurate indicator of household economic welfare, particularly given that the regressors in (2.3) only explain 11 per cent of the variation in average household per capita expenditure. Therefore, I treat this alternative method as an important robustness check.

Finally, identification of the causal effect of delayed onset is based on the assumption that monsoon onset is exogenous with respect to household expenditure for all households. Several previous studies have assumed that rainfall is exogenous with respect to household behavior (see, for example, Paxon, 1992; Munshi, 2003; Newhouse, 2005; Jayachandran, 2006; the literature is surveyed in Rosenzweig and Wolpin, 2000). This identification assumption may be threatened if some households are able to anticipate rainfall shocks, but

[19] Arguably, rainfall shocks could influence who the household identifies as the head, but there is no reason why the characteristics of the head would be systematically different for households residing in areas with a delayed onset in the past two years.

this is unlikely to be a serious concern in this context. Systematic dissemination of *El Niño* forecasts to rural farmers began only recently, as a pilot project in 2005.[20] In addition, 1992, 1996, and 1999 – the three years preceding the IFLS survey rounds – were not *El Niño* or *La Niña* years. Finally, the estimated effect of delayed onset is identified using within-year local variation in monsoon onset, which is difficult for models to forecast accurately. Given these factors, it would seem highly unlikely that farmers were able to anticipate variation in monsoon onset.

2.3.3 Descriptive statistics

Approximately 60% of rural households in the IFLS are farm households (see table 2.2).[21] Farm households are slightly more likely to be poor than rich, as the share of farm households is 64 per cent in the bottom expenditure tercile and 57 per cent in the top tercile. Approximately half of the farm households cultivate rice as their main crop.[22]

Table 2.2. Main variables, IFLS data.

	Mean	s.d.[a]
PCE (in logs)[b]	12.089	0.710
Share of farm households	0.596	0.491
Farm profits pc (in rupiahs)[bc]	31,711.8	56,959.3
Farm profits pc, 1st tercile	19,5628.3	28,526.5
Farm profits pc, 2nd tercile	29,174.2	47,762
Farm profits pc, 3rd tercile	48,366.2	81,455.4

Notes: [a] denotes for standard deviation. [b] Household per capita expenditure and farm profits are expressed in December 2000 Jakarta prices. [c] Conditioning that household owns a farm business.

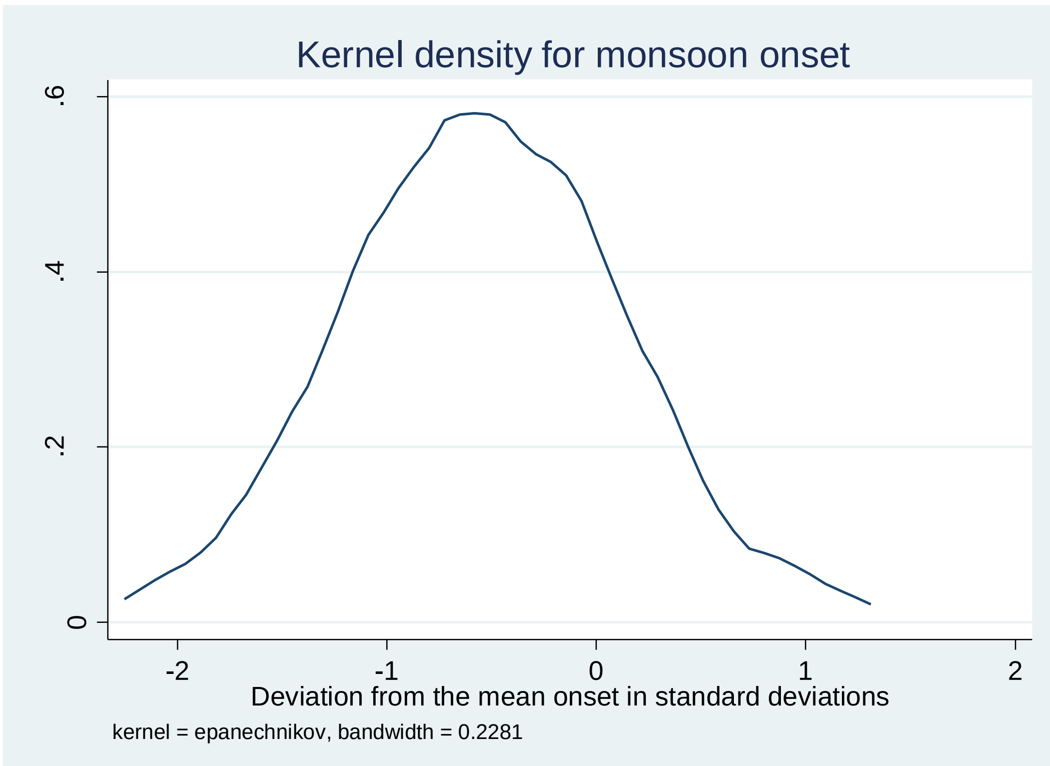
[20] This is in the form of Climate Change Field Schools organized for farmers. See for example: http://www.agrometeorology.org/topics/accounts-of-operational-agrometeorology/climate-field-schools-in-indonesia-coping-with-climate-change-and-beyond. Accessed 10 October 2009. These schools may be beneficial for the farmers: Climate change field schools surveyed in the main rice production kabupatens in West and East Java in 2007-2009 indicated that formal climatic data were used in the timing of farming activities (Natawidjaja et al. 2009).

[21] Households are classified as farm households if at least one member of the household was reported as working on a farm business on household-owned land in 1993.

[22] This is taken from 2000 IFLS data. Unfortunately, farmers were not asked to list their main crops in prior waves.

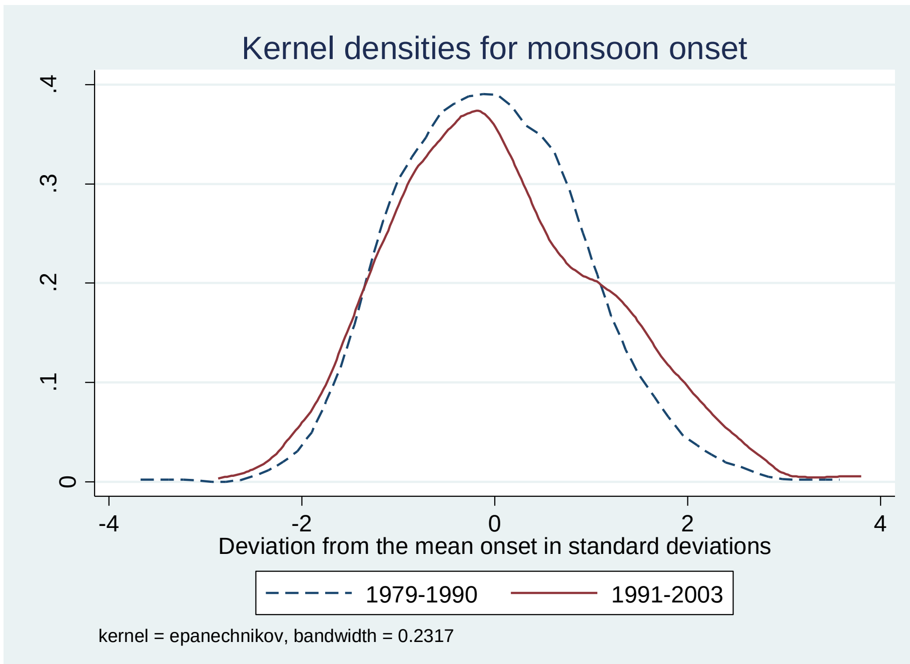
Figure 2.3 presents the density of the monsoon onset prior to IFLS years. None of the years (i.e. 1992, 1996 and 1999) experienced a strong *El Niño* event to an extent that monsoon onset could be delayed as much as nearly four standard deviations (up to three months). Notwithstanding this, monsoon onset shows meaningful variation for the proposed analysis.

Figure 2.3. Density of the monsoon onset prior to the IFLS years.



Key factors motivating this essay are the projected increases in climate variability and extreme weather events. This raises the question of whether there are detectable changes in the distribution of monsoon onset in the rainfall data. Figure 2.4 shows the kernel densities for monsoon onset for two time periods: 1979 to 1990, and 1991 to 2003. The figure shows that the right tail of the distribution expands to a small degree in the later period, suggesting that delayed monsoon onset has become slightly more common. The Kolmogorov-Smirnov test for equality of the distribution function rejects the null hypothesis of equality, with a p-value of 0.008. The choice of period matters, however, as changing the cut-off year to 1991 increases the p-value to 0.129.[23]

Figure 2.4. Kernel densities of monsoon onset for 1979-1990 and 1991-2003.



2.4 Estimation results

This section discusses the estimated effects of onset timing on household expenditure, and for farm households also farm profits using the IFLS data. To facilitate the interpretation and comparability of the results of different models, I focus on reporting the estimated effects of early and late onset. Early onset refers to the monsoon arriving one standard deviation early, relative to the station's historical mean, while late onset refers to a one standard deviation delay. Tables 2.3 and 2.4 present summary estimates of the effect of late

[23] Placing the cut-off point to 1989 or 1992 renders a statistically significant test results, i.e. rejecting the null of equal distributions, but now the two time periods have different amount of years which might affect the results.

and early onset on per capita farm profit and household expenditure for a variety of subgroups. The quadratic regressions are implemented as a robustness check. For the sake of brevity I only present quadratic estimates of per capita expenditure and farm profits for all rural households (tables A5 and A7 in the Appendix), but I have implemented quadratic regressions for poor, middle-class and rich households as well. These robustness checks confirm my main findings.

2.4.1 All rural households

The top row of table 2.3 indicates that delayed onset the previous year reduces the per capita expenditure of all households by a moderate amount, 6.9 percent, and the estimate is not statistically significant (complete results are available in table A4 in Appendix A). The non-parametric estimates, presented in Appendix A, figure A1, show a larger effect. In the non-parametric smoothing the delayed onset reduces per capita expenditure of all rural households by 10 per cent and early onset increases by 5 percent, and both effects are statistically significant at the 5% level.[24] These results are also fairly similar to the quadratic approach (table A5).

[24] The non-parametric estimate does not control for the first lag of monsoon onset.

Table 2.3. Impact of late monsoon onset on rural households' expenditure and farm profits, spline function approach with two lags of monsoon onset, IFLS data.

		Dependent variable	One year ago		Two years ago	
All			Coef.	S.E.	Coef.	S.E.
	All	Log PCE	-0.069	0.056	0.086	0.039**
	Farm	Profit pc ('00 rp)	-8,939	4,890*	-469	2,949
		Log PCE	-0.084	0.067	0.129	0.042***
	Non-farm	Log PCE	-0.049	0.047	0.050	0.036
Poor	Farm	Profit pc ('00 rp)	-7,247	3,735*	2,683	2,509
	All	Log PCE	-0.071	0.044	0.075	0.042*
Middle-Class	Farm	Profit pc ('00 rp)	-13,477	3,633***	560	4,902
	All	Log PCE	-0.151	0.071**	0.050	0.047
Rich	Farm	Profit pc ('00 rp)	-8,273	12,095	-5,573	5,157
	All	Log PCE	0.014	0.085	0.117	0.047**

Notes: Late onset refers to a one-standard deviation delay in monsoon onset, onset defined as the date after August 1st that cumulative rainfall exceeds 20 cm. Estimates are obtained using a spline regression with a single knot at -0.6 standard deviations, including year and household fixed effects. Farm households report owning a farm business in 1993, while household economic class refers to the tercile of the average per capita expenditure during the three surveys.

There is no evidence that these moderate falls in per capita expenditure persist. In fact, the estimates in table 2.3 suggest that late onset two years ago increases per capita expenditure by 8.6 percent. In the non-parametric estimation of expenditure on twice lagged onset, which does not control for lagged onset, the association between delayed onset and per capita expenditure is weak (figure A2 in the Appendix). This set of results, on the whole, rules out the possibility that households experience a substantial reduction in expenditure two years following a delayed onset.

In general, early onset has much weaker estimated impacts. The top row of table 2.4 indicates that early onset previous year reduces expenditure by only 2.5 percent. While the estimated effect of early onset two years ago is large (10.6 percent) and statistically significant, this result is not robust. Finally, non-parametric estimation (figure A2 in Appendix A) also suggests a smaller effect of early onset, no greater than 6 per cent and imprecisely estimated.

Table 2.4. Impact of early monsoon onset on rural households' expenditure and farm profits, spline function approach with two lags of monsoon onset; IFLS data.

		Dependent variable	One year ago		Two years ago	
All			Coef.	S.E.	Coef.	S.E.
	All	Log PCE	-0.025	0.036	0.106	0.049**
	Farm	Profit pc ('00 rp)	-2,289	2,449	-3,027	3,717
		Log PCE	-0.017	0.036	0.057	0.068
	Non-farm	Log PCE	-0.032	0.049	0.177	0.035***
Poor	Farm	Profit pc ('00 rp)	-3,755	3,223	1,354	3,719
	All	Log PCE	-0.075	0.023***	0.086	0.051
Middle-Class	Farm	Profit pc ('00 rp)	-1,747	2,285	1,420	2,280
	All	Log PCE	-0.006	0.031	0.078	0.054
Rich	Farm	Profit pc ('00 rp)	-2,226	4,111	-13,237	9,545
	All	Log PCE	0.007	0.072	0.143	0.059**

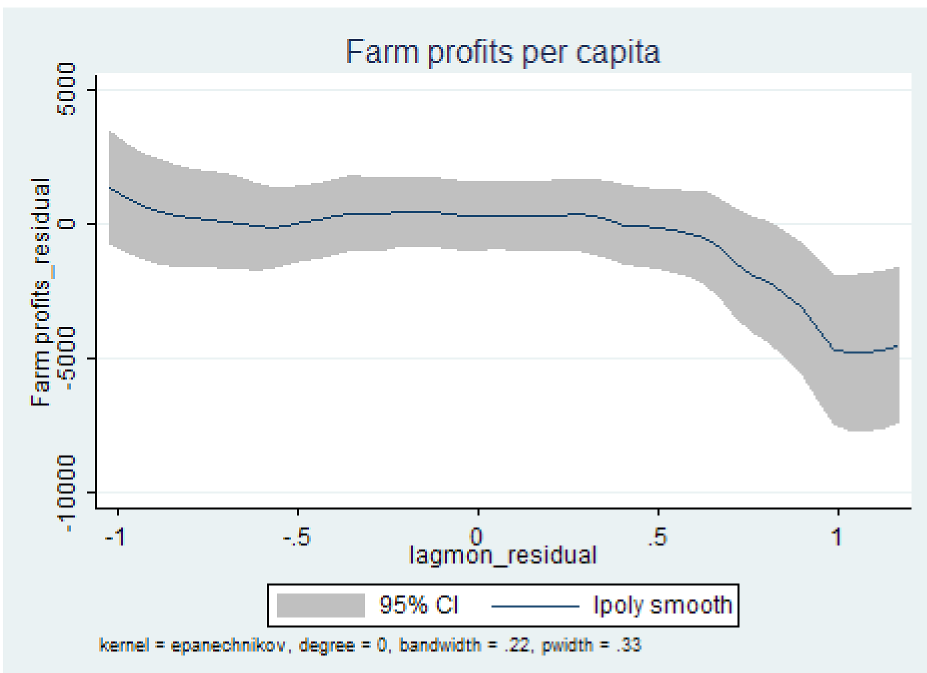
Notes: Early onset refers to a one-standard deviation advance in monsoon onset, onset defined as the date after August 1st that cumulative rainfall exceeds 20 cm. Estimates are obtained using a spline regression with a single knot at -0.6 standard deviations, including year and household fixed effects. Farm households report owning a farm business in 1993, while household economic class refers to the tercile of the average per capita expenditure during the three surveys.

2.4.2 Farm versus non-farm households

Farm households, defined as those that reported owning a farm business in 1993, may be more vulnerable than non-farm households to delayed onset. This could occur, for example, if farm households tend to be dependent on agricultural profits and delayed onset reduces the area harvested. On the other hand, the negative effect of delayed planting may be at least partially offset by higher prices for crops in years when monsoon arrives late. Investigating farm households also allows me to examine the effect of delayed onset on per capita farm profits, in addition to household expenditure. To assess the effect of delayed onset on farm and non-farm households, I re-estimate equation (2.2) for rural farm households, and (2.1) for both farm and non-farm households.

Delayed onset reduces farm profits by a considerable amount. Table 2.3 above indicates that farm profits fall by 8,900 rupiah per month, and the effect is marginally statistically significant at the 10% level. This reduction represents 28% of the average monthly farm profits per capita of all farmers. Non-parametric estimation gives estimated impacts that are slightly smaller, but nevertheless substantial. Figure 2.5 presents the non-parametric estimation of per capita farm profits smoothed against monsoon onset, using local polynomial smoothing and controlling for household and year fixed effects.[25] Late onset decreases per capita farm profits by approximately 5000 rupiah, or 16 per cent of average monthly per capita farm profits, and this is statistically significant at the 5% level.

Figure 2.5. Local polynomial smoothing for per capita farm profits and monsoon onset.



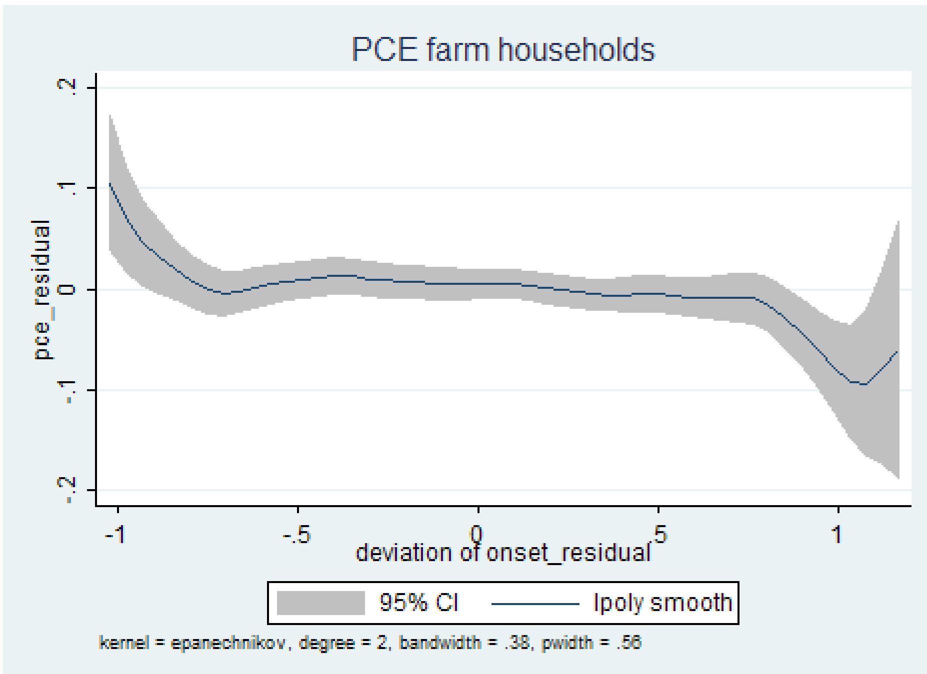
Note: Local polynomial smooth, conditional on household and year fixed effects.

Per capita expenditure, on the other hand, is less responsive to delayed onset for both farm and non-farm households. For farm households, delayed onset in the previous year reduces per capita expenditure by only 8.4 per cent (table 2.3). This is comparable to the estimated impact on non-farmers (4.9 percent). Neither estimate is statistically significant at the

[25] It is noted that the measure of farm profits in the IFLS data is rather crude. The section on profits aims to cover also production for own consumption but it has no disaggregated information on costs and revenues.

conventional levels, however. Figure 2.6 below show the non-parametric estimation of per capita expenditure of rural farmers.

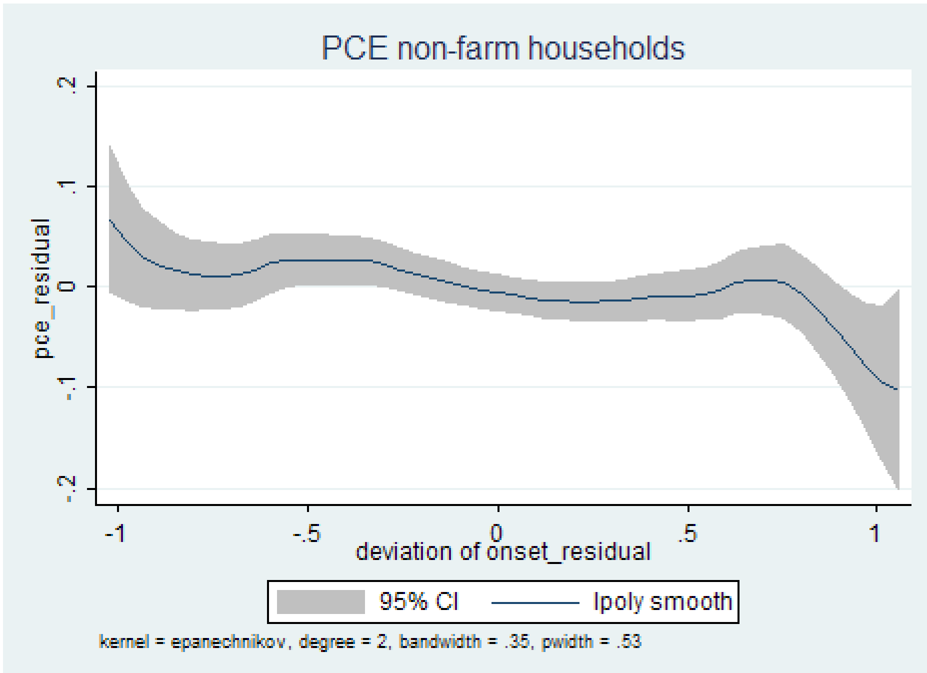
Figure 2.6. Local polynomial smoothing for per capita expenditure and monsoon onset; rural farm households.



Note: Local polynomial smooth, conditional on household and year fixed effects.

Respectively, figure 2.7 below shows the non-parametric estimation for rural non-farmers. Non-parametric estimation confirms minor differences between farmers and non-farmers. Full set of results of the impact of monsoon onset on per capita expenditure of the rural farm and non-farm households are given in table A8 and table A9 in the Appendix.

Figure 2.7. Local polynomial smoothing for per capita expenditure and monsoon onset; rural non-farm households.



Local polynomial smooth, conditional on household and year fixed effects.

Farm households appear to fully recover from late onset within two years, as late onset two years ago has a large (12.9 percent) and positive association with farm households' expenditure. However, delayed onset two years ago is only weakly associated with farm profits, suggesting that any benefit due to delayed onset two years ago may be generated through mechanisms other than increased profits (table 2.3).

Table 2.4 shows that the estimated effects of early onset are small for both farmers and non-farmers. Early onset the previous year is associated with a 3.2 per cent decline in expenditure for non-farmers and a 1.7 decline for farmers. Early onset two years ago is associated with expenditure gains of 5.7 per cent for farmers and a large gain of 17.7 per cent for non-farmers. Given the similarity between the estimated effects of early onset in the prior year, this result appears to be an anomaly and is difficult to explain.

With the exception of early onset two years ago, the estimated effects of variation in onset are broadly similar for farm and non-farm households. This similarity is consistent with price increases in crops and knock-on effects from smaller harvests affecting all households in a community. Farm households may be slightly more vulnerable to delayed onset the previous year, but if anything, benefit more from delayed onset two years ago. Early onset the previous year has small effects on both farm and non-farm households. While the differences between farm and non-farm households are limited, there may be larger differences between poor, middle-class, and wealthier households. The next section examines this important question.

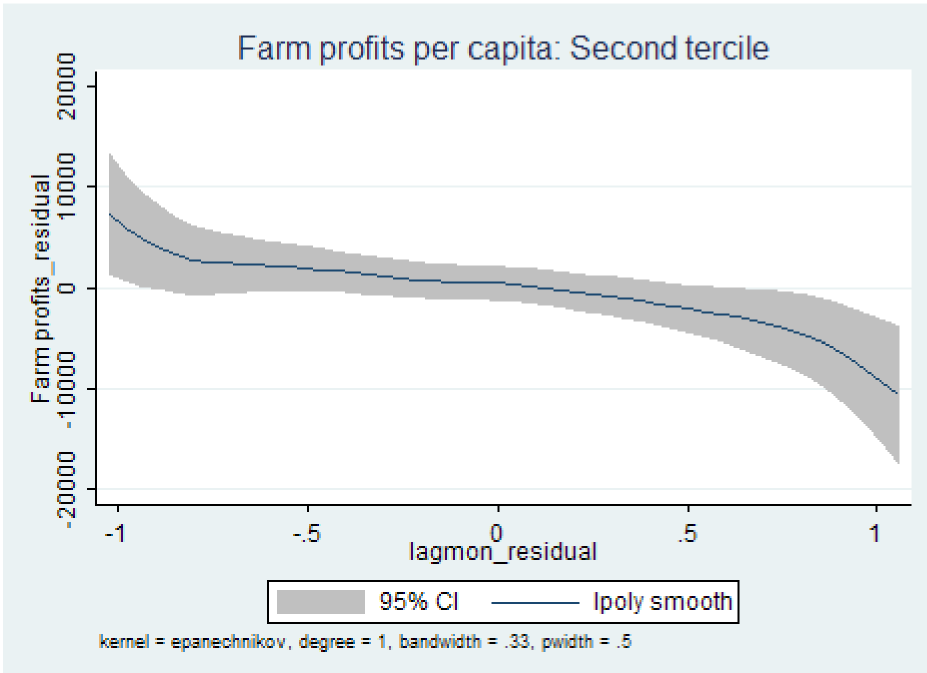
2.4.3 Analysis on distributional implications

As discussed above, theory offers ambiguous predictions with regards to the effect of delayed onset on poor households. The bottom portions of tables 2.3 and 2.4 and shed light on this by presenting estimation results for poor, middle-income and rich households.[26]

The estimates suggest that the effect of delayed onset on farm profits and household expenditure is largest for middle-class households. For these households, late onset reduces farm profits by 13,500 rupiah per month. The reduction in farm profits is considerable, as it represents 44% of the monthly per capita farm profits, or 11% of average total profits, of the average middle-income farmers. The estimate is also robust to non-parametric estimation (see figure 2.8 below).

[26] Full results are available in Appendix A, tables A10, A11, A12, A13, A14 and A15.

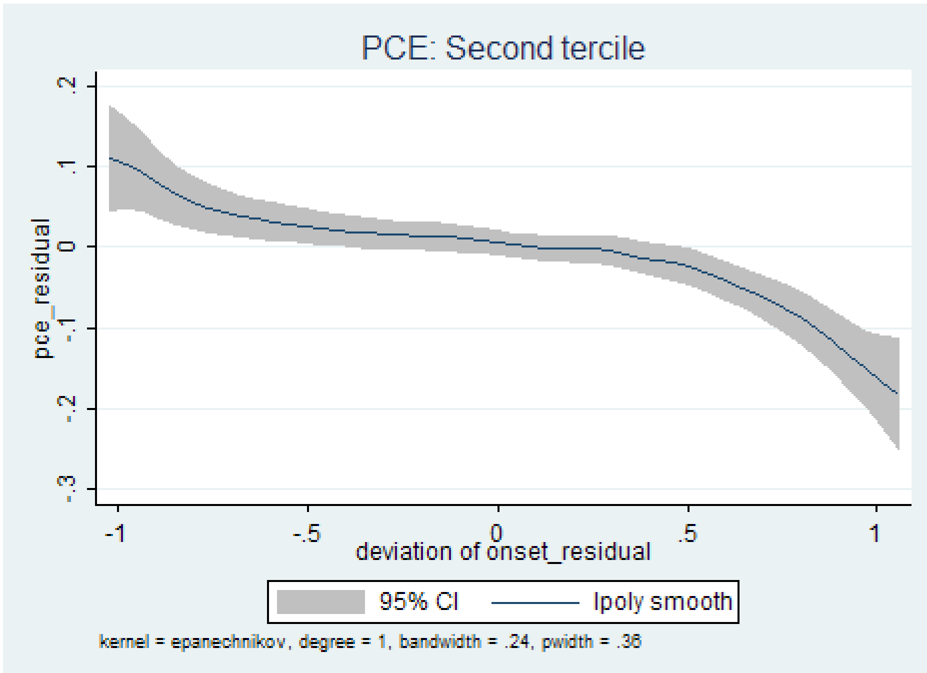
Figure 2.8. Local polynomial smoothing for farm profits per capita and monsoon onset; households in the second expenditure tercile.



Note: Local polynomial smooth, conditional on household and year fixed effects.

The reduction in per capita expenditure of the middle-income households reflects the reductions in per capita farm profits. For these households delayed monsoon onset reduces per capita expenditure by 15.1 percent. The estimate is similar and statistically significant in the non-parametric estimation as well (see figure 2.9 below).

Figure 2.9. Local polynomial smoothing for per capita expenditure and monsoon onset; households in the second expenditure tercile.



Local polynomial smooth, conditional on household and year fixed effects.

Poor farmers, on the other hand, suffer less than middle-income farmers following late onset. Late onset reduces farm profits by 7, 200 rupiahs for the poorest farmers, and expenditure by 7.1 per cent for poorest households. Moreover, only the estimated effect of farm profits is marginally statistically significant at the 10% level.[27]

Finally, rich households are the least vulnerable to late onset. The estimates in table 2.3 above suggest that late onset reduces per capita profits for the wealthiest farmers by only 8,300 rupiah per month, which is nearly the same in absolute terms as the poorest farmers and not statistically significant. Furthermore, the impact of delayed onset on per capita expenditure of the wealthiest households is negligible.

Early onset, unlike late onset, appears to effect poor households most. Table 2.4 shows that early onset reduces per capita profits by nearly 3,800 rupiah per month, and per capita expenditure by 7.5 percent. The estimated effect on expenditure is statistically significant at

[27] The estimated effect is statistically significant at the 5% level when only the first lag of monsoon onset is included in the regression.

the 1% level. Middle-class and wealthy households are much better protected from early onset. For middle-class households, early onset is associated with a small (1,750 rupiah) reduction in farm profits and a negligible reduction in per capita expenditure, and estimated magnitudes are similar for the top tercile of households. Early onset is associated with abnormally high rainfall, which could lead to minor flooding that reduces production for poor households.

As noted above in section 2.3.2, the division of households into terciles based on their average per capita expenditure could bias the estimates. As a robustness check, I estimate the portion of household per capita expenditure that is predetermined to the rainfall shocks. The results when using this alternative classification are broadly similar to the original results. In terms of household per capita expenditure, middle-income households remain most vulnerable to delayed onset. A delayed monsoon onset in the previous year reduces the per capita expenditure of these households by 17.8 per cent and the effect is statistically significant at the 5% level. For poor households, the effect is close to zero and not statistically significant. Middle-income households also suffer the most both in absolute and relative terms when looking farm profits per capita. Farm profits of the middle-income farmers are reduced by 13,800 rupiah following a delayed monsoon onset, compared to 10,700 rupiahs of the poor farmers. However, the estimated effect of delayed onset on the profits of middle-income farmers is only marginally statistically significant.

An alternative method to address the possible endogeneity of the household economic classification is to use wealth instead of household expenditure. To the extent that wealth is less responsible to rainfall shocks than expenditure, this would mitigate the bias. Classifying households based on wealth instead of expenditure does not change the primary finding that households in the middle of the per capita wealth distribution face the greatest loss in terms of farm profits and per capita expenditure after late onset, providing suggestive evidence that the magnitude of any bias of this nature is low. The reductions when using household assets as wealth indicator are of the same order of magnitude as when using household expenditure. The reduction in farm profits of the middle-income farmers became 15,000 rupiahs (instead of 13,500) and the reduction in per capita expenditure remained 15.1 percent. However, the results that poor households are most vulnerable to early onset, as shown in table 2.4 is not confirmed when using average assets rather than average expenditure as the main indicator of household wealth.

2.4.4 Household size and total expenditure

The timing of the monsoon's arrival can affect per capita expenditure either through the numerator, total expenditure, or through the denominator, i.e. household size. Weather shocks may increase household size by encouraging households to combine and by discouraging the formation of new households. Shocks, however, may also decrease household size by encouraging out-migration from large households (Henry et al., 2004). During the financial crisis in the late 1990s, average household size increased both in rural and urban areas (Thomas et al., 2004). This suggests that increases in household size may be an important coping mechanism during financial distress among Indonesian households.

To better understand the potential role of household size, I decompose the effect of delayed onset into its effect on household size and its effect on total household expenditure, by re-estimating equation (2.1) with log household size and log total expenditure as dependent variables. Table A16 in the Appendix presents the results for household size, for all households and each expenditure tercile, and table A17 shows comparable results for total expenditure. In each table, the bottom portion of the third column presents the estimated impacts of delayed onset for middle-income households. Delayed onset increases household size by roughly 7.4 per cent for these households, and the estimate is statistically significant at 1% level (table A16). Delayed onset in the prior year has a slightly larger negative effect on household expenditure, roughly 7.7 percent, although the estimate is not statistically significant (table A17). Therefore, in this specification, approximately 50 per cent of the total effect of delayed onset on per capita expenditure is manifested through an increase in household size.

2.4.5 Intensity of rainfall versus timing of rainfall

This section turns to the secondary question of whether the results are sensitive to the use of monsoon timing rather than rainfall intensity as the main indicator of climactic variability. To examine this, I re-estimate equation (2.1) using rainfall intensity instead of monsoon onset. The measure of rainfall intensity is the standardized deviation from the annual historical mean precipitation and the knot is located at -0.6 standard deviations as before. Annual means are taken from August to July, which coincides with the crop season.

Table 2.5 below presents the estimated effects of early and late onset from a model using rainfall intensity. For comparison purposes, the results from the monsoon onset model displayed in tables 2.3 and 2.4 are redisplayed here (full results are given in tables A18 and A19). The most striking result is that the relatively large effect of delayed onset on middle-income households also holds for rainfall intensity. For these households the estimated impact of a one standard deviation delay in onset the prior year is virtually identical to the estimate for a one standard deviation shortfall in rainfall amount. Both measures indicate a positive effect of delayed onset and low rainfall two years ago. However, other results for monsoon onset, such as the moderate negative effect of early onset on poor households, are not apparent when using rainfall intensity.

Table 2.5. Estimated effects of onset and intensity variables on per capita expenditure from separate regressions, IFLS data.

	All households	Poor households	Middle-class	Rich households
Prior year				
Early onset	-0.025	-0.075***	-0.006	0.007
High rainfall	-0.011	-0.001	-0.003	-0.010
Late onset	-0.069	-0.071	-0.151**	0.014
Low rainfall	-0.098*	0.010	-0.149**	-0.108
Two years ago				
Early onset	0.106**	0.086	0.078	0.143**
High rainfall	0.041	0.032	0.055	0.030
Late onset	0.086**	0.075*	0.050	0.117**
Low rainfall	0.025	-0.030	0.122	-0.020

While confirming that both late onset and below-average rainfall harm expenditure of the middle-class, these results provide limited indication about which measure is more strongly associated with expenditure. To address this question, I re-estimate a version of equation (2.1) that includes both monsoon onset and rainfall intensity as independent variables. The results are presented in table 2.6 below. They show that, in general, rainfall intensity has smaller impacts than monsoon onset, particularly in the prior year. For example, the estimated impacts of low rainfall the prior year are around half as large as the estimated impact of late onset, for all households as well as middle-class households. The pattern holds for farm profits as well: after controlling for the timing of rainfall, the intensity of rainfall has little explanatory power.[28]

Table 2.6. Estimated effects of onset and intensity variables (same regression), IFLS data.

	All households	Poor households	Middle-class	Rich households
Prior year				
Early onset	-0.068*	-0.109***	-0.080**	-0.004
High rainfall	0.005	0.026	0.006	-0.002
Late onset	-0.077	-0.086	-0.146*	0.015
Low rainfall	-0.036	0.072	-0.074	-0.063
Two years ago				
Early onset	0.060	0.048	0.017	0.117
High rainfall	0.042	0.039	0.061**	0.019
Late onset	0.079*	0.081	0.040	0.106*
Low rainfall	0.064	0.044	0.131*	-0.001

2.4.6 Monsoon onset and local rice prices

Similarly to farm households suffering lower profits the year following delayed onset, onset timing affects the expenditure of non-farm households as well. This suggests that rainfall patterns could affect household welfare partly by influencing local rice prices (Falcon et al., 2004). Since rice is the largest element in the consumer price index, and 72 per cent of the

[28] The rainfall measure used in this study, annual average rainfall, is fairly crude, compared to more sophisticated measures taking the account to rainfall in different seasons. Exploring these measures is left for future research.

rural population are net consumers of rice, increases in the relative price of rice tend to have widespread effects on household welfare.[29]

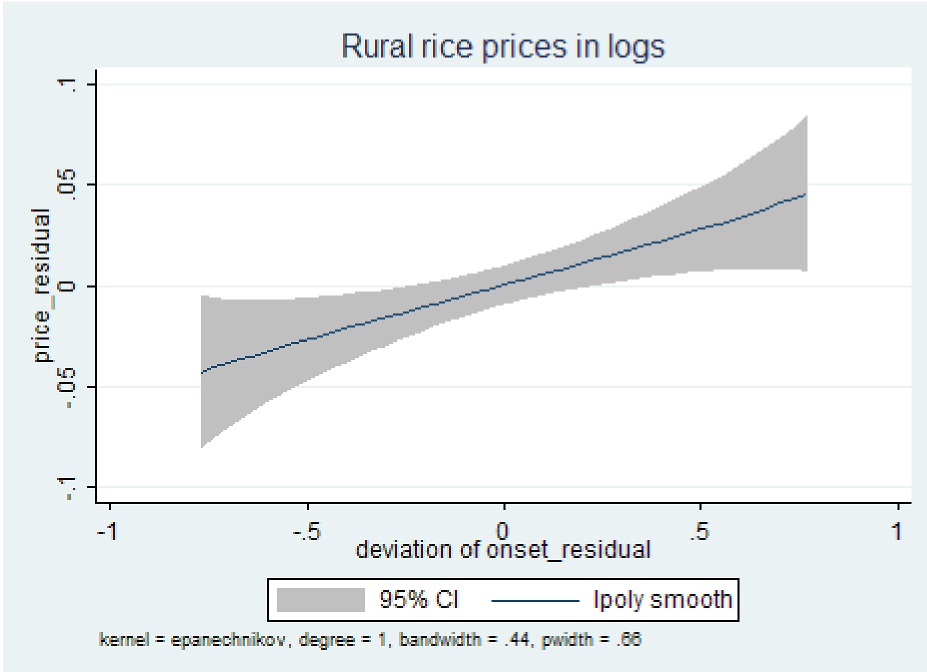
The measure of local rice prices comes from price information provided in the IFLS community questionnaire. In each village the interviewer visited three markets and marked down the price (per kg) of average quality rice. Since the information provided in IFLS1 is not comparable to that in IFLS2 and IFLS3, I only use information provided in the latter two waves. My preferred specification includes community and year fixed effects, which is equivalent to regressing change in local rice prices on change in onset timing.[30] To test for robustness, I also estimate a pooled cross section regression of onset on rice prices.

The estimates indicate that late onset increases rice prices by 6.2 per cent in my preferred specification with community fixed effects, and 3.5 per cent in the pooled cross-section (see tables A20 and A21). These results are also confirmed by non-parametric estimation (see figure 2.10). The figure shows a linear relationship between onset timing and the average consumer price of rice, as late-arriving rains raises the price of rice while early arriving rains has an opposite effect.

[29] BPS does not disclose rice's share in the consumer price index, but raw and processed foods make up 36 per cent of the index. http://www.thejakartaglobe.com/business/indonesia-inflation-hits-seven-month-high-due-to-rising-food-prices/356040. Accessed 21 June 2010

[30] Monsoon onset is in linear form, instead of using the splines, in the rice price specifications.

Figure 2.10. The effect of monsoon onset on rural rice prices, years 1997 and 2000.



Local polynomial smooth, conditional on community and year fixed effects.

Data on spending patterns confirm that while poor households are those most directly vulnerable, the direct welfare effect of rice price increases due to delayed onset is mild. Table 2.7 shows that the purchased rice accounts for 14.8 per cent of the total expenditure in the bottom tercile. Therefore, the estimated six per cent increase in rice prices due to a one standard deviation delay in onset only reduces welfare by approximately one percent.

Table 2.7. Share of rice in the household budget by terciles, IFLS data.

	Share of rice (both self-grown and purchased) in total expenditure	Share of self-grown rice in total expenditure	Share of purchased rice in total expenditure
1st TERCILE	20.3	5.5	14.8
2nd TERCILE	15.8	4.5	11.3
3rd TERCILE	11.3	3.4	7.9

2.5 SUSENAS Data

This section turns to the impact analysis of monsoon onset on expenditure using repeated cross-sections from the national socioeconomic survey. Each year, in late February or early March, a new set of roughly 190,000 households are interviewed as part of the core of the national socio-economic census (SUSENAS). The dataset includes results from a small consumption module, consisting of 15 food items and 8 non-food items, that combines purchased and self-consumption. To utilize these data, a five per cent sample of all rural households is randomly drawn from 1993 to 2004, stratified by district, year and per capita expenditure quintile. These 12 datasets were then combined into a single dataset consisting of 118,500 households. In the SUSENAS data households were matched to rainfall station closest to their district.

The SUSENAS core data has both advantages and disadvantages as a data source, compared to the IFLS. The main advantage is that annual data are available covering 1993 through 2004 rather than only 1993, 1997, and 2000. In addition, since SUSENAS covers the entire country, 49 weather stations can be matched per year. The IFLS, in contrast, only covers 13 provinces and 36 matched weather stations. However, the methodology used for SUSENAS does not imply re-interviewing the same households, which means that the estimated effect of rainfall on expenditure will be less precise. In addition, the SUSENAS consumption module covers relatively few goods and may not be sufficiently detailed to accurately detect variation in household consumption. Further, SUSENAS core interviews are implemented in February, implying that the effects of early or delayed onset might not yet have realized in household behavior.

2.5.1 Methodology and results

To assess the effect of delayed onset on expenditure using the SUSENAS core data, I estimated the following equation:

$$\ln C_{ijt} = \alpha + \phi_1 spline_{1,t-1} + \phi_2 spiline_{2,t-1} + \phi_3 spline_{1,t-2} + \phi_4 spline_{2,t-2} + \gamma Z_{ijt} + D_j + D_t + \varepsilon_{ijt}. \quad (2.4)$$

In equation (2.4), as before, C_{ijt} represents the per capita expenditure of household i located in province j in year t and splines represents the linear splines of lagged monsoon onset measured in standard deviations as explained earlier. Z contains a vector of household characteristics that are assumed to be unaffected by rainfall shock (gender, age and education of the household head). D_j is a set of province dummy variables and D_t is a set of year dummies. For the sake of brevity, I only present results of spline function approach for SUSENAS. Unfortunately SUSENAS data does not contain any information on farm profits, and therefore only per capita expenditure is included as a dependent variable. It is also not possible to study the heterogeneity among poor and rich households due to the cross sectional nature of the data.

The results are shown in table A22 in Appendix. The timing of monsoon onset has little impact on household per capita expenditure using the SUSENAS data. The point estimates for delayed onset previous year are close to zero, and the 95 per cent confidence interval can rule out reductions greater than one percent.[31]

2.5.2 Comparing the SUSENAS and IFLS results

In the IFLS data, delayed onset the previous year has a moderately negative effect on per capita expenditure for all rural households, though the estimates are only statistically significant in some specifications. Further, delayed onset two years ago has a positive relationship with per capita expenditure. In the SUSENAS specifications, neither early nor delayed onset has a meaningful effect on per capita expenditure in any year. There are five

[31] Similar results were obtained using the quadratic approach.

main differences between the two sources of data that could account for the greater variability in the IFLS results.

The first and most apparent potential factor is that the surveys cover different years and provinces. SUSENAS is a nationally representative survey, although rainfall data is only available for 22 of Indonesia's 32 provinces.[32] The IFLS data covers only a subset of the area and years covered by the SUSENAS. To test the importance of this limited coverage, I re-estimated the effect of variation in onset in the SUSENAS, but limited the sample to only the provinces and years covered by the IFLS. Table A23 in the Appendix presents the results. When limiting the analysis to the IFLS years and provinces, the results change only little and the overall conclusion that variation in monsoon onset has no meaningful impact on expenditure in the SUSENAS data holds. Therefore, I conclude that the estimated effects using the IFLS data are unlikely to be the result of a subsample of years and provinces covered by the IFLS.

Another factor that might affect the results obtained in the IFLS data is the fact that the construction of the expenditure aggregate in the IFLS1 differs from that in the later waves due to changes in the questionnaire. However, these changes are rather small and mostly concern the wording on whether the reported non-food expenditure amounts are for the entire household or not. Nonetheless I test whether these changes affect the results by excluding the households in the 1993 that stated that the reported amount is not for the entire household or that they don't know whether it is. The proportion of households reporting that the amount of one or more non-food item (total of 13) is not for the entire household is high, 50%. However, excluding these households from the econometric analysis changes the point estimates only little. For example, delayed onset reduces now the per capita expenditure of the middle-income households by 14.6 per cent (previously 15.1 percent). The results indicate that restricting the sample mostly affects the standard errors: the second spline is now statistically significant at the 10% level but previously at the 5% level. This is confirmed by randomly excluding 50% of the households in 1993.[33]

---

[32] Provinces were omitted from the analysis if they did not contain a weather station; these include North Sulawesi, Central Sulawesi, South-East Sulawesi, Maluku, Maluku Utara, West Irian Jaya and Papua.

[33] I repeated the exercise for several times. By bootstrapping the standard errors I get very similar point estimates and significance level for the unrestricted and restricted sample.

Therefore, I conclude that the estimated effects using the IFLS data are unlikely be driven by changes in the questionnaires.

There are three other sets of potential explanations for the discrepancy between the two datasets. However, in contrast to differences in coverage, these are not easily testable. The first is differences in the timing of the survey. The monsoon rains typically arrive in September-November, depending on the location; crops are planted in the beginning of the rainy season, and harvested in March. SUSENAS is fielded in late February or early March, raising the possibility that households are short-sighted and only react to delayed onset by cutting back on their expenditure after the March harvest. It is also possible that households are better insured six months after late onset than they are a full year later.

The second possibility is that differences between the two datasets reflect the non-representative nature of the IFLS sample, especially as the panel ages. In particular, the IFLS makes no effort to re-contact young persons (born after 1968) that exited households after 1993, unless they were the two children of the head in 1993 that were selected as “main respondents”. In addition, not all new split-off households are successfully re-contacted and re-interviewed. As a result, successive rounds of the IFLS increasingly under-represent households with young heads.

The final set of explanations revolves around differences between the two datasets in how expenditure is ascertained and deflated across time and space. In this respect, the IFLS expenditure is likely to be more accurate, for three reasons. First, the IFLS consumption questionnaire is far more detailed than the SUSENAS core. It asks households to report consumption on 37 food items and 19 non-food items, whereas the SUSENAS core questionnaire only asks about 15 food items and 8 non-food items. Second, the IFLS questionnaire asks separately about purchased food, and food that was given or self-produced. In contrast, the SUSENAS asks households to estimate the total amount that was purchased, self-produced, or given. Third, the IFLS likely uses more accurate deflators. In 1997 and 2000, rural price deflators were constructed for each province, using data from Rural Consumer Price Survey, whose respondents were farmers or farm workers. For the first IFLS round, as well as all SUSENAS rounds, consumption was deflated using price data published by the Central Bureau of Statistics (BPS). Unfortunately, price data is only published for major cities, meaning that the SUSENAS data is deflated by the capital city of the province, and may not accurately reflect changes in rural prices as a result of delayed onset.

In summary, IFLS data has more accurate information on household consumption and better deflators. In addition, it has information on farm profits and the panel dimension allows me to study heterogeneity among poor and rich households. On balance, IFLS data is the preferred data in the given context and therefore the IFLS results are considered as the leading results of this chapter.

2.6 Conclusion

The major contribution of this chapter was to present new empirical evidence on whether delays in monsoon onset reduce households' farm profits and expenditure in rural Indonesia. Previous research suggests that delayed onset reduces aggregate rice production by reducing the area planted for rice. How households respond to these shocks, is, however, largely unknown. Evidence from other countries suggests that households' ability to maintain consumption in the face of shocks varies from country to country, but this is the first study to examine this question in Indonesia, and to my knowledge, the first to examine how household vulnerability to climate variation depends on households' economic status in the context of a developing country.

Using the IFLS data the results indicate that delayed monsoon onset reduces farm profits per capita by a moderate amount, and reductions are most significant for middle-income farmers. For these farmers, delayed onset reduces monthly per capita farm profits by 13,500 rupiahs (around half of their monthly per capita farm profits) and expenditure by roughly 15 percent. For poor farmers the estimated reduction in farm profits is smaller (7,200 rupiahs) and only marginally statistically significant. Rich farmers are relatively well insured against variation in monsoon onset, which may stem from their ability to self-insure against rainfall shocks and their access to irrigated farmland, which is heavily skewed towards large landowners.[34] Meanwhile, the more moderate impacts faced by poor farmers are consistent with the use of lower-risk and lower-return crops and farming strategies (Rosenzweig and Binswanger, 1993; Zimmerman and Carter, 2003).

Approximately half of the reduction in the household per capita expenditure is realized through increases in household size. Earlier studies have shown that Indonesian households increased the household size during the financial crisis in the late 1990s in order to benefit from economies of scale (Thomas et al., 2004). This essay suggests that adjustment in household size is an important coping mechanism also when household experiences a delayed monsoon onset.

Monsoon onset, however, appears to have no lasting impact on farm profits and expenditures. In particular, there is no empirical evidence indicating that delayed onset two years ago has a substantial negative effect and, in fact, in some specifications it is positively associated with per capita expenditure. Thus, in light of these estimates, I conclude that households do not experience a substantial reduction in expenditure two years following late monsoon onset.

The findings suggest that the timing of rainfall affects households' behaviour more than the amount of rainfall. After controlling for the timing of monsoon onset, the intensity of the annual precipitation have negligible impact on household per capita expenditure. Finally, I also find that delayed monsoon onset increases local market price of rice in rural areas: a one standard deviation delay in onset increases the price of rice by 6.2 per cent and the relationship between onset and local rice prices is linear. The increase in the price of rice disproportionately harms the poor, but the effect of the price increase may be diluted, given that purchased rice only accounts for about 15 per cent of their budget.

The results are robust to various sensitivity checks, including different functional forms, non-parametric estimation, and the division of household into wealth terciles according to the consumption predetermined to the rainfall shocks and according to their assets instead of expenditure. However, the SUSENAS data does not confirm the adverse effect of

---

[34] Descriptive analysis, however, suggests that differences in access to irrigation are moderate, and cannot explain the greater vulnerability of middle-income households. Around 55% of the poor farmers live in communities where no technical irrigation is available, while 47% of middle-income and 44% of rich farmers.

delayed monsoon onset on household per capita expenditure. Using the SUSENAS data I find that monsoon onset, neither early nor delayed, has virtually any effect on per capita expenditure. Even though it is not possible to test all the possible factors behind the discrepancies, different timing of the surveys, as well as different ways to measure expenditure and price changes, might have a role to play. Unfortunately, the SUSENAS survey does not contain information on farm profits. I have discussed the strengths and deficiencies of the two datasets in the current application, and as a result I have taken the IFLS estimates as my main and concluding results.

An important limitation of this study is the amount and quality of the rainfall data. Only 36 rain stations were matched to each wave of the IFLS, which reduces the variation in the main independent variable. The years prior to the IFLS interviews provide only moderate variation in monsoon onset, and none of the years captured by the IFLS follow a strong *El Niño* year. This problem is overcome in the SUSENAS analysis, however. The measure of late onset is based on daily rainfall data, which is undoubtedly measured with some error. Since this measurement error is independent of local household characteristics, however, if anything, my estimates under rather than overstate the impact of delayed onset. Future work could investigate the accuracy of rainfall data, as well as alternative definitions of monsoon onset and rainfall shocks.[35] Despite these data limitations, the analysis demonstrates that households in the second tercile face a substantial adverse impact from delayed onset.

Further research could shed light on how community characteristics affect households' ability to cope with weather shocks. In particular, households with access to irrigation and access to credit may suffer less from climatic volatility. Finally, formal weather-based insurance mechanisms could also mitigate loss from climate variation. Households in the middle tercile of the expenditure distribution in rural Indonesia are not wealthy by international standards, and the results of this study suggest that they could benefit from further intervention.

[35] These could include definitions that take into account false starts, i.e. threshold is followed by a dry spell (see for example Moron et al., 2009). However, onset dates presented in Moron et al. (2009) are in good agreement with Naylor et al. (2007a).

Chapter 3 The effect of weather shocks and risk on schooling and child labour in rural Indonesia

3.1 Introduction

The objective of this chapter is to study the impact of weather shocks on schooling and child labour in rural Indonesia. The weather shock is measured as the deviation of monsoon onset, the start of the rainy season, from its historical mean date. In addition, this chapter studies whether ex-ante risk affects parents' decision to send their children to school. The weather-related risk is measured as the coefficient of variation of monsoon onset.

Weather variables have been commonly used in the literature as a means of identifying the effects of permanent and transitory components of income. However, despite the many advantages of this method, mainly the strong correlation between weather and farm income and the randomness of the weather events, the models have been based on theoretical frameworks with somewhat strong assumptions about the operation of rural labour markets, preferences and technology (Rosenzweig and Wolpin, 2000). Accordingly, some recent papers have used weather variables to study the direct relationships between income risk and income shocks and the outcome of interest using reduced form specifications (Kochar, 1999 and Rose, 2001 in Rosenzweig and Wolpin, 2000). Further, Rose (2001) argues that the direct method enables her to eliminate the possible endogeneity problem related to weather shock and production decision.[1] This study builds on this more recent body of literature and examines the effect of weather shocks and weather risk on schooling and child labour. It is noted that I am not able to distinguish the mechanism through which the effects materialise. Nevertheless, the objective of this essay is to document the impacts of past weather shocks and weather risk and therefore the mechanisms are of secondary interest.

[1] For example, Walker and Ryan (1990) find that farmers commonly increase the acreage of drought-resistant crops relative to that of water intensive crops if their expectations of the rainfall conditions are poor.

The research questions addressed in this chapter are important from a policy perspective for a number of reasons. The expected changes in climate patterns represent a serious threat to agricultural productivity in developing countries, which undoubtedly affect livelihoods and incomes of rural population (see for example Cline, 2007; Easterling et al., 2007). Rice farming in Indonesia is greatly affected by the variation in the timing of the rainy season (monsoon) as *El Niño* events can delay rice planting by up to two months, reducing the area harvested and often driving up domestic and international rice prices (Falcon et al., 2004; Naylor et al., 2007a). Further, the current consensus predicts that the Asian monsoon will intensify in the future with climate warming, implying more and longer droughts in Indonesia (Overpeck and Cole, 2007). Agriculture continues to be an important source of livelihood while 60% of the work force in rural Indonesia engages in agriculture. Except for the well-documented relationship between monsoon onset and rice production (se, for example, Naylor et al., 2001; Naylor et al., 2007a) the socio-economic implications of climate variability in Indonesia are relatively unknown. This study seeks to fill this gap by studying the effect of climate variability on schooling and child labour.

Most of the previous literature has focused on ex post effect of a shock on education outcomes (see, for example, Kruger, 2007 and Maccini and Yang, 2009) while one of the contributions of this chapter is to enhance understanding on how schooling is used as a measure to cope with risk, that is the realization of a shock ex ante. [2] Children's schooling in risky environments might be adversely affected by households' need to build-up buffer stocks to cope with future shocks.

Schooling and other investments in human capital play an important role in escaping poverty; yet there are many factors that may interrupt schooling or prevent children from starting school. A broad body of literature has examined the interaction between exogenous shocks, such as unemployment, illness, crop loss, income loss, and investments in children (see, for example, Jacoby and Skoufias, 1997; Jensen, 2000; Thomas et al., 2004; Kruger, 2007). These questions are of particular relevance in developing countries where missing and/or imperfect credit markets may hinder investments in human capital. Along with human capital investments economists have been interested in the determinants of child

[2] On the effect of risk on schooling, see for example Fitzsimons (2007) and Kazianga (2005).

labour and the impacts of exogenous shocks on child labour (see, for example, Kruger, 2007; Beegle et al., 2008; Yang, 2008).

My prior hypothesis is that early onset has either a positive or neutral impact on school attendance whereas delayed onset has a negative effect. This is based on previous research which shows that delayed onset decreases the amount of rice harvested in the following calendar year (Falcon et al., 2004).[3] The reduced harvest affects households' farm profits, which in turn might have implications on households' investment in human capital, especially in rural settings where credit is likely to be scarce. However, both good harvest and bad harvest (delayed onset) could increase the demand for child labour because the child wage has both substitution and income effects. The increase in the child wage rate increases the demand for child labour because of the substitution effect, while the income effect has a negative sign. Therefore, the total effect depends on the relative strength of these two factors. Indeed, Kruger (2007) finds that coffee boom raises child labour in Brazil, whereas Beegle et al., (2008) argue that self-reported crop loss lead to increased hours worked by children in Tanzania.

This study contributes to the literature on households' coping mechanisms when facing an exogenous shock, with a particular emphasis on weather shocks and weather-related risk. Children's schooling is at risk when household faces a rainfall shock affecting its income. Grimm (2008) analyses the impact of food price inflation on children's schooling in Burkina Faso. The findings suggest that a loss in purchasing power had a negative effect on enrolment rates. Jensen (2000), using data on Côte d'Ivoire, finds that enrolment rates for children aged 7-15 declined by 14 and 11 percentage points among boys and girls, respectively, in areas that experienced adverse weather conditions, and actually increased at the same time in all other areas. Jacoby and Skoufias (1997), using the ICRISAT data on rural India, find that child labour and school attendance play a significant role in the self-insurance strategy for poor households. Björkman (2006) finds that negative rainfall, and thus income shocks, reduces female enrolment in primary school in Uganda. In respect to weather risk, Rose (2001) finds that ex ante risk, measured as the coefficient of variation of rainfall, increases the probability of a household participating in the labour market in rural

[3] Also in Chapter 2 of this thesis I find that delayed onset has an adverse effect of expenditure and farm profits of the middle-income households.

India.[4] In addition the author finds that also unexpected bad weather increases labour force participation.

Past weather shocks have adversely affected also Indonesian households. Self-reported crop loss is associated with reduced education expenditure (Cameron and Worswick, 2001). Aggregate village-level risk, measured as past rainfall variability, was found to have reduced educational attainment of rural children (Fitzsimons, 2007). Finally, early-life drought between 1953 and 1973 adversely affects health, educational attainment, espousing quality, and adult socioeconomic status in rural Indonesia (Maccini and Yang, 2009).

The overall effect of child labour on individual welfare is ambiguous in theory. Child labour may itself be harmful for child's education and health, and these adverse effects might be lasting (see for example O'Donnell et al., 2005; Beegle et al., 2008). On the other hand, child might gain essential work experience that could be rewarded in the labour market (see for example Beegle et al., 2005). Nonetheless, several empirical studies have revealed negative consequences of child labour (Kruger, 2007; Beegle et al., 2008).[5] However, it is important to make a distinction between different types of child labour. Studies on India demonstrate that child labour in rural areas is often 'light' in a sense that children ought to be able to educate themselves and work, provided that schools were available. However, the story is very different for organized child labour (Basu, 1999).[6]

Using three rounds of the Indonesia Family Life Survey, IFLS data, I find that delayed monsoon onset has an increasing impact on the incidence of child labour. A one standard deviation delay in monsoon onset in the previous year increases the probability of a child working by 5.8 percentage points using data for 2000 only and 9.5 percentage points in the course of the three surveys. With respect to education, I find that delayed monsoon onset only in particular years is harmful for school attendance: delayed onset in the transition

[4] The dependent variable in the main analysis is a dummy variable indicating whether a member of the household participated in the labour market.
[5] It is notable that Beegle et al. (2005) find that the loss in education attainment due to child labour is offset by increased earnings from wage and farm work. However, the authors argue the lasting effect of the reduced education may only realize in the long term when the return to education increase and return to work experience decrease.
[6] Basu and Van (1998) argue that in the multiple equilibrium parents choose to send their children to work when additional income is needed (bad equilibrium), but refuse to do so when adult wages are sufficiently high.

year from primary to secondary school reduces the probability of attending school in following years by 2.8 percentage points. Finally, young children aged 6-10 years are less likely to enter primary school in riskier environments.

The remainder of the chapter is structured as follows. Section 3.2 provides a short introduction on rice farming in Indonesia and overviews of the education system and child labour in Indonesia. Section 3.3 introduces the data, and section 3.4 describes the empirical approach. Results are presented and discussed in sections 3.5 and 3.6. The final section concludes, discusses some policy implications and outlines areas for future research.

3.2 Background

3.2.1 Farming and rice production in Indonesia

In Indonesia, rainfall patterns vary greatly within a year across districts as well as within districts over time. As the country is located close to the equator, the variation in temperature is very small, both within years and across them, implying that rainfall patterns are the most important dimension of the weather variation. The climate in Indonesia consists simply of one wet season (October-May) and one dry season (June-September) each year.[7] In the 20 years before 2004, a 30-day delay monsoon onset occurred nearly 18 per cent of the time in West/Central Java and 10 per cent in East Java/Bali (Naylor et al., 2007a).

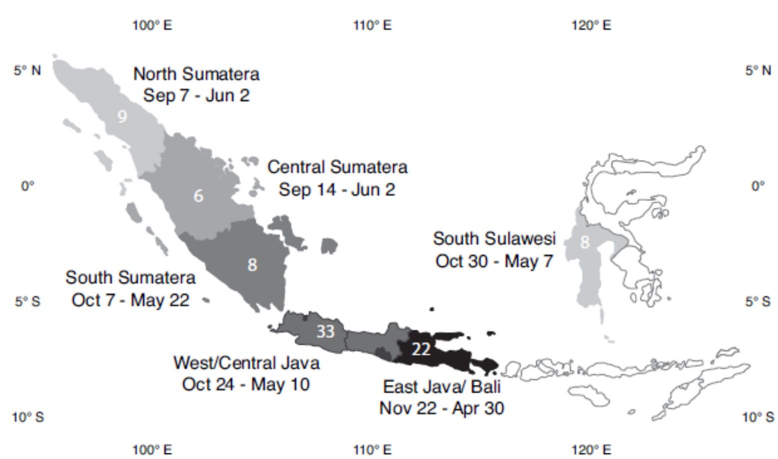
Agriculture, despite its declining contribution to GDP (from 47 per cent in 1969 to around 13 per cent in 2006) employs most rural Indonesians. Agriculture currently accounts for 60 per cent of rural employment, having declined only slightly from 70 per cent in 1990. Two thirds of the households in the bottom two consumption quintiles work in agriculture (Kishore et al., 2000; World Bank, 2008).

[7] See figure 3.1 for local variation.

Only a minority of farm households controls irrigated rice land (*sawah*), implying that rain-fed agriculture continues to play a very important part. The distribution of *sawah* is skewed towards larger landholders: three quarters of agricultural households controlling *sawah* have less than 0.5 hectares of *sawah* each, and together they control only 38% of all *sawah* in Indonesia (McCulloch, 2008).

In Indonesia most of rice is typically planted at the beginning of the rainy season between October and December (see figure 3.1 for regional variations), when there is enough moisture to prepare the land for cultivation and to facilitate the early rooting. However, the main planting period occurs before the peak of the monsoon, because excessive water hampers rooting. During the 3-4 months grow-out period from planting to harvest, rice requires 600-1200 mm of water depending on the agro-ecosystem and the timing of the rainfall or irrigation. A smaller, dry season planting take then place in April and May after the wet season crop has been harvested (De Datta, 1981 in Naylor et al., 2001). Figure 3.1 also shows the share of rice that is produced in each of the regions.

Figure 3.1. The timing of the rainy season and the share of rice production out of total production in selected provinces of Indonesia.



Note: Onset date is the date past August 1 when accumulated rainfall equals 20 cm, averaged over reporting rainfall stations in the region for the years 1979–2004; termination date is the date on which 90% of that year's rainfall has accumulated. The number on each region indicates the share of region's rice production out of total production.
Source: Naylor et al., (2007a).

The timing of the onset of the monsoon is affected by the El Niño Southern Oscillation (ENSO), which causes anomalies in the sea surface temperature and sea-level pressure. *El Niño* events can delay rice planting by up to two months, reducing the area cultivated and delaying the plantings of next year's dry-season crop (Naylor et al., 2007a).[8] In addition to delays in rain, *El Niño* events are associated with reductions in the length of the rainy season (Cook et al., 2001). Monsoon timing affects the total amount of land planted for many crops, but is particularly important for rice. A sea-surface temperature index explains 60 per cent of rice planted in Java, and 40 per cent of the variation in rice production (Naylor et al., 2001). From 1983 to 2004, a 30-day delay in monsoon onset caused rice output to fall, on average, by 580,000 metric tons (11.6%) in East Java/Bali and 540,000 metric tons (6%) in West/Central Java during the main rice harvest season between January and April (Naylor et al., 2007a). Also, studies on rice farming in India have found that variability in area cultivated is higher than yield variability (Walker and Ryan, 1990).

3.2.2 Education system in Indonesia

Indonesia has invested considerably in education in the recent decades. After achieving almost uniform enrolment in primary education, the policy focus has switched to increase the enrolment in secondary education. The secondary enrolments lagged behind, increasing slowly to just over 50 per cent in 1990. To address this problem, the Government of Indonesia extended the obligatory school-going age to 15 years in 1994, implying that the compulsory education was extended to nine years (six years of primary education and three years of secondary, junior high school).[9,10] The junior secondary enrolment rates reached 58 per cent in 1998 and were 65 per cent (net) and 82 per cent (gross) in 2004. However, there remains variation across provinces as well as within them. After completing junior

[8] During El Nino events, the warmer ocean water shifts eastward away from Indonesia causing rain to fall over the central Pacific Ocean.

[9] It is beyond the scope of this study to assess the impact of the reform on school attendance.

[10] The current Government is planning to increase the compulsory education to 12 years by 2014. The critics argue that the Government should first finish the earlier reforms as the net attendance rate in the junior high school was only 67% in 2007 (Jakarta Post 28.6.2010, http://www.thejakartapost.com/news/2010/06/28/analysis-indonesia%E2%80%99s-12year-compulsory-education-program.html, accessed 14.1.2011).

secondary school the child can continue to senior secondary school, subject to, however, competitive entry. Within both junior and senior high school a distinction exists between general and vocational schools (see, for example, Pradhan, 1998, pp. 413-414; Manning, 2000, p. 26; del Granado et al., 2007).

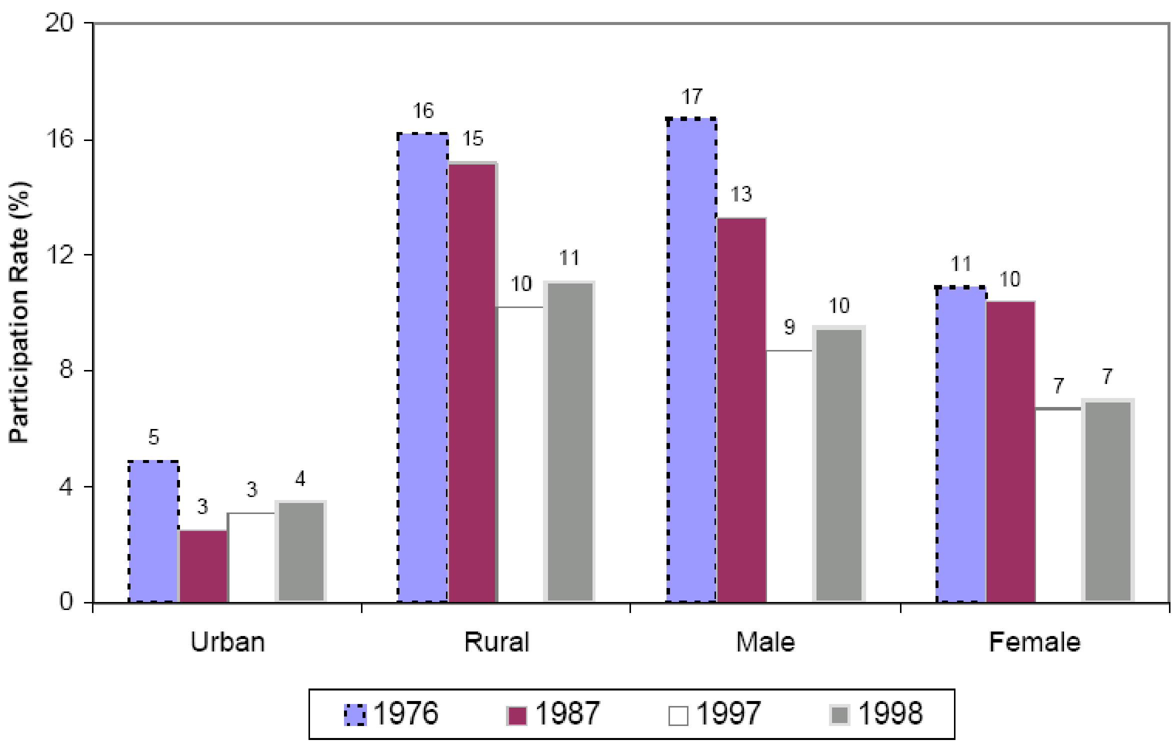
The education expansion in Indonesia has kept up with that of most East Asian countries. However, there remain obstacles to universal education, which could also affect children's engagement in labour force. Firstly, whilst primary school enrolments rates are very high, a significant proportion of children drops out from primary school before completing grade six (nearly 20% in 1993 and 15% in 2004). Also, the continuation from primary to secondary school remains a problem, resulting in a significant loss in terms of educational attainment. As a result, approximately 30% of the primary school completers did not continue to secondary school in the late 1980s and the corresponding figure was 25% in the 1990s (Behrman and Deolalikar, 1991; Manning, 2000).

3.2.3 Child labour in Indonesia

There are a few, albeit descriptive, studies on the prevalence of child labour in Indonesia (see, for example, Manning 2000). Priyambada et al. (2005) compare years 1998 and 1999 in order to shed light on the question about the extent to which Asian financial crisis affected child labour and school attendance in Indonesia. However, the authors do not use any exogenous variation for identification but instead compare two subsequent years. The findings suggest that the probability of a child participating in the labour force is higher for males and for children from poor families and children living in rural areas. The probability of working is also higher in female-headed households, in households with a high dependency ratio and in households where the head of the household is working in agriculture. On the other hand, probability of working decreases with the education of the household head.

From a historical perspective, the labour force participation of children aged 10-14 years declined steeply since the mid-1970s mainly due to supply side factors, such as the increase in the supply of primary education and the improvements in living standards allowing parents to better support the education of their children. Along with the increase in education opportunities, demand side factors further contributed to the reduction; in particular, the shift from agriculture and small-scale manufacturing that had been the most significant employers of child labour (Manning, 2000). The steady decrease in child labour slightly reversed in 1998 following the Asian crisis. According to the national labour force survey, SAKERNAS, approximately eight per cent of the Indonesian children aged 10-14 were reported in the labour force. The share was higher in rural areas, 11 per cent (see figure 3.2).[11]

Figure 3.2. Labour force participation of children aged 10-14 years in 1976-1998.



Source: National Labour Force Surveys in Manning (2000).

[11] Manning (2000) argue that SAKERANS understates the extent of child labour because it does not adequately take into account the economic work within the household, which is common particularly in rural areas.

3.3 Data

3.3.1 Data sources

In this chapter the household survey employed is the Indonesia Family Life Survey, IFLS, run by RAND Corporation and the Demographic Institute of the University of Indonesia (IFLS1 and IFLS2) and the Rand Corporation and the Center for Population and Policy Studies of the University of Gadjah Mada (IFLS3). It is a panel survey covering years 1993 (IFLS1), 1997 (IFLS2) and 2000 (IFLS3) and surveyed in autumn.[12] A fourth round of the survey was conducted in 2007-2008, but this has not been used in this study, due to inability to obtain rainfall data beyond 2004. There are also some other caveats for using the IFLS4 in this study.[13] Moreover, panel data techniques are not used in this study in order to maximize the number of observations. There are approximately 4400 children aged 6-16 years for whom two or three rounds of data is available resulting in the total number of observations in the panel model approximately 10,000 children, that is roughly 25 per cent less than in the pooled cross section. This would be a significant reduction in the number of observations. Moreover, I would lose, for example, all children older than 12 years in the 1993 round. Other advantages of the pooled cross section are the ability to estimate marginal effects of the time-invariant characteristics and to maintain comparability of the linear probability model (LPM) and probit estimates. Fixed effect probit estimate is not available and the random effect probit estimator does not allow clustering the standard errors. On the other hand the conditional logit estimator drops all the observations for which the outcomes are only ones or zeros. Moreover, in the child labour section I also estimate a single cross section for the year 2000. Hence in order to maintain comparability of the estimates panel data techniques are not used in this study.

---

[12] Additional information about the survey is provided in Strauss et al., (2004), Frankenberg and Thomas (2000), and Frankenberg and Karoly (1995).

[13] Firstly, IFLS4 was fielded at different time of the cropping season compared to the earlier waves. Second, as explained more in detail later in this section, information about the timing of the rainfall has been disseminated to rural areas since 2005. And finally, the year 2007 was considered more of La Niña rather than El Niño and therefore including the 2007 would likely not increase the variation in the main independent variable.

IFLS data is a rich data set that provides detailed data at the individual and household level on, among others, health, education, migration, employment, income and consumption. The first IFLS round sampled 311 villages, covering approximately 7,200 households in 13 provinces in Indonesia, representing approximately 83% of the Indonesian population.[14] Subsequent rounds attempted to re-contact all households interviewed in 1993, and households' attrition rates were generally below five per cent. The IFLS survey covers virtually all of the provinces highlighted in figure 3.1, which account for roughly 85 per cent of the national rice production. Out of the total sample approximately 52.4% of the households were located in rural areas. The exclusion of urban areas limits the total sample to 13,348 children aged 6-16 years and 6,792 children aged 6-19 years who have already completed primary school.

The rainfall data employed is from the NOAA (National Oceanic and Atmospheric Administration) Global Summary of the Day combined with additional data obtained from the Indonesian Meteorological Agency (*Badan Meteorologi dan Geofisika, BMG*). Imputed values for the missing values in the Global Summary of the Day data were provided by CEREGE, *Centre Européen de Recherche et d'Enseignement des Géosciences de l'Environnement*.[15] The rainfall data set contains daily rainfall data for the period 1979-2003 for 52 stations, of which 36 stations match with the IFLS data. Original IFLS households were matched with the closest weather station at the community level and households that have moved location were matched to the rainfall station closest to the geographic centre of their new district.

The start of the monsoon is defined as the number of days past August 1 when cumulative rainfall exceeds 20 cm, following Naylor et al., (2007a). The rationale for this definition is that 20 cm of cumulative rainfall is needed to moisten the ground for rice planting. For each station, I calculated the start date of the monsoon and onset was then standardized using each station's 'leave-out' mean and standard deviation across years. In other words, data from the onset year was excluded when calculating the mean and standard deviation used to standardize each year's onset. A value of O_{ijt-1} equal to zero

[14] There are currently 33 provinces in Indonesia, at the time of the first survey there were 27 provinces.
[15] More information on the share of imputed values in the data is presented in Appendix A, Appendix to Chapter two.

would indicate that the nearest station's monsoon onset last year was equal to its historical average, while a value equal to one would indicate that last year's monsoon arrived one standard deviation late. The standard deviation of monsoon onset across the entire sample is 24 days. Alternative definitions of monsoon onset are also available, such as in Moron et al., (2009) that takes into account false starts, that is, dry spells occurring after the threshold has been reached. Nevertheless, Moron et al., (2009) argue that their estimations of mean onset dates for various regions in Indonesia are consistent with Naylor et al., (2007a).

3.3.2 Descriptive statistics

In the literature school attendance has been a widely used measure of education outcomes (see, for example, Al-Samarrai and Reilly, 2000; Lavy 1996; Kruger 2007). The relative merit of focusing on attendance in the current application is also the fairly straightforward way of linking the outcomes with the monsoon onset variables.[16] From table 3.1 we can see that the approximately 80 per cent of the rural children during the study period are attending school. Girls and boys aged 6-16 years are equally likely to attend school. In the school attendance specification I first focus on children aged 6-16 years who have not yet completed the compulsory education, i.e. children with less than nine years of education. Furthermore, approximately 6.4 per cent of children aged 6-16 years have never attended school, while the corresponding figure for children aged 6-10 years is 10.8 per cent.

Table 3.1. School attendance by gender, children aged 6-16 years who have not yet completed compulsory education (grade 9), IFLS1-IFLS3.

Currently attending school	Boys	Girls	Total
Yes	5,389 (80.4% of boys)	5,342 (80.4% of girls)	10,731 (80.4% of total)
Observations	6,706	6,642	13,348

[16] Educational attainment is part of the future research.

Later I also study the effect of monsoon onset in the transition year (from primary to secondary) on the school attendance in the following years. The decreased enrolment rates after primary school in table 3.2 reveal that drop-outs and class repetition are a problem in rural Indonesia as only 56.3 per cent of the children aged 6-19 years who have completed primary education are still attending school.[17] Another important observation in table 3.2 is that girls' enrolment rate for post-primary school is slightly lower than boys' enrolment rate.

Table 3.2. School attendance by gender, children aged 6-19 years who have completed primary school, IFLS1-IFLS3.

Currently attending school	Boys	Girls	Total
Yes	1,945 (57.9% of boys)	1,880 (54.8% of girls)	3,825 (56.3% of total)
Observations	3,359	3,433	6,792

Information provided in the IFLS household roster give significantly lower rates for children who have worked in the past 12 months compared to the child labour figures for rural areas provided by the Sakernas survey (see figure 3.2). According to the national labour force survey, approximately 11% of rural children were working in 1998, compared to the average of 4.8 per cent in the IFLS survey (see table 3.3 below).

Table 3.3. Labour force participation in the past 12 months, children aged 10-14 years, IFLS1-IFLS3.

Did the child work?	Boys	Girls	Total
Yes	176 (5.5% of boys)	130 (4.0% of girls)	306 (4.8% of total)
Observations	3,188	3,222	6,410

[17] The corresponding figure for children aged 6-16 years is 70.4%.

On the other hand, the wave 2000 has a separate section on child labour in the child book and according to this data approximately 17.1% of rural children aged 10-14 years worked either for wages or as a family worker in the past month in the year 2000 (see table 3.4 below).[18] The proportions of boys and girls reported to have worked seem approximately the same.

Table 3.4. Labour force participation in the past month, children aged 10-14 years, IFLS3.

Did the child work?	Boys	Girls	Total
Yes	199 (17.5% of boys)	183 (16.6% of girls)	382 (17.1% of total)
Observations	1,138	1,101	2,239

Comparing the data on child labour in different sources my judgement is that information presented in the child book is more accurate than the information available in the household roster. The child book contains information on both the wage work and work on family business, and moreover, the respondent is the carer of the child or the child her/himself, who likely have the best information about the work engagement. Therefore, in the child labour analysis I will focus on the data in wave three (2000). However, for comparison, I will also present the pooled model using information from the household roster from the waves 1993, 1997 and 2000.

Of those children who are not attending school in 2000, approximately 38 per cent reported to have worked in the past month. The figure is smaller for children who are attending school, approximately 14 per cent. Further, 27 per cent of children who worked in the past month are not attending school and accordingly 73 per cent are attending school (see Table 3.5).

[18] It is worth noting that the in the household roster the question refers to the past 12 months where as in the child book, wave three, the question refers to the past month. It is also notable that in 2000 the child book contain separate questions on work for wages and work on family business and I have combined these 2 question in order to construct an overall measure of child labour (including both family labour and wage work). In section 3.6.1 I present more detailed information on work for wages vs. family labour. The correlation coefficient between work definition in household roster and child book in year 2000 is 0.37 and it is statistically significant at the 1% level.

Table 3.5. Working and attending school, children aged 10-14 years, IFLS3.

Working	Attending school		
	No	Yes	Total
No	174	1,680	1,854
Yes	105 (37.6% of not attending)	277 (14.2% of attending)	382
Observations	279	1,957	2,236

Regarding monsoon onset, none of the years prior to the IFLS years (i.e. 1992, 1996 and 1999) experienced a strong *El Niño* event to an extent that monsoon onset could be delayed as much as nearly four standard deviations (up to three months). Notwithstanding this, monsoon onset shows meaningful variation for the proposed analysis (see figure 2.3 in chapter two).

3.4 Empirical approach

I now discuss the specifications that I use to study the effect of weather shocks and risk and other conventional variables, such as parental education and household wealth, on school attendance and child labour. There are several aspects that need to be taken into account when choosing the specification. These include pooled vs. panel regression model, non-linear effect of monsoon onset, treatment of standard errors and the endogeneity of the per capita expenditure, a proxy for permanent income.

The empirical approach aims to exploit the information on whether a child is currently attending school or has worked in the past month or past 12 months.[19] The probit model is the most common method in the literature to estimate demand for schooling/child labour in this context. Another option is a linear probability model (LPM), i.e. ordinary least square

[19] Therefore the specifications in this study do not capture the intensity of schooling/working but only the extensive margins.

regression. In the following I discuss the relative merits of these models for the purposes of this study.

First, a reason to prefer the probit model over the LPM is that the estimated probability lies between [0,1]. On the other hand LPM better enables the implementation of instrumental variable regression (IV-regression), which is needed to address the possible endogeneity of the household expenditure measure. Instrumenting the expenditure also enables me to reduce the potential measurement error related to the expenditure measure (see, for example, Al-Samarrai and Reilly, 2000). LPM also enables the use of fixed effects, which is part of the future research.

Finding suitable instruments for per capita expenditure is not straightforward, however.[20] A common method in the literature is to use household asset measures as instruments for household expenditure. Assets are correlated with household expenditure and therefore fulfil the criterion of relevance. However, the criterion of validity, requiring instruments to be uncorrelated with the error term, is more difficult to assess. Valid instruments start a unique causal chain; i.e. create exogenous variation in the endogenous variable that in turn changes the outcome variable. Instruments should not be directly correlated with the outcome variable, only though the endogenous variable (see, for example, Murray, 2006). Therefore the validity of the asset as an instrument could be questioned.[21] Even though assets are clearly correlated with household expenditure it could be argued that higher household expenditure enables higher assets and not vice versa. This is could be the case particularly with durable assets and housing conditions. However, the market for land is relatively narrow in rural Indonesia and the legal and institutional framework is very complex (World Bank, 1994). Therefore, I argue that the value of land is a suitable instrument for per capita expenditure.[22] An IV estimate captures the causal effect for those households whose behaviour (per capita expenditure in this case) can be manipulated by the instrument (real value of land). The effect is generally known as a local

[20] It is notable that lagged monsoon onset does not have sufficient power as an instrument for household expenditure.
[21] Sargan test can be used to assess the validity of the instruments when the number of instruments exceeds the number of endogenous variables. However, Sargan test relies on the assumption that at least of the instruments is valid (see Murray 2006).
[22] IFLS data provide information on the value of land, judged by the household itself and not on the amount of land.

average treatment effect (LATE). The households in this group are called compliers (see, for example, Angrist and Krueger, 2001). It is notable that the LATE is not informative on those households whose per capita consumption is not affected by their land holdings.

However, the linear probability model is likely to overestimate the impact of delayed monsoon onset in the pooled child labour specification. Only approximately 5 per cent of the children aged 10-14 years were reported to have worked implying that very few children engaged in labour experienced delayed monsoon onset and therefore the IV-regression in the linear probability framework may overstate the effect (see section 3.6.2 for further discussion). Therefore, my main specification is a pooled probit model where the share of food in the budget proxies per capita expenditure.[23] As a robustness check, I also present an IV regression in the linear probability model where the per capita expenditure is instrumented by the real value of land the household owns.[24] Endogeneity of household expenditure is ultimately not testable. However, Stata reports an endogeneity test for the IV estimation which can give some indication about the endogeneity.[25] In the current application the test statistic suggests that household expenditure is endogenous in the pooled school attendance model but exogenous in the child labour specifications.[26] This finding is not in line with my prior expectations and therefore the IV estimation is used mainly as a robustness check, and the share of food proxies wealth in my main specifications.

Another important issue is to allow a non-linear relationship between the outcomes of interest and the timing of monsoon onset. Findings of chapter two of this thesis suggest that monsoon onset has a non-linear impact on household per capita expenditure and farm profits. As in chapter two I use the linear spline function (see, for example, Gujarati, 2003) in this study. Linear splines replace the onset variable by a set of piece-wise linear segments allowing monsoon onset to exert differential effects on school attendance and

[23] Correlation between share of food in the budget and per capita expenditure is -0.185, and it is statistically significant at the 1 per cent level.
[24] Per capita expenditure excludes education expenditure because household expenditure on education is used to construct the community average cost of schooling.
[25] With clustered standard errors the test is equal to inclusion of the residuals of the first stage to the model as additional regressors. If the coefficient of the residuals in this augmented regression is statistically significant then the endogenous variable is considered as endogenous.
[26] Similar results on the endogeneity are obtained using the IVPROBIT model. Importantly, IVPROBIT confirms the main results of this study.

child labour at different locations of the monsoon onset distribution, determined by the threshold values, knots. The key empirical issue here relates to the choice of the knots; more specifically, the location and the number of knots. In this study the choice is made by experimentation and the best fit is obtained using three splines.[27, 28]

In the following equations O_{ijt-1} stands for monsoon onset and indicates the timing of monsoon onset (in standard deviations compared to the historical mean) for individual i living in province j at the nearest weather station previous year to the survey. For example, if we assume three linear splines, the form is expressed as follows:

$$f(O_i) = \phi_1 spline_1 + \phi_2 spline_2 + \phi_3 spline_3\,, \tag{3.1}$$

Where below the first threshold value O^*_{t-1}, $O_{ijt-1} \leq O^*_{t-1}$:

$$spline_1 = O_{ijt-1}$$
$$spline_2 = 0$$
$$spline_3 = 0.$$

Between the two threshold values O^*_{t-1} and O^{**}_{t-1}, $O^*_{t-1} < O_{ijt-1} < O^{**}_{t-1}$:

$$spline_1 = O^*_{t-1}$$
$$spline_2 = O_{ijt-1} - O^*_{t-1}$$
$$spline_3 = 0.$$

[27] The use of splines allows the early onset to exert a different (in absolute terms) effect compared to late onset. This might be of interest especially in the child labour specification because the coefficients of the first and third splines are evidence of the relative strengths of the underlying income and substitution effects.

[28] In chapter two of this thesis I use two lags of monsoon onset. However, the second lag proved to be unnecessary in the current application. I checked that the results presented in this chapter are robust to including the second lag but because the estimates of the second lag are close to zero and statistically insignificant I dropped the variable.

And above the second threshold value, O^{**}_{t-1}, $O_{ijt-1} \geq O^{**}_{t-1}$:

$$spline_1 = O^{*}_{t-1}$$
$$spline_2 = O^{**}_{t-1} - O^{*}_{t-1}$$
$$spline_3 = O_{ijt-1} - O^{**}_{t-1}.$$

In the equations above O^{*} and O^{**} are threshold values, knots, determined in advance. The final specifications have two knots and they are located at -0.5 and 0.5 standard deviations. The estimated effect for ϕ_1 provides the average effect, *ceteris paribus*, of monsoon onset on attendance/child labour if the onset falls within the first linear segment, i.e. if monsoon onset is less than the value $O*$. Respectively, the estimated effect for ϕ_2 provides the average effect for a monsoon onset that falls between the two knots and, finally, the estimated effect for ϕ_3 provides the average effect for an onset that falls within the third segment, i.e. an onset greater than the value $O**$.

My main specification could be expressed as follows:

$$PR(S_{ijt} / L_{ijt} = 1) = \alpha + \phi_1 spline_{1,t-1} + \phi_2 spline_{2,t-1} + \phi_3 spline_{3,t-1} + X'_{ijt}\beta_1 + \gamma_j + T_t + \varepsilon_{ijt}, \tag{3.2}$$

where S_{ijt} is a binary variable, taking the value one if the child i, living in province j is enrolled in school in year t and L_{ijt} is a binary variable taking value of one if child has worked in the past 12 months either for wages or as a family worker. X'_{ijt} is a vector of individual and household characteristics, including age and age squared of the child, gender of the child, parental education, dummies for maternal and paternal orphans, religion, gender and age of the head of the household, share of food in the budget as a proxy for permanent income, community average travel time to school, community average cost of schooling[29] and household demographic structure defined as number of below school age children, number of school age children, number of adults and number of old people in the household. Also dummies indicating households' farm ownership or engagement in non-farm business are included in the characteristics. γ_j is a vector of province dummies and T_t is a vector of year dummies. ε_{ijt} is a stochastic error term, which is robust to heteroscedasticity and clustered on rain stations.[30]

Wave three in the year 2000 has more detailed and accurate information on child labour. To take advantage of this information I estimate a probit specification for the year 2000 where the monsoon onset variable appears in linear form because linear splines may overestimate the effect of delayed onset due to the decrease in sample size.[31] The estimated equation is:

$$PR(L2_{ij} = 1) = \alpha + \delta_1 O_{ijt-1} + X'_{ij}\beta_3 + \gamma_j + \varepsilon_{ij}, \qquad (3.3)$$

where $L2_{ij}$ is a binary variable, taking the value of one if the child worked for wages or as a family worker during the past month. O_{ijt-1} stands for standardized monsoon onset in the previous year, X_{ij} is a vector of individual and household characteristics as described earlier.

Primary school participation is almost uniform in Indonesia. However, drop-outs appear to be a major problem in later stages. In this context delayed monsoon onset might have a more significant role to play during the transition year from primary to secondary school.

---

[29] I constructed household average cost of schooling (tuition fees and other costs, including supplies, uniforms, registration fees) per household member attending school and then calculated the community average of this cost measure.

[30] Climate shock might also suffer from spatial autocorrelation, that is, shocks in two regions are likely to be correlated if the regions are in close proximity. However, no attempt is made to correct the standard errors for spatial autocorrelation in this study.

[31] When placing the knots at -0.5 and 0.5 standard deviations delayed monsoon onset increases the probability of child aged 10-14 years working by over 40 percentage points in the IV regression and over 30 percentage points in the pooled probit model.

To test this hypothesis I estimate the following specification for children aged 6-19 years who have completed primary school:

$$PR(S_{ijt}=1) = \delta_2 OT_{ij} + X'_{ijt}\beta_5 + \gamma_j + \lambda_{bc} + T_t + \varepsilon_{ijt}, \tag{3.4}$$

where S_{ijt} is a binary variable, taking the value of one if the child is attending school at the time of the survey. In this specification, OT_{ij} is the standardized monsoon onset, in linear form, in the transition year, i.e. in the year when the child was supposed to transfer from primary to secondary school and λ_{bc} represents birth cohort fixed effects. In Indonesia children normally start school at the age of six, implying that children who are born in August-December start school in the year they turn to seven. Using this information I construct the year each birth cohort started primary school. I further assume that the transition year is six years after starting the primary school.[32] It is notable that in this case the timing of the IFLS surveys does not restrict the analysis and therefore we can use the monsoon onset in the transition year instead of the first lag.

Finally, I study whether the riskiness of the weather affects parents' decision to send their children to school. The riskiness of weather is measured as the coefficient of variation of monsoon onset (see for example Rose, 2001).[33] By definition, the coefficient of variation is time-invariant. Therefore, in my view, the most relevant research question is whether the child has ever attended school. The estimated equation is:

$$PR(SE_{ijt} = 1) = \alpha + \theta_1 COEF + X'_{ijt}\beta_7 + \gamma_j + T_t + \varepsilon_{ijt}, \tag{3.5}$$

[32] IFLS survey also includes a question on the year/age child completed primary school. Because of the high proportion of missing values in this question (approximately 50 percent) I decided to construct the transition year assuming that it occurred six years after the starting of the primary school. Because of the class repetition this measure is undoubtedly measured with error to some extent.

[33] Coefficient of variation is the standard deviation over the mean.

where the variable SE indicates whether the child has ever attended school and the variable COEF is the coefficient of variation of monsoon onset.

The empirical approach is based on the assumption that monsoon onset, as any rainfall variable, is exogenous. Therefore we can interpret the changes in schooling and child labour as causal effects due to the timing of monsoon onset, and any omitted variables of the model should not bias the estimate of monsoon onset. Several previous studies have assumed that rainfall is exogenous with respect to household behaviour (see, for example, Paxon, 1992; Rose, 2001; Munshi, 2003; Newhouse. 2005; Jayachandran, 2006; the literature is surveyed in Rosenzweig and Wolpin, 2000). However, some caution should be exercised. In the past decade the *El Niño* phenomenon has been extensively modelled and relatively accurate predictions about the occurrence of *El Niño/La Niña* are available. As a result, it could be argued that monsoon onset cannot be treated as purely exogenous. Put differently, rainfall itself is exogenous but forecasts alter farmers' behaviour. However, I argue that this is unlikely a problem in the current application. Firstly, even if strong *El Niño/La Niña* years are possible to predict, at least to some extent, my data does not cover these 'big' events and there remains local variation in the timing of the monsoon that is not covered in the national aggregate forecasts. Moreover, the systematic dissemination of the available information and forecasts to rural areas started only as a pilot project in 2005.[34] Therefore, as far as I could judge, these farmers did not possess accurate forecasts on monsoon onset.[35]

In the rest of the study I focus on the effects of early and late onset. I define early onset as a monsoon onset that arrives one standard deviation before the historical mean, and late onset as a monsoon onset that arrives one standard deviation after the historical mean. The multiplication of the coefficient of the first linear spline/coefficient of linear monsoon onset by the negative one gives us the effect of the early onset and correspondingly, the

[34] Information has been disseminated via Climate Change Field Schools organized for farmers. For further information see: http://www.agrometeorology.org/topics/accounts-of-operational-agrometeorology/climate-field-schools-in-indonesia-coping-with-climate-change-and-beyond. Accessed 10th October 2009.

[35] Field study surveyed in the main rice production kabupatens in West and East Java in late 2000s decade indicated that formal climatic data were used in the timing of the farming activities (Natawidjaja et al., 2009).

multiplication of the coefficient of the third linear spline/coefficient of linear monsoon onset by one gives the effect of the delayed onset.

3.5 Empirical results schooling

3.5.1 The effect of weather shocks on school attendance

Equation (3.1) is estimated for children aged 6-16 years, who have not yet completed the compulsory education. I also divide sample into young children aged 6-10 years and older children aged 11-16 years in order to examine whether the impact of timing of monsoon onset on school attendance is constant across the age groups. The division of the analysis into two age groups is supported by the finding that younger children were at a disadvantageous position to elder children in terms of school attendance during the economic crisis in late 1990s. Thomas et al. (2004) argue that given the higher rates of return to secondary education in Indonesia, protecting the education of elder children is a prudent choice in resource-scarce households. Accordingly, the authors' results suggest that older children, males aged 16-19 years and females aged 14-19 years were more likely to be enrolled in school in 1998 relative to 1997.

All children aged 6-16 years

Estimates for rural children aged 6-16 years are presented in table B4 in Appendix B, for a brief summary see table 3.6 in section 3.6.2. As robustness check I also present results from the IV regression in the linear probability model where the per capita expenditure has been instrumented with the real value of household's land.

Overall, variation in monsoon onset does not have a significant impact for school attendance. For all children aged 6-16, monsoon onset arriving 0.5 standard deviation early has no impact on the probability of a child attending school. The impact of monsoon onset on school attendance is increasing in the interval between -0.5 and 0.5 standard deviations, but the estimated effect does not gain statistical significance. Monsoon onset arriving 0.5 standard deviations or later compared to the historical average has a decreasing impact on school attendance, but the effect is again not statistically significant at the conventional levels.[36] Separate regressions for boys and girls do not change this overall finding (results not reported here).

As expected, the probability of attending primary school is increasing with age but at a decreasing rate: according to the estimates for age and age squared the turning point is 10.34 years. The estimated turning point could be an indicator of both delayed enrolment and the problem of drop-outs. The marginal effects for the gender of the child and the gender of the household head are not well determined. Children living in Christian households are 5.4 percentage points more likely to attend school compared to children living in Muslim households, but the effect is only marginally significant at the 10% level.

In line with previous studies parental education is an important determinant of children's education (see, for example, Lavy, 1996; Al-Samarrai and Reilly, 2000)[37]. For instance, if the mother has some primary education or has completed primary education, this increases the probability of a child attending school by 4.9 or 9.4 percentage points respectively, compared to children whose mother has no formal education. Children whose father have completed primary education or some secondary education are 5.2 or 7.1 percentage points more likely to be attending school compared to ones whose father has no formal education. However, it is notable that the highest categories of parental education are not well determined in the IV specification even though they are highly significant in the pooled probit regression.[38] One possible explanation is that the fitted values of per capita expenditure absorb the effect of high parental education. The important role of parents in children's education is also captured by the orphans in the sample. Children who have lost their mother or father are 3.7 or 5.9 percentage points less likely to attend school compared to children whose parents are still alive.

[36] The coefficient for monsoon onset in linear form is close to zero and statistically insignificant at the conventional levels.
[37] In this specification only the marginal effect for father having some primary education is not statistically significant at the conventional levels.
[38] In the LPM the marginal effects of parental education are similar to the pooled probit model.

The marginal effect of the share of food in the budget, a proxy for household wealth in the pooled probit regression, indicates that one percentage point increase in the share of food in the budget decreases the probability of attending school by 0.00271 of a percentage point. The estimated elasticity is -0.212.[39] As anticipated, the distance to school is an important determinant of school attendance in rural Indonesia. Increasing the distance to school (measured as one-way travelling time) by 10 minutes decreases the probability of attending school by 0.02 of a percentage point. The corresponding elasticity, -0.04, is very low, though. The estimates for the standard demand variable, logged household per capita expenditure is explored as follows. The IV regression suggest that a 10% increase in per capita expenditure increases the probability of attending school by 3.35 percentage points, and the corresponding elasticity is 0.42, which is reasonable.[40,41]

Generally the estimates from the pooled probit are rather consistent with the IV specification. However, there are some notable differences. Firstly, as already discussed earlier, per capita expenditure is likely absorbing the effect of parental education reducing the magnitude and/or statistical significance of the effects of parental education in the IV specification.

Second, the effects of household demographic characteristics differ. The effect of the number of children below school age is negative and statistically significant in the pooled probit model but positive and statistically significant in the IV-regression. The effect in the IV regression seems rather counterintuitive given that young children need care which commonly is the task of the siblings. The marginal effect of the number of school age children in the family is negative and statistically significant in my main specification but positive in the IV model. The result from the pooled probit suggests that there is some evidence that parents are trading ´quality´ for ´quantity´ (see for example Montgomery et al. 1995). The marginal effects of number of adults and number of old persons in the households are positive in both specifications, but statistically significant only in the IV-specification. This finding suggests that children and adults do not compete for the same

[39] The elasticity is computed using the formula: marginal effect*(mean share of food/sample proportion).
[40] The elasticity is computed using the formula: marginal effect/sample proportion.
[41] On estimations of income/expenditure elasticities on education in developing countries, see for example Behrman and Knowles (1999).

resources. Finally, the marginal effect for children living in households that own a non-farm business is only statistically significant in the IV-specification: these children are 7.0 percentage points less likely to attend school compared to children in households not owning a non-farm business.

Young children aged 6-10 years

Estimation results for young children aged 6-10 years are presented in table B5, columns 1 and 2, and they look fairly similar to those of all children. The timing of monsoon onset does not have a significant impact of the school attendance of young children. Monsoon onset arriving earlier than 0.5 standard deviation compared to the historical mean has virtually zero effect, whereas monsoon onset arriving later than 0.5 standard deviations decreases the probability of attending school. The estimated effect of monsoon arriving one standard deviation late is -4.5 percentage points but again, the effect does not gain statistical significance at the conventional levels.

Among young children, girls are more likely to attend school. The estimated impact effect indicates that girls are 1.6 percentage points more likely to attend school than boys. Travel time to school is an important factor determining young children's attendance but the estimated effect is of the same magnitude as for all children. Younger children's attendance is increasing with parental education but there seems to be no notable difference between mother's and father's education. The estimated marginal effects of the share of food in the budget, -0.087, and per capita expenditure, 0.191, are also smaller for young children compared to -0.271 and 0.326 for all rural children. Both the number of below school age children and school age children in the family exert a negative impact on younger children's school attendance.

Older children aged 11-16 years

The estimation results for children aged 11-16 years are presented in table B5, columns 3 and 4. Estimation results show that the lagged monsoon onset has a stronger impact on older children. The second spline is positive and marginally statistically significant at the 10% level, implying that between the interval of -0.5 and 0 standard deviations, attendance of older children is decreasing and increasing between 0 and 0.5 standard deviations.[42] The coefficient of the third spline is negative but not statistically significant at the conventional levels.

Generally the impact effects for maternal education are positive and also higher than for all children.[43] For example, mother having some primary or completed primary increases the probability of attending school for older children by 6.0 or 14.4 percentage points, compared to 4.9 or 9.4 for all rural children, respectively. Also the impact effects of parental education for completed primary education and above are large and significant. However, impact effects for paternal education are not well determined in the IV specification, likely undermined by the expenditure absorbing some of the effects as explained earlier. The estimate for maternal and paternal orphans are marginally statistically significant, impact effects being -0.051 and -0.064, respectively. The number of school age children exerts a negative impact and the number of adults and old people in the household exert a positive impact on school attendance for older children. Older children living in Christian households are 10.6 percentage points more likely to attend school compared to Muslim households and the impact effect is statistically significant at the 1% level. Furthermore, the marginal effect of the share of food in the budget and per capita expenditure are high and also highly statistically significant: a one percentage point decrease in the share of food in the budget increases the probability of attending by 0.0045 of a percentage point, and correspondingly, a 10% increase in the per capita expenditure increases the probability of attending school by 5.24 percentage points.

[42] The significance of the second spline might indicate that monsoon onset in linear form could fit the data better. The estimate of the pooled probit with monsoon onset in linear form suggests that one standard deviation delay in monsoon onset increases the school attendance of the older children by 2.6% and it is statistically significant at the 5% level. However, the estimate of the linear monsoon onset is not statistically significant in the IV-model. Together this gives weak evidence that parents may protect the education of their older children when facing delayed onset, similar to the finding of the education outcomes during the financial crisis (Thomas et al., 2004).

[43] Somewhat surprisingly the impact effect for mother having university education is negative and statistically significant in the IV specification. However, only 110 mothers of children above 10 years have university education.

3.5.2 Transition from primary to secondary school

In the earlier sections I have provided evidence that monsoon onset plays limited role in the participation in compulsory education in rural Indonesia. However, monsoon onset could have a larger effect during the years when children are more vulnerable to drop out. The transition from primary to secondary school has been identified as a crucial turning point in school progress in Indonesia (see, for example, World Bank, 2006b, p. 69). To test this hypothesis, I estimate equation (3.4) for children aged 6-19 years who have completed primary education, the results are presented in table 3.6 (for further details, see table B6 in the Appendix B, columns 1, 2 and 3). I assume that the transition year is six years after starting the primary school.

I find that monsoon onset in the transition year indeed has a decreasing impact on school attendance in the following years and the estimated effect is statistically significant at the 1% level. Monsoon onset arriving one standard deviation late in the transition year reduces the school attendance in the following years by 2.8 percentage points, and respectively, monsoon onset arriving one standard deviation early increases the school attendance by 2.8 percentage points.[44] The result is robust to the IV regression, although in the IV specification the effect is slightly smaller (-2.1 percentage points) but also statistically significant at the 1% level. The result is also robust to inclusion of children aged 6-16 years only and fitting a spline function rather than monsoon onset in linear form.[45]

Other variables determining the continuation from primary to secondary education are wealth (permanent income, proxied by the share of food in the budget), parental education, and gender, among others. Girls are 7.7 percentage points less likely to attend school after primary education compared to boys. Controversially, the community cost of schooling

[44] I have included observations regardless of whether they are present in more than one wave after the transition year. For example, persons who transferred to secondary school in 1992 might be present both in 1993 and 1997 waves. This enables capturing the lasting effects of the monsoon onset in transition year, i.e. I am able to control whether they return to school later even if dropping from school for some period of time. However, this does not change the results. After eliminating the duplicate observations the number of observations decreases from 6786 to 4810 children but the estimated effect is the same and statistically significant at the 1% level.

[45] For the children aged 6-16 years the estimated effect is -0.026 and it is statistically significant at the 1% level. When using the splines the estimated effect for the early onset is 0.064 (statistically significant at the 5%) and the estimated effect for the delayed onset is -0.031 (statistically significant at the 10% level).

exerts a positive impact on continuation from primary to secondary. However, as the cost measure is constructed by using the community average household expenditure on education, the measure likely captures also wealth of the community, as the coefficient is not statistically significant in the IV specification.

The results presented above are grounded on the assumption that the birth cohort correctly represents the school starting age. However, due to delayed enrolments the school starting age, and therefore the transition year is likely to be, at least to some extent, measured with error. As an additional robustness check I re-estimate equation (3.4) using information on school starting year based on parents' recall on the age/year their children started primary school.[46,47] Generally I consider the variable representing children's age more reliable than parents' recall on the year their children started school, and therefore the birth cohort approach is my main specification.[48] Again, I assume that transition year follows six years after the year children started primary school. The results are presented in table B6 in the Appendix, columns 4, 5 and 6.

Monsoon onset arriving one standard deviation late compared to the historical average in the transition year reduces the probability of attending school in the following years by 1.9 percentage points, and the estimated effect is statistically significant at the 1% level. The effect is slightly smaller in the IV specification, -1.0 percentage points, and marginally statistically significant at the 10% level. Together these results suggest that timing of monsoon onset, which is a crucial factor in rice planting and production, play an important role determining children's continuation from primary to secondary school.

3.5.3 The impact of weather risk on school entry

In this section I present results derived from equation (3.5). The dependent variable is a dummy variable indicating whether a child has ever attended school. The main independent

---

[46] Persons aged 15 and above personally answered this question.
[47] In this specification I do not control for birth cohort fixed effects.
[48] The birth cohort approach is also robust to age group of 6-16 years and fitting the splines instead of the monsoon onset in linear form.

variable is the coefficient of variation of monsoon onset, representing the riskiness of the environment. The estimation results are presented in table B7 in the Appendix.

The estimation results suggest that the riskiness of the weather does not play a significant role in determining entry of all children age 6-16. However, risk is an important factor affecting entry into school for young children aged 6-10 years. A 10% increase in the coefficient of variation decreases the probability of a child entering school by 0.3 of a percentage point, ceteris paribus.[49] This is an important finding as delayed enrolment could adversely affect the education attainment or deter the entry completely. Accordingly, Fitzsimons (2007) find that village risk inversely affected years of schooling in rural Indonesia.

Other important factors determining entry into school are wealth measures, parental education, ownership of non-farm business, and gender; factors that decrease the probability of entering include distance to school, number of below school age and number of school age children in the household.

3.5.4 Discussion

In Indonesia, the transition year from primary to secondary school particularly has been identified as an important year affecting future education outcomes. The findings of this study suggest that timing of monsoon onset is a major factor determining children's continuation from primary to secondary school. I construct the variable indicating monsoon onset in the transition year and find that delayed monsoon onset in this particular year reduces the probability of attending school in the following years by 2.8 percentage points. The size of the effect is notable and it is robust to various specifications.

The object of this paper is also to examine the relative importance of weather risk, measured as the coefficient of variation of monsoon onset in school progress. The results

[49] The result is obtained using the formula: 0.1*mean coefficient of variation*marginal effect, the mean coefficient of variation being 0.322.

suggest that parents delay the entry into school of young children in riskier environments. Therefore, to conclude, both ex post weather shock, i.e. delayed monsoon onset during the transition year and ex ante weather risk, play an important role in education outcomes in rural Indonesia. In other respects, monsoon onset in previous year has little impact on compulsory school attendance for all rural children. However, it is notable that I am only measuring school attendance, while monsoon onset could also affect the intensity children are attending school (hours spent in school, among others) and education attainment. Moreover, there seems to be some notable differences between the factors determining the school attendance for young and old children. Family characteristics, such as share of food in the budget as a proxy for wealth, parental educational attainment and religion have stronger impact on older children than for younger ones. This evidence suggests that enrolment in primary school is almost universal in Indonesia, but drop-outs are a serious problem at later stages. Weather shocks, as well as family resources and characteristics, play an important role in the continuation and completion of schooling in rural Indonesia.

3.6 Empirical results on child labour

To study the effect of monsoon onset on child labour I first estimate equation (3.3) with a single cross section data using information only from the year 2000. As explained earlier in section 3.3, there are two sources of information on child labour in the IFLS data, and I will start with the definition that is most consistent and least problematic. Wave three in 2000 has a separate section on child's work history in the child book, including children aged 6-14 years. The section contains separate questions on whether child worked for wages or in family business in the past month. The respondent of the child file is the carer of the child or the child her/himself, who are likely have the best information about the work engagement. I estimate a probit model where the share of food in the budget proxies household wealth. As a robustness check, I also present results from an IV specification in the linear probability framework.

In addition, I present a pooled probit specification for all years 1993, 1997 and 2000 where the dependent variable is a dummy variable indicating whether child worked (the variable is meant to cover both the work for wages and the work in the family business) in the past 12 months. The variable is defined for children aged 10-14 years. Unfortunately, this variable has a large share (16.7%) of missing values in 1993. Therefore, I also present results restricting the sample to years 1997 and 2000 only.

3.6.1 Monsoon onset, work in family business and wage work

Table B8 in the Appendix presents estimation results for the cross section specification, using only information provided in wave three (summary of the main results are presented in table 3.6 below). Table B8 presents marginal and impact effects from pooled probit specification, LPM and IV specification for all children aged 6-14 years, and old children aged 10-14 years. Linear splines do not fit the data well and might overestimate the effect of delayed onset.[50] Therefore, I use monsoon in the linear form and focus on specifications for all children and old children aged 10-14 years. Only 2% of children below 10 years reported to have worked in the past month which constrains a meaningful estimation for this age group. The corresponding figure for children aged 10-14 years is 17 per cent. Working on family farm or family business is much more common compared to wage work: almost 14 per cent of the children aged 10-14 years reported to have worked in the family business in the past month, compared to 4 per cent for wage work.

For children aged 10-14 years, a one standard deviation delay in monsoon onset increases the probability of a child working by 5.8 percentage points. The finding is statistically significant at the 5% level, and the estimate is similar in the IV specification. This finding suggests that monsoon onset arriving later than historical average is associated with increased child labour. Delayed onset could cause crop loss due to reduced area harvested and which, in turn, could increase the price of rice. In chapter two of this thesis I

[50] Restricting the sample to the year 2000 only significantly reduces the number of observations (the number of children aged 10-14 is reduced from 6321 to 2240). For example, in the probit model for children aged 10-14 years the estimate for the third spline, i.e. delayed monsoon onset is 0.336

find that one standard deviation delay in monsoon onset increases the local market price of rice by 6.2 per cent. Increased child labour may have an economic underpinning both in the event of either crop loss or an increase in rice prices, depending, for example, on the net producer status of the household.

Interestingly, the share of food in the budget as well as the per capita expenditure in the IV specification are statistically insignificant, suggesting that child labour in rural Indonesia might not be a result of poverty. This finding contrasts the findings in previous studies (see for example Manning, 2000 and Priyambada et al., 2005).

In respect to parental education, higher parental education is associated with smaller probabilities of child labour. For example, mother or father having some or completed senior high school decreases the probability of a child working by 7.4 or 6.6 percentage points, respectively. As anticipated, the ownership of a farm and non-farm business is an important determinant of children’s engagement in work. The fact of the household having farm business or non-farm business increases the probability of child labour by 6.4 or 8.0 percentage points, respectively.

Further disaggregation reveals that both family work and wage work are increased by delayed onset, but the increase in wage work is slightly better determined (see table B9). One standard deviation delay in monsoon onset increases the probability of a child working on family business by 3.6 percentage points and for wages by 3.2 percentage points. It is notable that neither the share of food in the budget nor the per capita expenditure has a statistically significant effect; this holds both with the wage and family work specifications. As expected, the ownership of farm and non-farm business only increases the probability of a child engaging in family work. Children living in female headed households are more likely to engage in wage work but this relationship does not hold for work on family business.[51] An interaction term between gender and lagged monsoon does not gain statistical significance implying that there are no gender differences in labour supply when children are exposed to delayed onset (see table B10).

[51] Average per capita expenditure of female headed households is slightly smaller than of male-headed households.

Finally, I do not find any evidence that riskiness of weather, measured as the coefficient of variation of monsoon onset, would affect the probability of a child working (results not reported here).

3.6.2 Monsoon onset and child labour, pooled model

Estimation results for pooled cross section for years 1993, 1997 and 2000 are presented in table B11, in columns 1, 2 and 3. Estimation results confirm the earlier finding that delayed monsoon onset increases the probability of a child working; specifically, monsoon onset arriving one standard deviation late increases the probability of child labour by 9.5 percentage points.[52] The estimated effect is higher (0.186) in the IV specification. However, a word of caution is appropriate. Only approximately 5 per cent of the children aged 10-14 years were reported to have worked implying that very few children engaged in labour experienced delayed monsoon onset and therefore the IV regression in the linear probability framework may overstate the effect to some extent.[53]

Interestingly, early onset also increases the probability of a child working, although by a much smaller rate: the marginal effect for the first spline is -0.005. However, it is worth re-emphasizing that monsoon onset in this segment takes only negative values and therefore the estimation result implies that monsoon onset arriving one standard deviation early compared to the historical mean increases the probability of a child working by 0.5 percentage points. Assuming that the early monsoon onset is associated with an increase in the rice harvest and household incomes, this result is consistent with a strong substitution effect.[54]

[52] Placing the monsoon onset in linear form in the pooled probit specification suggest that delay in monsoon onset decreases child labour. However, the coefficient is close to zero (-0.005) and only marginally statistically significant. Further, the result is not robust to IV specification or excluding the year 1993.

[53] The coefficient of the third spline in the standard OLS regression is 0.173. Therefore the higher IV-estimate compared to the pooled probit is likely due to the linear probability model than to the IV regression per se.

[54] This assumption is not entirely plausible, however, as I did not find robust evidence for the positive effect of early onset on farm profits and household expenditure (see chapter 2 of this thesis).

The marginal effect of the share of food in the budget is positive and statistically significant implying the children living in poor households are more likely to work. However, the estimate of the per capita expenditure in the IV regression is not statistically significant. The share of food in the budget is likely an imperfect proxy for income and therefore I cannot conclude that children living in poorer households are more likely to engage in child labour in the pooled model.

Parental education has a negative impact on the probability of a child working, as expected, and the estimated effects are highly significant. Only the highest categories lack statistical significance, likely due to the small number of mothers/fathers who have completed higher degrees. Reflecting the findings in the schooling specification, the fitted value of per capita expenditure is likely to absorb the effect of high parental education in the IV specification. The variables for distance to school and community average schooling costs do not have explanatory power in the demand for child labour.

In the pooled model for 1993-2000 the dummies indicating the ownership of farm and non-farm business are not well determined, confirming the argument that the work variable in the pooled model does not capture work on family business adequately.

Estimation results for pooled cross section for years 1997 and 2000 are presented in table B11 in the Appendix in columns 4, 5 and 6. Restricting the sample to the years 1997 and 2000 only reduces the statistical significance of the first spline, while the third spline remains highly significant. The estimate for the third spline suggest that monsoon onset arriving one standard deviation late increases the probability of a child working by 11.6 percentage points. Again, the estimate for the third spline is much higher in the IV specification (0.267). Nevertheless, the estimation results enable us to conclude that delayed monsoon onset is associated with an increase in child labour.

Table 3.6. Summary results for early and delayed monsoon onset on school attendance and child labour.

	School attendance, children aged 6-16 years	School attendance, monsoon onset in the transition year	Child Labour in 2000, children aged 10-14 years	Child Labour in 1993-2000, children aged 10-14 years
Early monsoon onset	-0.003	0.028***	-0.058**	0.005*
Delayed monsoon onset	-0.063	-0.028***	0.058***	0.095***

Notes: Early monsoon onset refers to an onset that arrived one standard deviation earlier than historical average, and delayed monsoons onset to a one standard deviation delay.

3.6.3 Discussion

The analysis presented above suggests that delayed monsoon onset increases the incidence of child labour. This is confirmed using both the more detailed information on child labour in wave 2000 and in the pooled cross section. The data for 2000 reveals that both work on the family business as well as wage work increases as a result of delayed onset. Further, the results cast doubt whether child labour is a result of poverty as the estimates on share of food in the budget and per capita expenditures in the IV specifications fail to gain statistical significance in the specification using data on the year 2000. In the pooled probit model for 1993-2000 and 1997-2000, the estimate on the share of food in the budget suggests that children living in poorer households are more likely to work but this cannot be taken as robust evidence because the share of food is likely an imperfect proxy for income: the estimate on per capita expenditure in the pooled IV specification is statistically insignificant. The results using data on 2000 suggest that children living in households owning a non-farm business are more likely to work. It seems that these households are wealthier than average households which could, at least partly, explain the result that households' wealth is not necessarily inversely related to child labour.[55]

3.7 Conclusion

The major thrust of this study was to document the effects of delayed monsoon onset on schooling and child labour in rural Indonesia, in an environment of incomplete insurance and capital markets. Previous studies have found that parents might withdraw children from school as a coping mechanism to an exogenous shock to household incomes. The prior assumption about the effect of monsoon onset on child labour is less clear cut. The conventional perception has been that negative income shocks increase child labour. However, recent studies have emphasised the positive relationship between economic upturns and child labour. In this context the key question is whether parents see the shock as a temporary one that could be exploited by increasing the labour supply of their children. Furthermore, this study has also examined the effect of risk on education outcomes.

I find that delayed onset is associated with an increase in child labour. The estimates using data on 2000 suggest that one standard deviation delay in onset increases the probability of a child working by 5.8 percentage points. Further disaggregation reveals that both family work and wage work are increased by delayed onset. The spline functions in the pooled cross section suggest that monsoon onset arriving one standard deviation late compared to historical average increases the probability of a child working by 9.5 percentage points in the course of the surveys. Finally, I do not find any gender differences in labour supply when studying children's exposure to delayed onset.

Previous studies have argued that children might be particularly vulnerable to drop-out from school in specific years. Accordingly, I study the effect monsoon onset in the transition year from primary to secondary school on the probability of attending school in

[55] Using the data for 2000 the interaction term between the indicator variable for non-farm business and share of food in the budget is negative and statistically significant at the 5% for all children aged 6-14 years implying that wealth is not inversely related to child labour in households that owns a non-farm business.

the following years. Indeed, delayed monsoon onset coinciding with the transition year reduces the probability of attending by 2.8 percentage points. The estimated effect is notable and robust to various specifications. On that account, monsoon onset is one factor, among many, explaining drop-out and continuation from primary to secondary school.

In other respects, I find that monsoon onset in the previous year does not affect the compulsory school attendance of children aged 6-16 years. However, the riskiness of the environment, measured as the coefficient of variation of monsoon onset, plays a role in parents' decision to send their children to school. The findings of this study suggest that an increase in the weather-related risk reduces the probability of ever attending school for children aged 6-10 years. Thus, uncertainty about weather, and hence production, is associated with delayed enrolments in rural Indonesia. This finding may suffer from omitted variable problem. For example, due to data availability, I am not able to control for the quality of household's landholdings or access to irrigation. To the extent that the land quality is negatively correlated with riskiness of the weather, the estimate of the education response to weather risk is downward biased. Therefore, if anything, the results presented here underestimates the effect of weather risk. To conclude, both ex ante weather risk and ex post weather shock on a specific year adversely affect school progress and education outcomes in rural Indonesia.

One limitation of this study is that the reduced form specification does not allow me to examine the mechanisms through which late monsoon onset affects households and lead to an increase in child labour. Another limitation is the amount and quality of the rainfall data. Only 36 rain stations can be matched to the IFLS data and none of the years captured by the IFLS follow a strong *El Niño* year. Further, the measure of late onset is based on daily rainfall data, which is measured with a degree of error. Since measurement error in rainfall is independent of household characteristics, the true effect of delayed onset, both positive and negative, is greater than the estimates presented here.

Despite the aforementioned limitations, this study makes an important contribution to the understanding of the nature of child labour in rural Indonesia, and particularly its response to changes in agricultural conditions. The study has also identified two important factors that threaten school enrolment and progress in rural Indonesia: delayed monsoon onset in transition year and weather risk. An important policy conclusion that emerges is that better insurance policies and/or credit opportunities might help households to cope with the delay in the rainy season. Further, the study provides evidence suggesting that enhancing the weather forecasting systems and the distribution of weather-related information potentially help rural households to cope with weather shocks. Finally, further research on the intensity of schooling and education attainment is needed to complete the analysis on the impact of monsoon onset on education outcomes and to assess the interaction between child labour and schooling.

Chapter 4 A Household welfare perspective on the expansion of palm oil production in Indonesia

4.1 Introduction

The bulk of the worldwide expansion of oil palm plantations in recent decades has occurred in South-East Asia, and in particular in Indonesia, where the total land area devoted to palm oil has increased more than 2100 per cent since the early 1980s (Sheil et al., 2009). The welfare impacts of these plantations are being debated. There is evidence suggesting that oil palm plantations and other biofuel sources bring additional income for Indonesian households living in remote areas (see, for example, Peskett et al., 2007), and the proponents of palm oil claim that the industry has strong spillover effects. Also, the Government of Indonesia has subsidized smallholder production by providing credits and land. On the other hand, palm oil producers are blamed for extreme forest degradation, forest fires due to land clearing, and soaring food prices (see, for example, Naylor et al., 2007b; Sheil et al., 2009; Rist et al., 2010). Despite being one of the most important topics in Indonesia, in terms of environmental policy, climate change and rural development, it is striking how little systematic empirical research exists on how households are affected by the expansion of palm oil production.

The objective of this paper is to evaluate the welfare implications of the expansion of palm oil production in Indonesia, primary focus being smallholder production. However, as I am interested in evaluating the nationwide impact on society as a whole, I will not limit the study exclusively to producers. The rapid expansion of palm oil production suggests that its effects will also be distributed to non-producers. Therefore, the aggregate effects are of crucial importance and should be taken into consideration when planning any future expansions. My approach is to study the impact of the increase in the area of oil palm plantations and the level of palm oil production, at the district level, on the welfare of the households and individuals located in these districts, regardless of whether they produce palm oil or not.[1]

The main welfare indicator used in this study is the household per capita expenditure. In addition, I will study the impact of the expansion of palm oil production on the probability of a household member reporting symptoms of asthma. This latter indicator was selected to evaluate the indirect costs of palm oil expansion. In particular, the conversion of tropical forests often involves forest fires, which in turn could have adverse effects on health, and particularly on breathing (see, for example, Osterman and Brauer, 2001). Palm oil production also involves toxic waste being released by refineries, and previous studies suggest that proximity to toxic waste correlates with increased levels of asthma (National Research Council 1991, p. 171). In addition, odours coming from these refineries might prove harmful to locals' health.

Forest fires are a frequent phenomenon in Indonesia. Although long dry spells related to the *El Niño* phenomenon typically worsen the situation, the prime cause of such fires is often land clearing for plantations, as burning the land is still regarded as the quickest and cheapest method. The study by Frankenberg et al. (2005) finds that the 1997 forest fires due to *El Niño* had a negative impact on the health of those individuals affected.[2] Also Osterman and Brauer (2001) report several studies that have documented the association between respiratory problems and both indoor and outdoor wood burning. However, the health impacts of the expansion of palm oil production have not yet been studied. Importantly, although land clearing using fire is now prohibited by law, smallholders continue to use this method due to a lack of machinery for alternative land clearing methods (Casson et al., 2007 in World Bank, 2010).

To my knowledge, this is the first study seeking to evaluate the costs and benefits of the expansion of palm oil production using large samples of survey data. The existing, albeit descriptive, studies have focused only on few villages at a time (see, for example, Feintrenie et al., 2010; Rist et al., 2010). Another descriptive study by Kessler et al., (2007)

[1] Oil palm (elaeis guineensis) is the plant where as palm oil refers to the oil that is extracted from the palm: crude palm oil from the fruit and palm-kernel oil from the seed.
[2] In this study health status is measured as an ability to carry a heavy load, for example.

examines the socioeconomic impacts of the production of selected agricultural commodities, by comparing province-level outcomes in the mid 1990s with those in the early 2000s. The evidence regarding palm oil production in Indonesia is mixed; some indicators, such as employment, performed better than the national average, while others performed worse, particularly GDP per capita, poverty rates[3], and food security.

Indonesia has a sound track record in poverty alleviation. During the period of the late 1970s to the mid 1990s the poverty rate was halved (from 33 percent in 1978 to 17.6 percent in 1996), but it subsequently rose again to 23 percent in 1999 due to the financial crisis. Poverty rates declined again after a strong stabilization programme, but then increased again in 2006 following the ban on rice imports (World Bank, 2006a).[4] Some descriptive village-level studies suggest a link between increased palm oil production and poverty reduction (see, for example Susila, 2004 in Rist et al., 2010), but this has not been verified in any larger scale study.

I employ annual data from the national household socioeconomic survey, SUSENAS, matched with district-level data on palm oil production and area planted. Standard OLS estimates are unlikely to provide consistent estimates, due to possible omitted factors correlated with both household incomes and palm oil production. District fixed effects allow me to control for all time invariant factors that affect both household welfare and palm oil, such as soil type. However, there could still remain time-varying factors that are omitted from the regression and therefore cause biased estimates, such as infrastructure and rainfall, among others. There could also be reverse causality, grounded on the fact that palm oil production requires some financial investments and knowledge, implying that districts with higher average incomes could be more likely to produce palm oil. In order to tease out the causal estimate of the impact of palm oil production, an instrumental variable strategy is used. I use the historical values of palm oil production and cultivated area as my main instrument, relying on the fact that palm oil production is most likely to expand in areas where the suitable conditions and knowledge are already present. To avoid losing data, the historical values are taken from the national agricultural village survey (PODES,

[3] HPI, Human Poverty Index. This index measures deprivations in life expectancy, education, and standard of living.
[4] The poverty rate increased from 16 percent in 2005 to 17.7 percent in 2006.

Potensi Desa Agriculture Survey). In addition, I recognize the fact that not all districts are suitable for palm oil production. The relationship between palm oil production and forest degradation has been well documented. According to the FAO (2005), it is estimated that more than 56 percent of the expansion of oil palm plantations in Indonesia between 1990 and 2005 occurred at the expense of natural forest cover.[5] I use the district-level forest area prior to the study period as an alternative instrument for palm oil plantations/production. The underlying intuition is that forest area creates the potential for large oil palm plantations.

The results suggest that smallholder production in Indonesia as a whole, and also in the main production regions, Sumatra and Kalimantan, has a weak negative impact on household per capita expenditure but this effect is not present among households in rural areas. Therefore, the findings of this study indicate that there is no evidence of positive spillover effects put forward by the proponents of palm oil. With respect to health implications, smallholder area and production do not have an impact on the incidence of asthma either at the national level or in the main production regions. However, the total area and production of palm oil, i.e. including both smallholders and large plantations, increase asthma in West, South and East Kalimantan. The estimated impacts are rather small but, nevertheless material, given the scale of the expansion of the palm oil production in Indonesia.

The rest of the chapter is organized as follows. Section 4.2 provides background information on the expansion of palm oil production in Indonesia. Section 4.3 describes the data sources and descriptive statistics. Section 4.4 introduces the methodology and section 4.5 discusses the results. Section 4.6 concludes.

[5] However, proponents of the palm oil industry claim that palm oil plantations are indeed forests, and that therefore the debate over deforestation is meaningless. Nonetheless, it is generally accepted that oil palm plantations do not provide the same degree of biodiversity and level of environmental services, such as carbon storage, as do natural forests. Palm oil proponents also claim that current plantations mostly use degraded forests. However, research suggests that even degraded forests retain more biodiversity than do plantations (Gillison and Liswanti, 1999; Maddox, 2007 in Sheil et al., 2009).

4.2 Background

There is a long history to the debate over cash crops vs. staple crops, and the welfare impacts of cash crop production. For example, there are several studies on poverty alleviation and cash crop production in the context of Africa; see for example Bigsten et al., (2003) on coffee and chat in Ethiopia; and Glewwe (1991) on cocoa in Côte d'Ivoire. Moreover, rubber plantation owners acquired large plots of land in Indonesia during the 1980s and 1990s, and natural rubber was an important export product of the country (Angelsen, 1995).[6] Some of the concerns attributed to palm oil production, such as forest degradation, were already raised during the rubber boom. However, what distinguishes palm oil production from that of other cash crops is the scale of its expansion. In Indonesia, the area harvested for natural rubber was 2.9 million hectares in 2008, compared to 5-6 million hectares for palm oil (FAOSTAT). Palm oil production is also expanding to a much wider geographic area.

In this section I first discuss the facts and trends of palm oil production internationally, and then focus on smallholder production in Indonesia, specifically. Finally, I discuss the costs and benefits related to palm oil production, and the relationship between palm oil production and deforestation.

4.2.1 Palm oil production: global and local trends

Oil palm is planted for commercial purposes in over 40 countries and accounts for almost 10 per cent of the world's permanent crop land (FAOSTAT). Indonesia is currently the largest producer of palm oil in the world (19,500 thousand tonnes in 2008/2009), as

[6] In Indonesia, rubber is mainly planted and produced by smallholders; in 1998, smallholders controlled over 85 per cent of the area planted and 76 per cent of the total production (Purnamasari et al., 1999).

Indonesian production exceeded that of Malaysia in the mid-2000s. Indonesian palm oil production accounts for over 40 per cent of the world total production.[7]

The demand for palm oil has increased in the past two decades, initially for use in the chemical industry, food production and consumer goods. The soaring demand for biofuels explains the more recent boom. The largest importers of palm oil are China, India and the EU-27 bloc (USDA 2009). Palm oil has a high yield in terms of oil production; one hectare of oil palm produces 4000-5000 kg of oil, compared to 1000 kg for rapeseed, 800 kg for sunflower and 400 kg for soya bean and coconut. (Sheil et al., 2009, p. 11; 20). Currently, palm oil is the main source of vegetable oil, representing nearly 30 per cent of the world's vegetable oil production (World Bank 2010, p. 5). Besides being the world's leading exporters of palm oil, both Indonesia and Malaysia also have large domestic markets.

There has been an upward trend in the price of palm oil over the past two decades, albeit with a relatively high volatility. The world price of palm oil soared during the 2000s, peaking at US$1,146 per tonne in March 2008. However, subsequently, during the global financial crisis, the price of palm oil plummeted down to US$400 per tonne (Sheil et al., 2009, p. 19).

In Indonesia, palm oil is cultivated and produced by large private plantations (50 per cent of total production), smallholders (40 per cent), and large public plantations (10 per cent). The main production area is Sumatra, especially the provinces of North Sumatra, Riau, South Sumatra and Jambi. However, the plantations have recently been expanding eastwards; Kalimantan has become another major production area and Papua is expected to become the third major production area, due to its abundance of land (see, for example, Sheil et al., 2009; World Bank, 2010). In 2008, the region of Sumatra accounted for 78 per cent of Indonesian palm oil production, Kalimantan 18 per cent, Sulawesi 3, and Papua 1 per cent (Departementen Pertanian, 2009). During the period 1997-2007, the area devoted to oil palm cultivation grew fastest among smallholders, whose annual growth was 12 per cent, compared to 3 per cent for public plantations and 6.7 per cent for private plantations. Over this period, also the level of palm oil production grew fastest among smallholders.

---

[7] http://www.pecad.fas.usda.gov/highlights/2007/12/Indonesia_palmoil/ (accessed 28th July 2011).

However, in prior decades the fastest growth came from private plantations (World Bank, 2010).

According to the Government of Indonesia, the rapid growth in oil palm plantations will also continue in the future. Plantation areas are projected to expand from 7.4 million hectares in 2008 to 9.3 million hectares by 2015, a 25 per cent increase. Smallholders are expected to account for the largest share of this increase, although their annual growth rates will be smaller than in the past decades. According to this scenario, smallholder plantation areas will reach that of the private plantations. As land constraint limits expansion in Sumatra, future growth is projected to occur predominantly in Kalimantan and Papua (World Bank, 2010).

According to the Government statistics, there were 477 palm oil mills in Indonesia in 2006, of which majority (400) were located in Sumatra. Most of the mills are located within the plantation areas and are owned by companies and organized smallholders. There are only 57 independent mills in Indonesia which serve the independent smallholders, and these are all located in Sumatra (World Bank 2010, p.8)

Palm oil contributed 1.5-2 per cent to Indonesian GDP in the early 2000s, rising to 4.5 per cent in the late 2000s (Barlow et al. 2003).[8] It is estimated that approximately half of crude palm oil production is exported, with palm oil accounting for approximately 6 per cent of the country's non-gas export earnings (World Bank, 2010). There are no accurate statistics on employment in the palm oil industry, but according to one estimate approximately 1.2 million labourers were employed in this industry in the early 2000s (Barlow et al., 2003)[9]. In terms of the monetary value of production, palm oil is in the second place after rice of all agricultural products (FAOSTAT).

[8] The late 2000s figure is taken from: http://www.istockanalyst.com/article/viewiStockNews/articleid/3660667, accessed on 18th September 2011.

[9] According to another estimate, oil palm cultivation accounts for 1.7 million to 3 million jobs in Indonesia, and the jobs in processing come on top of these figures. Moreover, the employment effects of smallholder plantations are likely to be larger than those for large plantations. It is estimated that in smallholder plantations one person is employed in every 2 hectares, compared to approximately every 7 hectares in large plantations (World Bank, 2010).

4.2.2 Smallholder production in Indonesia

Palm oil requires tropical conditions (an average annual precipitation of 1,780-2,280mm and a temperature range of 24-30°C), and an altitude less than 600 meters. Furthermore, palm oil thrives in disturbed forests and close to rivers, and is tolerant of various different soil types (Deasy, 1942; Sheil et al., 2009). Other factors that explain the expansion of oil palm plantations in Indonesia are infrastructure and population density, among others (Angelsen, 1995; Miyamoto, 2006).[10] Improved infrastructure increases land rent, and therefore increases the incentive to expand production. Various studies have also found that smallholders in the developing world respond to changes in relative prices (see for example Godoy, 1992). In Indonesia, there is also another specific factor, related to land rights and titles, which partly explains the conversion of forests into plantations. All forest in Indonesia is *de jure* owned by the state, but according to common law clearing the forest to agricultural land gives *usufruct* rights to this land (Angelsen, 1995; Sirait, 2009).

Smallholder palm oil production has expanded rapidly over the past decade and plays an important role in total production. It is estimated that around 30 per cent of smallholder production is produced by individual farms and the remainder is by joint partnerships with large plantations (Barlow et al., 2003, p. 9). National data do not distinguish between independent smallholders and joint partnerships, but according to available estimates the recent growth can be largely attributed to independent smallholders (World Bank 2010, p. 4).[11]

One of the oldest arrangements for joint ventures between smallholders and large estates is the *Perkebunan Inti Rakyat* (PIR), which was introduced in the late 1970s by the transmigration programme.[12] PIR was initially targeted at rubber plantations, but was later

---

[10] It is notable that these studies did not try to estimate any causal relationship, i.e. whether infrastructure causes the plantation expansions, or whether plantations provide the incentive to improve infrastructure.

[11] This is particularly the case in Sumatra, where the land constraint has restricted the expansion of large plantations (World Bank, 2010, p. 8).

[12] The aim of the transmigration programme was to reallocate people from the densely populated islands of Java and Bali to the less densely populated areas of Sumatra, Kalimantan and Papua. The programme was originally introduced during the Dutch colonial era and peaked during Suharto’s regime. However, in the 2000s only a few families were relocated.

expanded to oil palm plantations. This scheme provided a good opportunity for large companies to exploit both the large land areas conceded by the Government and the abundance of low-cost labour offered by migrants relocated from other areas in Indonesia (transmigrants).[13] In some cases, land acquisitions have been accused of offering inappropriate compensation. The Government allocated land from a land category called conversion forests. However, on many occasions the land allocated was previously managed and used by local communities. Other reported problems were dependence on a single crop, deteriorated food security and limited income sources during the 4-5 year unproductive period (Vermeulen and Goad, 2006; Feintrenie et al., 2010; World Bank, 2010).[14]

The PIR scheme was later followed by the Primary Cooperative Credit for Member's scheme, *Koperasi Kredit Primer untuk Anggota* (KKPA).[15] For example, the KKPA arrangement in Bungo district, in Jambi province in Sumatra, is based on a contract signed between the company, smallholders grouped in cooperatives, and banks, under the supervision of the Government. Smallholders allocate part of their land to the company, which plants, manages and harvests the crops. This part of the land forms the *nucleus* of the plantation, and landowners are paid back a share of the harvest revenue after deducting the management costs. On the other hand, the planting costs for land that remains with the smallholders, *plasma*, have to be paid by the smallholders. However, smallholders can opt to also entrust the management of the *plasma* land to the cooperative and in turn receive a monthly rent. Smallholders organized as cooperatives have more autonomy under KKPA than under the traditional PIR arrangement although the bulk of the decision-making is still in the hands of the company. (Vermeulen and Goad, 2006; Feintrenie et al., 2010).

The advantages of the KPPA arrangement for the smallholders include access to improved seedlings and technical advice from the plantation manager. However, the drawbacks reported are similar to the PIR arrangement, such as high debt accumulated before the production period, disallowance of intercropping, dispute over land rights, and environmental damage.

---

[13] Large companies could also work without smallholders, by purchasing the land and hiring workers.
[14] By comparison, the unproductive period for rubber is 6-7 years (Feintrenie et al., 2010).
[15] Generally the joint ventures are called the Nucleus Estate Smallholder (NES) system.

4.2.3 Costs and benefits of palm oil production

Palm oil plantations and production may have relatively immediate impacts on return to labour and land. On the other hand, there may also be longer term impacts due to changes in land use, such as effects on food security and the ecosystem, among others. As stated earlier, this study focuses on the aggregate welfare effects of palm oil production, irrespective of whether a household owns an oil palm farming business. Given the scale of the palm oil expansion, it is reasonable to assume that both the benefits and costs are distributed to a wider population. Oil palm has occupied large areas of land, possibly at the expense of other crops, and this might threaten food security. Moreover, the need for a quick processing after the harvest, as well as economies of scale in mills, necessitate mills having access to large areas of mono-cropped land, preventing local people from exercising mixed livelihood strategies (World Bank, 2010).

Palm oil production may have a different impact on household welfare in areas where land is abundant, as compared to those where it is constrained.[16] The PODES 2003 Agricultural Survey could provide some insight on this. PODES (*Potensi Desa*, Village Potential Statistics) is a census of all Indonesian villages surveyed by the Central Bureau of Statistics. PODES data on land use provides information on the hectares of rice fields that have been converted to other purposes during the past three years. I define no loss in rice fields if the reported area is zero, moderate loss represents villages where less than 100 hectares have been converted and heavy loss where more than 100 hectares of rice fields have been converted to other purposes. Table 4.1 below shows the village area of oil palm plantations according to these three categories. Descriptive evidence suggests that the relationship between rice field conversion and oil palm plantations varies across regions. In Sumatra, oil palm plantations are expanding onto converted rice fields, which could be indicative of a land constraint. By contrast, in Kalimantan the largest plantations are located in villages where no conversion of rice fields has taken place.

[16] Another important question relates to how palm oil affects off-farm employment opportunities. However, I do not have appropriate data to address this question.

Table 4.1. Village oil palm plantations (in hectares) and loss in rice fields.

	No loss in rice fields	Moderate loss in rice fields	Heavy loss in rice fields	All villages
Sumatra	58.9	64.4	345.6	63
Kalimantan	60.2	15.9	35.1	55
All Indonesia	26.5	13.2	151.8	25.2

Air pollution due to land clearing and forest fires will also affect areas other than the precise burning place. The final distribution of air pollution depends on the direction and speed of the wind, but it is clear that the consequences are felt in a rather wide area (Frankenberg et al., 2005). In addition, crude palm oil production generates large amounts of waste. Refineries produce both liquid and solid waste as well as noxious odours and smoke pollution (McCarthy and Zen, 2010).

There are also several channels through which palm oil cultivation and production could benefit all households in the region. First, palm oil processing, such as mills and refineries, brings employment opportunities to the area, because the fresh fruit bunches must reach the mill within 24-48 hours of harvesting. Another possible channel is improved infrastructure that could benefit all households and industries in the region (see, for example World Bank 2010). And finally, spillover effects may be present. Proponents of palm oil defend the production expansion by promising increases in rural welfare and improved infrastructure. For example, the chairman of the Indonesia Palm Oil Association (GAPKI) argues that, "The development of oil palm plantations also plays a key role in rural development. - - With extraordinary multiplier effects, oil palm plantations in turn will become new centers of economic activities in rural areas. The development of road infrastructure provides access for isolated areas, allowing fast and dynamic economic activities".[17]

[17] Jakarta Post, 12th February 2009 (available at http://www.thejakartapost.com/news/2009/12/02/palm-oil-economic-pillar-indonesia.html, accessed 15th October 2010).

4.2.4 Deforestation

Although the relationship between oil palm plantations and deforestation has been documented in various sources (see, for example, FAO 2005), there is a degree of uncertainty related to the measures and definitions of deforestation (see, for example, Angelsen, 1995). First, the range of deforestation is wide, from a complete removal of tree cover to small changes in the ecological composition. There is no universal agreement as to what should be considered as deforestation. Second, there is an issue concerning the difference between permanent and temporary conversions. Also, the estimates for the environmental costs of conversions into plantations vary. Houghton (1993) estimates that conversion into plantations will normally result in a 30-60 per cent reduction in carbon stock in the vegetation, whereas conversion into pasture or permanently cultivated land involves a reduction of over 90 per cent. This estimate of the loss in carbon stock is similar to Tomich et al., (1998). However, there is more variance in the estimations for the loss in fauna. In addition, oil palm plantations could also follow logging in which case palm oil is not the primary cause of deforestation (see also footnote 5).

A rather crude way to look at the relationship between oil palm plantations and deforestation in Indonesia is to use the data provided by the PODES 2003 survey. The section on land use contains a breakdown of forest land (*hutan*) in hectares converted to other uses over the past three years.[18] Using this information I construct three measures of deforestation: no deforestation, moderate deforestation and heavy deforestation, following Chomitz and Griffiths, (1996). The first category includes villages that report zero hectares of converted forests (90 per cent of villages surveyed), while the second category includes villages that report positive deforestation, but of less than 100 hectares (7.8 per cent), and, finally, the third category covers villages that report more than 100 hectares of converted forests (2.2 per cent). Comparing the total palm oil plantation area in these villages, we see that plantations tend to rise with increasing levels of forest conversion (see table 4.2

[18] This variable is not a perfect measure of deforestation. First, the respondent might only have information on the village forest land, and not the state forest land. Second, the respondent might have included secondary forests or plantations in the forest category. Finally, it may be difficult to make precise area estimates.

below). However, this observation cannot confirm whether oil palm plantations are causing the deforestation.

Table 4.2. Village area of oil palm plantations (in hectares) and deforestation.

	No deforestation	Moderate deforestation	Heavy deforestation	All villages
Sumatra	55.7	54.8	344.6	63
Kalimantan	43.5	44.1	224.4	55
All Indonesia	21.1	28.4	178.1	25.2

4.3 Data and descriptive statistics

To my knowledge, there are no available large-scale household surveys that include direct questions about households' engagement in oil palm production.[19] Therefore, in this study I use annual district level (*kabupaten*) palm oil data, together with the National Socioeconomic Survey, SUSENAS. This implies that I am not able to distinguish those households that are directly supported by the industry. On the other hand, the use of the large nationally representative household survey avoids problems related to small samples, among others. Moreover, I am able to study the effects of palm oil production nationwide. Moreover, because the SUSENAS is surveyed in the beginning of the year and the data on palm oil plantations and production reflect the situation at the end of the calendar year, I use the lagged values of the palm oil.

Palm oil data (both the area planted in hectares and production in tonnes)[20] come from the Indonesian Ministry of Agriculture, Directorate General of Estate Crops (*Departementen Pertanian, Direktorat Jenderal Bina Produksi Perkebunan*).[21] As stated earlier, oil palm is cultivated by smallholders, as well as by large public and private

[19] The agricultural household survey, PATANAS might be an exception. However, the second round of the panel survey would only have been available only in late 2010, and moreover, the geographic coverage of this survey means that it would most likely only cover a few households engaged in palm oil production.

[20] Area data include immature, mature and damaged plantations.

[21] I consider that the quality of the data is largely satisfactory. However, I dropped two observations from the analysis that were likely outliers, i.e. had inconsistent growth in the production level compared to the previous year.

plantations. However, I primarily focus on smallholder production, as this may have a stronger relationship with household expenditure than production on large plantations. Another advantage of using smallholder data is that it covers a longer time period (2003-2006) than do other available production data. Moreover, the smallholder data covers the whole area of Indonesia, including over 350 districts (using the 2002 definitions of districts). Following a decentralization process, a number of new districts were created in the 2000s. In this study, 2002 is therefore taken as a base year.

According to the national district-level data, the average area of smallholder plantations is 5,800 hectares and the average production is around 11,500 tonnes. Between 2003 and 2006, the average district smallholder oil palm plantation area in Indonesia increased by approximately 50 per cent, from 4,500 hectares to 6,800 hectares. However, approximately 67 per cent of the districts in Indonesia do not have any smallholder oil palm plantations. Restricting the sample to only those districts that do have smallholder plantations, the average district area of smallholder plantations is 18,100 hectares and the average production is around 41,400 tonnes, respectively. However, if we focus on the main production regions, in Sumatra only 28 per cent of the districts do not have any smallholder plantations, while the corresponding figure in Kalimantan is 31 per cent. The average district smallholder plantation area in those districts that do have smallholder plantations is 22,600 hectares in Sumatra and 11,000 hectares in Kalimantan. Table 4.3 below presents summary statistics on smallholder area and smallholder production for all districts in Sumatran and Kalimantan irrespective of whether the districts have smallholder plantations.

Table 4.3. Descriptive statistics for district-level smallholder data in Sumatra and Kalimantan, years 2003-2006.

	Average (2003-2006)	2003	2006
Smallholder area (ha)	13,300	10,200	15,700
Smallholder production (tonnes)	26,500	19,600	35,500

Nationwide, smallholder oil palm plantations represent, on average, one per cent of the district land area, but there are large differences across the regions. In Sumatra, there are districts where up to 15 per cent of the area is covered by smallholder plantations, while in Kalimantan, the proportion varies from 0 to 5 per cent.

In addition to the smallholder data I employ the complete production data (both smallholders and large plantations) for selected provinces in Kalimantan.[22] The area data cover the provinces of West Kalimantan, South Kalimantan and East Kalimantan (but not Central Kalimantan), and the production data cover the provinces of West Kalimantan and South Kalimantan (but not Central Kalimantan and East Kalimantan). Both the production data and the area data cover the time period 2004-2007, although data for the year 2004 is only available for West Kalimantan.[23,24] In these provinces the average district area devoted for oil palm plantations is 34,000 hectares when focusing only on those districts that do have plantations. Table 4.4 below presents summary statistics on total area of plantations and total production for selected provinces in Kalimantan irrespective of whether the districts have oil palm plantations.

Table 4.4. Descriptive statistics for district-level palm oil data for selected provinces in Kalimantan.

	Average (2005-2007)	2005	2007
Total area (ha)	25,700	21,100	30,700
Total production (tonnes)	49,200	41,600	55,200

Notes: Area data include provinces of West, South and East Kalimantan, and production data include provinces of West and South Kalimantan. For West Kalimantan data are also available for 2004 that are included in the regression analysis.

[22] Area data cover immature, mature and damaged plantations.
[23] Due to the inability to get comparable household expenditure data for the year 2008, the year 2007 is only used in health specifications.
[24] I also have data on complete production (both smallholders and large plantations) for the whole of Indonesia for the years 2005 and 2008. However, due to an inability to get household data for 2009 the use of these data will be left for future research.

The annually implemented SUSENAS survey is a nationally representative household survey. Each year, in late February or early March, a new set of roughly 200,000 households are interviewed as part of the core of the national socio-economic census (SUSENAS). The dataset includes results from a small consumption module, consisting of 15 food items and 8 non-food items, that combines purchased and own-produced items. The household per capita expenditure is deflated to 2007 prices using the consumer price index of the province capital. In addition there is data on household characteristics, such as education and health status of the household members, and housing conditions. The average log monthly per capita expenditure is 12.57 over the surveys, equivalent of 360,000 Indonesian rupiahs, or US$40. Very few household members report symptoms of asthma; only 1.7 per cent of the individuals aged 10 and above reported suffering from asthma over the preceding month, but the incidence of asthma increased from 1.4 per cent to 2.8 per cent over the survey period.

In addition to the core datasets, two other data sets are employed in order to construct the instruments for the IV estimation. First, PODES data is used to construct a historical measure of palm oil production in a district. The PODES data come from a survey of over 65,000 villages throughout Indonesia, which rotates themes such as agriculture, economic and population. The 2003 PODES Agriculture Survey includes a section on village-level plantation crop production, and both the area and production of the five most important plantation crops are listed here.[25] It is notable that the 2003 PODES survey was implemented in 2002 and therefore the production data refer to the year 2002. This is important as my study period starts in 2003. I aggregate palm oil production levels and oil palm plantation areas in all villages in a district, in order to construct a historical district-level measure of palm oil production and plantation.[26] Importantly, the 2003 PODES survey covers the whole geographic area of Indonesia.

Satellite data on district forest cover is used to construct an alternative instrument for oil palm plantations and palm oil production. The proportion of district area covered by forest

[25] As there are not too many plantation crops my judgement is that if oil palm is not listed among the five most important crops then the cultivation of oil palm is likely to be only a marginal activity.

[26] It is not stated specifically whether the area and production refer to smallholders or to total production. However, as the village head would probably not have access to data on area and production for private companies, it is therefore likely that this measure best refers to smallholders.

was provided by the Geographic Information Science Center of Excellence, South Dakota State University (see Broich et al., 2010 for further information). The district area in square kilometres was then employed in order to calculate the forest area of a district in 2000. As discussed earlier, large areas of forest have been converted into oil palm plantations and, therefore, district forest area prior to the study period is likely to be a good source of exogenous variation in palm oil production.[27] Because the forest area is calculated prior to the study period, the possible direct correlation between household welfare and forest cover is eliminated. However, despite of these advantages, there are some disadvantages related to this instrument. First, forest data used in this study is not pure, in that it also covers mature plantations. The satellite data employ a definition of forest as areas where there is a tree canopy exceeding 25 per cent coverage, and greater than five meters in height. Unfortunately, when using this definition, it is not possible to distinguish between natural forest and mature oil palm plantations. Second, data on forest cover are only available for Sumatra and Kalimantan. However, as Sumatra and Kalimantan are the main production regions, and together account for approximately 95 per cent of the national production I argue that this limitation is unlikely to bias the results to any significant extent.

Finally, I have monthly rainfall data covering the period of 1951-2007. This data is taken from two sources. Data for the period 1951-1998 was provided by Kirono et al. (1999), while that for the period of 1999-2007 come from the Indonesian Meteorological Agency (BMG). These data are used in robustness checks in sections 4.5.1 and 4.5.2.

4.4 Empirical strategy and identification

Using the time series oil palm plantation and palm oil production data as well as SUSENAS surveys, the impact of palm oil production expansion may be expressed as follows:

[27] Strictly speaking, large forest areas could also be used for other plantation crops, such as rubber, and therefore the effect of this IV specification could be a general plantation crop effect, not only specific to palm oil. However, no other plantation crop has expanded as aggressively as oil palm.

$$I_{idt} = \beta_1 + \beta_2 PO_{dt-1} + \beta_3 X_{idt} + D + \mu_{pt} + T + \varepsilon_{idt}, \tag{4.1}$$

where I_{idt} is the selected impact (expenditure, health) of household/individual i living in district d in year t, PO is the level of palm oil production in tonnes or the area of oil palm plantations in hectares in district d at time $t-1$, D a district fixed effect, μ_{pt} a province-year interaction term to control for any annual province-level shock affecting household expenditure or health, T year fixed effect, and ε_{idt} is the error term, clustered at districts. Because the SUSENAS survey takes place in the beginning of the year, while values for oil palm plantations and palm oil production levels reflect the situation at the end of the calendar year, I use the lagged values of the palm oil. X is a vector of household or individual characteristics such as household head's education level, age, gender, industry category, and occupation type, as well as household size in the household expenditure specification. In the health specification I control for the respondent's gender, age, education and industry category as well as some household characteristics related to the housing conditions that could also affect health status, such as dummy variables indicating whether household has its own toilet or uses tap water, or whether the dwelling is owned by household. In addition, I include a dummy for rural areas in both specifications.

The OLS specifications serve as benchmark estimates but could nonetheless be biased. First, there could be omitted variables that are correlated both with palm oil production and household welfare. The district fixed effects control for all time-invariant factors that affect both household welfare and palm oil production, such as soil type. However, there could still remain time-varying factors that are omitted from the regression and resulting, therefore, in biased estimates, such as infrastructure and rainfall, among others.[28] Also

[28] However, it is likely that the province-year interaction term captures the effects to some extent. Nevertheless, in sections 4.5.1 and 4.5.2 I implement some robustness checks where I include rainfall as an additional control.

some time-invariant factors such as soil type could have both level and trend affects.[29] Second, palm oil production could be endogenous in this context. Palm oil production requires large and expensive investments, such as roads and mills, and therefore districts with high average incomes are more likely to be engaged in palm oil production. However, it is notable that the dependent variable is at household or individual level and palm oil is measured at district level. Therefore to the extent that there is any reverse causality, it is likely to be weak. Third, there could be positive sorting; that is palm oil could attract wealthier households into the region.[30] And finally, both oil palm plantations and total production are likely measured with error.

Health status, in this case the probability of reporting symptoms of asthma, is measured at individual level. Therefore the problem of reverse causality is less likely to be a problem in the health specification compared to the expenditure specification. However, there could still be omitted factors correlated both with the palm oil production and health status.

One method that addresses all four possible problems related to the OLS estimation is the Instrumental variable approach, IV. Finding instruments for palm oil is, however, not straightforward. Many factors determining palm oil production (including rainfall, among others) could also be correlated with household welfare thorough other mechanisms than palm oil (say rice production). According to the land suitability assessments the requirements for cultivation of oil palm are not exclusive, in the sense that these areas are typically also suitable for rubber cultivation, among other crops (Ritung et al., 2007). As regards other geographical variables, altitude could be a potential instrument given that oil palm thrives at altitudes below 600 metres (see section 4.2.2). However, altitude did not have enough power as an instrument.[31]

---

[29] Strictly speaking extensive cultivation could exhaust the land and alter the soil type but in this paper I assume that soil type is time-invariant during the study period. This assumption is reasonable given the relatively short time period.

[30] Future research could address the relationship between palm oil production and migration. However, by controlling for the education level I am able to reduce the bias related to possible sorting.

[31] One plausible explanation for this may be that the relationship between the district altitude, on the one hand, and its palm oil production levels and oil palm plantations, on the other, is weak, in the sense that there is no clear cut-off point, but rather a gradual decline as altitude increases. Another possible explanation is that the data used in this study to calculate the proportion of district area that falls into four different elevation categories is not optimal. The 2003 PODES village survey has information on the altitude of each village,

In order to implement the IV estimation, I will construct a separate prediction model to predict the production of palm oil in a district, and then use the predicted values as instruments for actual values in the two-stage least squares estimation.[32] In the prediction model I will exploit the fact that not all districts are suitable for oil palm plantations. The principal idea is to include variables exogenous to household welfare in the prediction model, in order to generate exogenous variation in the predicted values. I have identified two potential sources of exogenous variation in the palm oil production and the relative merits of the instruments are discussed as follows.

First, lagged values of palm oil production could predict future production. The use of lagged values of an endogenous variable as an instrument is a standard method in the literature (see, for example, Jalan and Ravallion, 1999). The use of lagged values as instruments is based on the idea that areas suitable for oil palm cultivation are also likely to have high production levels in the future. Oil palm plantations also require special knowledge and skills, providing another reason why lagged values have good prediction power. However, using lagged values as instruments implies losing one round of data. As I already have a relatively short period of data, it would be preferable to not lose any data. Therefore, in order to avoid losing any data, the historical values of palm oil production are taken from the 2003 PODES survey.[33] This agricultural survey contains village-level information on the planted area of oil palm and level of production of palm oil throughout Indonesia. I aggregate these village-level data at the district level, which provides an approximate measure of district level data in 2002. These district-level measures of oil palm plantations and palm oil production in 2002 are time-invariant variables. Therefore they are interacted with predicted province-level palm oil area or production in order to predict the area of plantations and palm oil production in each district, this method to some extent follows the method by Duflo and Pande (2007):

---

which I then used to calculate the proportion of villages in each elevation category in the given district. This method relies on the assumption that the villages are of equal size and spread evenly across the district.

[32] Duflo and Pande (2007) use river gradient in order to predict the number of dams per district and then use the predicted number of dams in the district as an instrument for actual number of dams. Another paper that that use similar estimation strategy, that is, using predicted values as instrument for actual values, is by Saiz (2007).

[33] The 2003 Podes agricultural data uses 2002 data.

$$PO_{dt} = \alpha_1 + \alpha_2\left(PO2002_d * \overline{PO}_{pt}\right) + D + \mu_{pt} + T + \varepsilon_{dpt,} \quad (4.2)$$

where PO_{dt} is the palm oil production level (in tonnes) or oil palm cultivated area in hectares in district d at time t, $PO2002$ is the district palm oil measure in 2002, and $\overline{PO_{pt}}$ is the predicted palm oil production in province p at time t. This is constructed by multiplying the total production of palm oil in Indonesia in the given year with the proportion of production in the given province in 2003. In the area specification I use the predicted value of oil palm plantation, obtained by multiplying the total oil palm plantation area in Indonesia in the given year by the proportion of plantations in the given province in 2003. The use of predicted, rather than actual values of province production levels and plantation areas, ensures that the palm oil production is exogenous with the district palm oil production. D is the district fixed effect and T year fixed effect and μ_{pt} is the province-year interaction term included to account for any annual shocks that are common across districts in a province and that might affect palm oil production.

As expected, historical areas of plantations and levels of production are positively associated with current plantations and production. This finding reflects the fact that palm oil is expanding most strongly in areas that have been identified as suitable areas for planting oil palm and where there is appropriate knowledge and knowhow easily available. The F-test for historical palm oil variable interacted with predicted province measure of palm oil is 7.45-10.47, depending on the specification, implying that the chosen instruments have sufficient prediction power (see table 4.5 below). The prediction model for the total cultivated areas and production levels in Kalimantan is presented in table C6 in the Appendix (see columns 1 and 2).

Table 4.5. Prediction model for smallholder cultivation areas and production levels using the historical data for these measures as an exogenous source of variation.

	Area, All Indonesia	Area, Sumatra and Kalimantan	Production, All Indonesia	Production, Sumatra and Kalimantan
Oil palm area in 2002	0.0009***	0.0009***		
	(0.0003)	(0.0003)		
Palm oil production in 2002			0.0002***	0.0002***
			(0.0001)	(0.0001)
Observations	1863	788	1863	788
F-test for instrument	10.47***	10.46***	7.46***	7.45***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*year interactions	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Cultivation area is expressed in hectares and palm oil production in tones. Oil palm area is interacted with predicted province area and palm oil production interacted with predicted province production.
Regression coefficients are multiplied by 1000.
* p<0.1, ** p<0.05, *** p<0.01

There are some deficiencies related to this IV approach, however, as the historical values of plantation areas and production levels may still be correlated with the error term. If there are omitted variables that are serially correlated, the use of historical values might introduce bias into the IV estimates, and thus undermine the validity of the instrument (see, for example, Angrist and Krueger, 2001).[34] Therefore, I will also introduce an alternative instrument, district forest cover. As discussed earlier, the conversion of natural forest into oil palm plantations has been widely documented. However, the presence of forests could affect household expenditure and welfare in various ways, such as by providing firewood and other forestry products (such as natural rubber), or by providing hunting opportunities, among others (Angelsen, 1995). Therefore, to ensure that the correlation between forest area and household welfare realizes through palm oil, and not through these other mechanisms, data on forest area prior to the study period is used. I have used satellite data on the district forest cover (per cent) for the year 2000, together with data on district area (square kilometres), to calculate the district forest area in 2000. The district forest area may

[34] A potential serially correlated omitted variable is infrastructure.

be used as an indicator of the opportunity to plant oil palm that is exogenous to household characteristics such as knowledge and income.

One limitation of using the forest data is that it is not possible to distinguish mature plantations, including oil palm plantations, from natural forest, which introduces some noise to the forest measure. However, I argue that this is not a major issue, given that oil palm plantations are not a major contributor to the total forest area. Unfortunately, I do not have district level data on total oil palm plantation area (i.e. for both large plantations and smallholders) in 2000. However, province level data shows that the proportion of oil palm plantations out of total province forest cover is, on average, 11.2 per cent in Sumatra and 1.4 per cent in Kalimantan (the district forest cover data is only available for Sumatra and Kalimantan).

The prediction model using the forest area as an instrument can be expressed as follows:

$$PO_{dt} = \alpha_3 + \alpha_4\left(forest * \overline{PO}_{pt}\right) + \alpha_5(forest * t_t) + D + \mu_{pt} + T + \varepsilon_{dpt}, \qquad (4.3)$$

where PO_{dt} is the palm oil production (in tonnes) or plantation area in hectares in district d at time t, $forest$ is the forest area in a district (in square kilometers) in year 2000, $\overline{PO}_{pt}$ is the predicted palm oil production or palm oil plantation area in province p at time t as explained earlier. The interaction term between the forest area and year dummies allows for a varying impact of forest area on palm oil. This is to say, that the conversion rate could be different across the years. D represents district fixed effects, T year dummies while μ_{pt} is the province-year interaction term as before.

Table 4.6 below shows that, as expected, forest area in 2000 is positively correlated with palm oil area and production in subsequent years. The corresponding F-test is approximately 11.5 in the oil palm area specification and 17 in the production specification.

Prediction model for total area and production in Kalimantan using district forest area as an instrument is presented in the Appendix, in table C6 (see columns 3 and 4).

Table 4.6. Prediction model for smallholder area/production using district forest area in 2000 as an exogenous source of variation.

	Oil palm area, Sumatra and Kalimantan	Palm oil production, Sumatra and Kalimantan
Forest area in 2000	0.0206***	
	(0.0061)	
Forest area in 2000		0.0109***
		(0.0026)
Observations	773	773
F-test for instrument	11.54***	17.32***
District fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Province*year interactions	Yes	Yes

Robust standard errors in parenthesis
Oil palm area is expressed in hectares and palm oil production in tonnes. In the oil palm area model district forest area in 2000 is interacted with the predicted province area of oil palm, and in the production model district forest area is interacted with the predicted province palm oil production.
Specifications also include forest area*year interaction terms.
Regression coefficients are multiplied by 1000.
* p<0.1, ** p<0.05, *** p<0.01

I will estimate equations (4.2) and (4.3) in order to predict the cultivated area of oil palm and the level of production in each district. The predicted values will then be used as instruments for actual values, following Duflo and Pande, (2007) and Saiz (2007). The first stage regression for the specification with historical values of palm oil used as an instrument is as follows:

$$PO_{idt} = \theta_1 + \theta_2 \widetilde{PO}1_{dt} + \theta_3 X_{idt} + D + T + \varepsilon_{idt}. \quad (4.4)$$

And the first stage regression using forest area as an exogenous variation in palm oil is as follows:

$$PO_{idt} = \theta_1 + \theta_2 \widetilde{PO}2_{dt} + \theta_3 X_{idt} + D + T + \varepsilon_{idt}. \tag{4.5}$$

Finally I will estimate equation (4.1) with 2SLS, using fitted values from the first stage regressions, (4.4) and (4.5). In all regressions standard errors are clustered on a district level. [35]

4.5 Empirical results

In this section I discuss the empirical results of both the OLS specification and the IV specifications. Generally the estimated coefficients are consistent with the theory and previous literature. For example, in the smallholder area specification for Indonesia as a whole household per capita expenditure increases with the age of the head (but at a decreasing rate) and with the education of the head (see table C7 in the Appendix). On the other hand, female-headed households are poorer compared to male-headed ones. Households whose head works in mining, manufacturing, construction, electricity, wholesale, transportation, finance or in the public sectors have a higher per capita expenditure level compared to households whose head works in agriculture. Similarly, households whose heads are employers or employees have higher expenditure levels, while those whose heads are casual workers have smaller expenditure levels than do those whose heads are self employed. Somewhat surprisingly, the dummy variable for rural areas is statistically insignificant.

In the health status specification the probability of an individual reporting symptoms of asthma is greater in rural areas (see table C13 in the Appendix). However, there is no obvious explanation why this should be the case. Traffic pollution is likely to be worse in urban areas; on the other hand many activities related to agriculture may be correlated with asthma in rural areas. Most importantly, the positive sign of the rural coefficient is most likely a result of indoor pollution due to wood and other biomass being used as a cooking

[35] No attempt was made to use sample weights.

and heating fuel (Osterman and Brauer, 2001).[36] The probability of an individual reporting symptoms of asthma decreases with education, although education is most likely a proxy for income in this application, and also with age. Difficulties in breathing are less common among females, as well as among individuals who live in dwellings that are owned by the household and have their own toilet facilities. Next, I discuss the impact of the main independent variables, palm oil production levels and area of oil palm plantations, on household expenditure and individual health.

4.5.1 Smallholder area and production

Per capita expenditure. The first set of results relates to the impact of smallholder oil palm plantation area and smallholder palm oil production on household per capita expenditure. I first discuss the results for smallholder area. Estimates derived with the OLS suggest that smallholder area is negatively associated with household expenditure in Indonesia as a whole but the estimated coefficient is not statistically significant at the conventional levels (see column 1 in table C7 in Appendix). The IV estimates give very similar results to the OLS ones (see column 2 in table C7). There is no notable difference when restricting the sample to rural households only (see columns 3 and 4 of the C7).

OLS regressions give rather similar results when restricting the sample to the main production regions, Sumatra and Kalimantan. The summary of the main results for Sumatra and Kalimantan is presented in table 4.7 below (for further details, see table C9 and C10 in Appendix). The estimated effect of smallholder area is negative but statistically insignificant at the conventional levels. However, in this sample the IV estimate using historical forest area as an instrument indicates a negative effect (see column 3 in table C9). The estimated effect suggest that a thousand hectares increase in the district-level smallholder plantations decreases household expenditure by 0.45 of a percentage point and the estimated effect is marginally statistically significant at the 10% level. The

[36] In this setting respirable particulate levels could be 10-50 times higher compared in urban areas.

corresponding elasticity is 0.066.[37] The average size of the district smallholder area in Sumatra and Kalimantan is 13,300 hectares over the survey period, so an increase of thousand hectares is equal to six percent. It is notable that the IV using historical plantations as an instrument reveal a negative but statistically insignificant effect (see column 2 of table C9). In both IV estimations the first stage F-test is acceptable albeit higher in the specification when using historical values of plantation area as an instrument. Moreover, the IV estimation using historical forest cover as an instrument suggest a weak negative impact of smallholder area also in the rural areas. However, now the first stage F-test is low (4.18) and I therefore consider it is reasonable to conclude that the negative impact is not present among households living in rural areas (see table C10 in the Appendix).

There could be several factors underlying the negative relationship. For example, oil palm plantations may have been developed at the expense of crucial subsistence crops. There is also evidence of unfair agreements between landholders and palm oil producing companies, resulting in high debt burdens for farmers.

Next I will discuss the results for smallholder production. Using the data for Indonesia as a whole, OLS estimation results suggest that, similarly to smallholder plantation area, smallholder production has a negative but statistically insignificant impact on household per capita expenditure (see table C8 in Appendix, column 1). However, the IV estimate suggest a weak negative impact of smallholder production in all Indonesia, but again, the negative impact is not present among rural households (see table C8 in the appendix, columns 2 and 4). The estimated effect implies that a thousand tonnes increase in the district-level smallholder production decreases household per capita expenditure by 0.09 of a percentage point, while the corresponding elasticity is -0.01. It is notable that given the wide geographic area of Indonesia and the heterogeneity of livelihood strategies, as well as wide differences in standards of living across the country, pooling the nationwide data might cause problems in terms of comparability.

Similar results are obtained when restricting the sample to Sumatra and Kalimantan only. OLS regressions suggest a negative, but statistically insignificant, impact of

[37] The elasticity is calculated using the form: $\beta * \bar{x}$.

smallholder production on household per capita expenditure for households in Sumatra and Kalimantan (table C11 in Appendix, column 1). However, now both IV estimates suggest that smallholder production decreases household expenditure in Sumatra and Kalimantan and the estimates are marginally statistically significant at the 10% level (see Appendix table C11, columns 2 and 3). The estimated effects imply that a thousand tonnes increase in the district-level smallholder production decreases household per capita expenditure by 0.09 of a percentage point (using historical production from PODES 2003 as an instrument) or 0.24 of a percentage point (using district forest area as an instrument).[38] The corresponding elasticities are -0.024 and -0.068 respectively. Although these elasticities are rather low, the scale of the expansion suggests that the economic impact is moderate. For example, in Sumatra and Kalimantan the average district-level smallholder production increased around 16,000 tonnes between 2003 and 2006 (from around 19,600 tonnes to around 35,500 tonnes). Employing the estimated effects above, household per capita expenditure fell approximately 1.44 per cent (using historical production from PODES 2003 as an instrument) or 3.84 per cent (using district forest area as an instrument) over the study period resulting from the increase in the level of district smallholder production. Interestingly, restricting the sample to rural households only, the negative effect is no longer statistically significant (see table C12 in the Appendix).

Finally, it is notable that generally the coefficients in the IV regressions are larger in absolute terms than the OLS equivalents suggesting that the OLS coefficients are biased upwards. This is in agreement with prior expectations because, as discussed earlier, palm oil production requires large investments and knowledge.

Robustness check. As a robustness check I include the deviation of annual rainfall from its historical mean (for the period 1951-2007) as an additional regressor to the specifications evaluating the impact of smallholder area and smallholder production on per capita expenditure. For the sake of brevity I only present results for the specifications where my main results are obtained (see tables C17 and C18 in the Appendix). I use lagged values of rainfall because the SUSENAS survey is implemented at the beginning of the calendar year when the realization of the rainfall is unknown. Somewhat counterintuitively, the estimated

[38] The average level of district smallholder production in Sumatra and Kalimantan is around 26,000 tonnes over the survey period and therefore 1000 tonnes are equal to approximately four percent.

coefficient of the rainfall variable is negative. However, it is only statistically significant in some specifications. The results for my main variable, palm oil, change little and importantly the magnitude of the smallholder area and smallholder production coefficients remain the same. However, the only notable difference is that the IV estimate of smallholder production for Sumatra and Kalimantan using historical values of production as exogenous source of variation is now only statistically significant at the 10.6% level (previously 10% level).[39]

Health outcomes. The impact of palm oil expansion on health outcomes is measured by studying the probability of a household member reporting symptoms of asthma. Asthma and difficulties in breathing could be associated with forest fires and toxic odours as well as waste coming being released by refineries, although there could also be other relevant health indicators. Using the data for Indonesia as a whole the results from both the OLS regression and IV regression suggest there is no statistically significant relationship between smallholder area and the prevalence of asthma (see table C13 in Appendix).[40] The same holds also with smallholder production (see table C14 in Appendix). The estimated effects are always negative, but fall short of statistical significance at conventional levels.

Similarly, results from restricting the sample to Sumatra and Kalimantan only suggest that neither smallholder production nor smallholder area has a statistically significant impact on the incidence of asthma (see tables C15 and C16 in Appendix).[41]

---

[39] Because the sign of the estimated rainfall coefficient is not in line with my prior expectations the rainfall variable is excluded from my main specifications.

[40] It is notable that the individual data for all Indonesia is too large for estimation and therefore the data used in this study covers a 30% sample of individuals aged 10 and above, stratified by year and district.

[41] I also divided the sample into adults and children but could not find any statistically significant effect.

Table 4.7. The impact of smallholder area and smallholder production on log household per capita expenditure and asthma in Sumatra and Kalimantan; summary of the main results.

Dependent variable	LOG PCE		ASTHMA	
	All households	Rural households	All households	Rural households
		OLS		
Smallholder area	-0.0010	-0.0009	-0.000043	-0.000090
	(0.0009)	(0.0009)	(0.000044)	(0.000064)
Smallholder production	-0.0004	-0.0002	-0.000010	-0.000025
	(0.0004)	(0.0003)	(0.000019)	(0.000028)
		IV		
Smallholder area	-0.0016	-0.0015	-0.000112	-0.000264
	(0.0011)	(0.0010)	(0.000121)	(0.000208)
Smallholder production	-0.0009*	-0.0006	-0.000050	0.000101
	(0.0005)	(0.0004)	(0.000060)	(0.000096)

In the IV estimation historical values of smallholder production/smallholder area from PODES agricultural survey are used. Regression coefficients are multiplied by 1000 and therefore the estimated effect is for a 1000 ha increase in the area specifications and a 1000 tonnes increase in the production specifications.

4.5.2 Total production and area in Kalimantan

The second set of results relates to the total production and plantation area (including both smallholders and large plantations) in Kalimantan only. A summary of the main results is presented in table 4.8 below. There are some limitations in the data with respect to Central Kalimantan, for which the area and production data are unavailable, and also East Kalimantan where the production data is reported in fresh fruit bunches (FFB), not crude palm oil in tonnes. Therefore East Kalimantan is excluded from the production specifications. Given these limitations, results cannot be generalized to whole area of Kalimantan.[42] Importantly, the number of districts in the total production specifications is below 30 and the model test statistics and standard errors should therefore be interpreted with some caution.[43]

[42] It is noted that West Kalimantan is the main palm oil producing province in Kalimantan and it is included in the analysis. Moreover, due to the data available for this study I am only able to capture short-term effects.
[43] The cluster robust variance-covariance matrix is asymptotically consistent in the number of clusters. Ideally the number of clusters would therefore be relatively large.

Per capita expenditure. OLS regressions on the impact of total area and total production of palm oil give similar results to those for smallholders only. According to the OLS specifications both the total district area of palm oil and total district production of palm oil are negatively associated with household per capita expenditure but the estimates are not statistically significant at conventional levels (see tables C19 and C20 in Appendix). The instruments do not have enough power to assess the causal relationship between the total area of oil palm and household expenditure. For example, using the historical values of oil palm area as exogenous source of variation the F-test of the first stage (2SLS regression) is only 1.57. The instruments work better in the total production specifications, although the model test statistics should be interpreted with caution. Recognising this, the IV estimate is consistent in that there is no statistically significant relationship between total palm oil production and household expenditure.

Restricting the sample to rural households only does not change the overall finding that the total area and production of palm oil are only weakly associated with household expenditure, and the estimated coefficients do not gain statistical significance at conventional levels.

Health outcomes. The final set of results relates to the impact of total area and production of palm oil in Kalimantan on the probability of individuals reporting symptoms of asthma. The OLS estimation suggests that there is a positive relationship between the total area of oil palm and the presence of breathing difficulties but the estimated coefficient is not statistically significant at the conventional levels. The IV estimation using the historical values for plantation areas as an exogenous source of variation in the prediction model confirms the positive relationship between total plantation area the incidence of asthma and now the estimated effect is statistically significant at the 5% level (see table C21 in the Appendix, columns 1 and 2). The result is also robust to the inclusion of the rainfall variable (results not reported here). The estimated impact suggest that one thousand hectares increase in the total area of district oil palm plantations increases the probability of individual reporting breathing difficulties by 0.09 of a percentage point. The corresponding elasticity is 1.50.[44] The estimated effect for individuals living in rural areas is similar but now the effect is only marginally statistically significant at the 10% level. Also the F-test of the first stage is smaller 6.17, compared to 13.01 in the full sample (see table C21 in the Appendix, column 4).

Importantly, in focusing on the total palm oil production, the OLS estimates also suggest that total production is positively associated with incidence of asthma; the estimated coefficient is statistically significant at the 10% level. Moreover, the size of the coefficient is similar to that for the total area (see table C22 in the Appendix, column 1). The OLS estimate is not statistically significant when restricting the sample to rural households only (see table C22 in the Appendix, column 3). My IV estimates produce fairly similar results. Using the historical values from the PODES survey as an exogenous source of variation suggests that a thousand tonnes increases in the total production of palm oil increases the probability of an individual reporting symptoms of asthma by 0.01 of a percentage point and the effect is statistically significant at the 1% level (table C22 in the Appendix, column 2).[45] The corresponding elasticity is 0.44. The average district-level total production of palm oil in South and West Kalimantan increased around 33% between 2005 and 2007. Employing the elasticity listed above the results suggest that the prevalence of asthma increased by 14.5% in West and South Kalimantan resulting of the expansion in palm oil production. The estimated effect is also positive and statistically significant when restricting the sample only to those individuals living in rural areas (see table C22 in Appendix, column 4). However, as noted earlier the model test statistics and standard errors should be interpreted with caution in the IV specifications for total production.

---

[44] Elasticity is calculated: $\beta * \frac{\bar{x}}{\bar{Y}}$

[45] Both the OLS and the IV results are robust to the inclusion of the rainfall variable.

Table 4.8. The impact of total area and total production on log household per capita expenditure and asthma in Kalimantan, summary of the main results.

Dependent variable	LOG PCE		ASTHMA	
	All households	Rural households	All households	Rural households
			OLS	
Total area	0.0081	-0.0048	0.0002	0.0002
	(0.0060)	(0.0064)	(0.0001)	(0.0001)
Total production	-0.0007	-0.0006	0.000068*	0.000076
	(0.0009)	(0.0007)	(0.000035)	(0.000046)
			IV	
Total area	-0.0015	0.0053	0.0009**	0.0009*
	(0.0215)	(0.0204)	(0.0004)	(000006)
Total production	-0.0002	0.0000	0.000143***	0.000125**
	(0.0008)	(0.0006)	(0.000045)	(0.000051)

In the IV estimation historical values of smallholder production/smallholder area from PODES agricultural survey are used. Regression coefficients are multiplied by 1000 and therefore the estimated effect is for a 1000 ha increase in the area specifications and a 1000 tonnes increase in the production specifications. In the total area specifications, the provinces of South, West and East Kalimantan are included; in the total production specifications, the provinces of South and West Kalimantan are included.

4.6 Conclusion

In Indonesia, the area of oil palm plantations has increased by more than 2100 per cent since the 1980s. This expansion has been shown to have significant impacts on the environment, and particularly on forest degradation. There is also a general perception that palm oil production is an important driver of economic development in rural areas. However, there is little systematic empirical evidence on the welfare impacts of this expansion. There are a few small scale, location-specific studies looking at the effects of palm oil production on smallholder producers (see for example Feintrenie et al., 2010) and a descriptive study comparing province level outcomes (Kessler et al., 2007), but the aggregate welfare effects remain largely unknown. This study attempts to fill this gap.

In this chapter I have studied the welfare impacts of district-level areas of oil palm plantations and levels of palm oil production on households located in these districts. I have discussed that palm oil production requires investments and knowledge, and therefore that the possible endogeneity problem must be addressed in any credible evaluation. It is not straightforward to find a suitable instrument for palm oil. However, I argue, that given the importance and relevance of the topic, the OLS estimation combined with even suboptimal IV estimation will bring a notable contribution to the discussion. I have identified two sources of potentially exogenous variation in palm oil production. First, I use historical values of palm oil area and production, relying on the common assumption that lagged or historical values of an endogenous variable could be used as an instrument for current values. Second, I use district-level forest area data in the period prior to my study to predict the oil palm plantations and production levels in the study period. Generally these two sets of instruments provide similar results, which gives support to the chosen estimation strategies.[46] My primary focus is on smallholder production, both in Indonesia as a whole and separately in the two main production regions, Sumatra and Kalimantan. In addition, I study the welfare impact of total production in selected provinces in Kalimantan.

Two main findings emerge. First, the suggested positive spillover effects of palm oil production are not found in these data. The proponents of palm oil have argued that palm oil production could be the main driver of rural development in Indonesia. The findings of this study suggest that palm oil production has, if anything, a negative impact on household per capita expenditure both in Indonesia as a whole and in the main production regions, Sumatra and Kalimantan. Even if the negative effect is not present when the sample is restricted to rural households, I can rule out any positive impact. This finding is in contrast with the anecdotal evidence, which usually highlights a positive impact of palm oil production on farmers in the study villages (Feintrenie et al., 2010; Rist et al., 2010). Therefore the findings here are more in line with Kessler et al. (2007), who find mixed results on the socio-economic impacts of the palm oil production expansion. In addition, there is also evidence to suggest that some farmers might engage in unfair deals with the palm oil producing companies, in which case these farmers' debt may increase to unsustainable levels. It is also notable that previous studies have focused only on a few villages and therefore the results are difficult to generalize. However, it is important to note that the findings in this study do not suggest that individual farmers could not benefit from palm oil production. Rather, the results suggest that even if farmers had indeed benefitted

[46] Ideally the two instruments would be uncorrelated which is not the case in the current application, however.

from palm oil production, there is no evidence that these benefits would have been distributed to other households, at least in the short term.

The estimated elasticity appears low, but taking into account that the effect is estimated for all households in the district, not just the households directly supported by the industry, the estimated effect is moderate. Also the scale of the expansion suggests that the economic impacts are significant. For example, in Sumatra and Kalimantan the average district smallholder production increased over 80 per cent between 2003 and 2006.

Similarly to the findings regarding the impact of the smallholder production and area in Sumatra and Kalimantan, I have not found any evidence of a positive impact of total oil palm area or total palm oil production on household expenditure in selected provinces in Kalimantan. The OLS estimates point to a negative association, but the estimated coefficients fall short of statistical significance at the conventional levels. However, due to the lack of power of my instruments in the IV, estimation I cannot assess the causal relationship between total area and household expenditure in Kalimantan.

With regard to health impacts, smallholder area and smallholder production do not have an impact on the incidence of asthma, neither for Indonesia as a whole or in Sumatra and Kalimantan only. However, I do find evidence that both the total oil palm area and total palm oil production in selected provinces in Kalimantan adversely affect health, as measured by the probability of an individual reporting symptoms of asthma. Both the area devoted to oil palm and palm oil production could have adverse health effects, the former due to forest clearing by burning the land and the latter due to odours and toxic waste being released by mills and refineries.

As a result, the second important finding relates to regional disparities and heterogeneity. The results confirm that palm oil production cannot be taken as a panacea to increase rural welfare, and therefore it would seem advisable to consider the cost and benefits in relation to the local conditions and environment before deciding any future expansions.

This study has not addressed the question of whether palm oil production has benefited individual farmers. However, the results suggest that smallholder production is, if anything, welfare reducing in the short term, and also that households not directly involved in the production are affected. Moreover, the lack of evidence of positive effect for the total production in selected provinces in Kalimantan is not consistent with the arguments put forward by the proponents of palm oil production. If palm oil production alters ecosystems and water management, for example, households need additional resources to cope with these changes.

These findings lay out various areas for future research. A natural step forward would be to replicate the study using firm-level data. Another interesting area for future research would be to study district level outcomes in order to benefit from panel analyses, and here potential outcomes would be poverty rates and regional GDP, among others. A further important extension would be to expand the study period to capture long-term effects. And finally, the question related of effect of palm oil expansion on migration is left for future research.

Chapter 5 Conclusion

In this thesis I have studied changes in household welfare in relation to topical environment and climate change-related questions in Indonesia. The first two essays explore how household welfare is affected by the timing of monsoon onset. More specifically, the first essay investigates how household per capita expenditure and per capita farm profits are affected by delays in monsoon onset. Moreover, the first essay analyses the heterogeneity among households and finds that middle-income households are most vulnerable to delayed monsoon onset. Further finding of the chapter suggests that rainfall intensity explains little changes in household welfare after the timing of the rainfall is controlled for.

The second essay investigates schooling and child labour in relation to the delay in monsoon onset. As an additional contribution the essay seeks to better understand parents' decision to send their children to school in riskier environments. The results suggest that delayed monsoon onset increases child labour. With respect to schooling, monsoon onset coinciding with the transition year reduces the probability of attending school in the following years and parents postpone children's schooling in riskier environments.

The third essay investigates impacts of the expansion of palm oil production in Indonesia. The expansion of palm oil production is an important topic both in terms of rural development and climate change, as large areas of tropical forest are cleared for oil palm plantations. The results suggest that district-level smallholder production has a weak negative effect on household per capita expenditure and that district-level total production increases the prevalence of asthma in Kalimantan. Putting this evidence together it is argued that palm oil production is not a panacea to increase rural welfare.

In the following I will discuss briefly the main limitations of the methodologies chosen and why, despite of the limitations, I found these methodologies superior to any other alternative. In addition I will suggest extensions and areas for future research.

The main challenge of the first essay is to choose the methodology for addressing the heterogeneous impacts of delayed monsoon onset. A major objective of the essay is to study which households in the expenditure distribution are most affected by monsoon onset. The essay presents two methods. First, I construct an average household expenditure over the panel and take expenditure terciles of the average consumption. This method assumes a constant treatment effect within a group of households. However, the assumption of the constant treatment effect might be too strong. In order to investigate this further I present an alternative method for grouping households. I regress average household expenditure on selected predetermined household characteristics and then create predicted values of household expenditure. Finally I use the predicted values as the basis for grouping households into terciles. Importantly, the alternative method confirms my main finding, that is, middle-income households face the biggest losses from delays in monsoon onset.

Another possible problem of the chosen empirical strategy, household fixed effect regression, is that rainfall exhibits serial correlation. An alternative method could be the dynamic panel model. However, this would sacrifice one round of the data and is therefore inappropriate. I do address the problem of serial correlation by presenting regressions where only the first lag and then both the first and the second lag of monsoon onset are included. My main findings remain robust.

The findings of the chapter guide to interesting areas of future research. For example, weather-based insurance mechanisms and climate change field schools could help households and farmers cope with climate variability and increase their understanding about climate change.

Turning to the second essay one important limitation is that it is not possible to distinguish net producers and buyers of rice in the IFLS data. Therefore the reduced form specification does not allow me to examine the mechanisms through which late monsoon onset affects households and lead to an increase in child labour. On the basis of the findings of the first essay monsoon onset is likely to result in a negative income shock. However, changes in the price of rice might have different implications on net producers and buyers of rice. An interesting extension could be to employ alternative data where it is possible to identify net producers and buyers of rice. Another interesting extension could be to investigate the effect of monsoon onset on the intensity of schooling and educational attainment. An important policy lesson that emerges is that extra resources could be directed to ensure that children would have the opportunity to transfer from primary to secondary school. Finally, the findings of this essay suggest that the increase of child labour is not very harmful in terms of schooling but more research on the intensity of schooling is needed to assess the robustness of this hypothesis.

I recognise that some caution should be exercises in the interpretation of my results, given the limitations in the rainfall data. The amount of rainfall stations that match with the IFLS households is limited (36 stations) and moreover, daily rainfall data is undoubtedly measured with some error. Since measurement error in rainfall is independent of local household characteristics, the true effect of delayed onset, both positive and negative, is greater than the estimates presented here. Finally, none of the years prior to the IFLS survey years were an *El Niño* or *La Niña* year. However, I argue that the data generate meaningful variation in the monsoon onset prior to the IFLS study years for the purpose of this thesis.

Ty my knowledge the work presented in the third and final essay is the first study aiming at investigating the welfare effects of the palm oil expansion using large samples of survey data. The identification strategy presented in the chapter is not without limitations but I nevertheless argue that it presents valuable insight on the ongoing discussion about costs and benefits of palm oil production expansion. Palm oil production requires large investments and knowledge, raising the issue about endogeneity. However, finding suitable instruments for palm oil is not straightforward. I identify two sources of potentially exogenous variation in palm oil production. First, I use historical values of palm oil area and production and second, I use district-level forest area prior to my study period. The latter instrument is based on the observation that forest areas create the potential for large oil palm plantations. Both instruments are used to predict oil palm plantation areas and levels of palm oil production and then these predicted values are used as instruments for actual values. The limitation of the first instrument is that was there a omitted variable correlated with both household welfare and palm oil production the use of historical values of palm oil might not eliminate this problem. The limitation of the second instrument is that the definition of forest area used in the study includes mature plantations, including oil palm plantations. Nonetheless, I argue that this is not a major problem as mature plantations only form small part of the forest cover. I also tried some other plausible instruments, elevation of the district among others, but these did not have sufficient power. However, the elevation data was taken from the PODES village study and therefore the exercise could be replicated using district GIS data. More, if available, data on land suitability could be explored as a possible instrument for palm oil. However, rubber and oil palm thrive in similar conditions and in this case the estimated effect would be overall plantation effect and not solely oil palm.

Another limitation of the study is the availability of the palm oil data that only enables me to address short-term effects. Other plausible concerns, such as non-linear effect of the palm oil, are left for future research. Interesting extensions also include replication of the study using firm-level data and farm data as the critics of palm oil production claim that large companies have benefitted most from the expansion. The important policy lesson of the chapter is that more profound analysis on land use, compensation schemes between smallholders and large companies, and the relationship between palm oil and food production is needed in the areas affected.

Bibliography

Al-Samarrai, S. and B. Reilly (2000). Urban and Rural Differences in Primary School Attendance: an Empirical Study for Tanzania. *Journal of African Economies 9* (4), pp. 430-474.

Angelsen, A. (1995). Shifting cultivation and “Deforestation”: A Study from Indonesia. *World Development 23* (10), pp. 1713-1729.

Angrist, J. and A. Krueger (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives 15* (4), pp. 69-85.

Barlow, C., Z. Zen, and R. Gondowarsito (2003). The Indonesian Oil Palm Industry. *Oil Palm Industry Economic Journal 3* (1), pp. 8-15.

Basu, K. (1999). Child Labor: Cause, Consequence, and Cure, with Remarks on International Labor Standards. *Journal of Economic Literature 37* (3), pp. 1083-1119.

Basu, K. and P.H. Van (1998). Economics of Child Labor. *The American Economic Review (88)*, pp. 412-427.

Beegle, K., R. Dehejia, R. Gatti, and S. Krutikova (2008). The Consequences of Child Labour: Evidence from Longitudinal Data in Rural Tanzania. *Policy Research Working Paper Series 4677*, the World Bank.

Beegle, K., R. Deheija, and R. Gatti (2005). Why should we care about child labour: The education, labor market, and health consequences of child labor. *NBER Working Paper No. 10980,* National Bureau of Economic Research.

Behrman, J.R. and A.B. Deolalikar (1991). School Repetition, Dropouts, and the rates of return to Schooling: the Case of Indonesia. *The Oxford Bulletin of Economics and Statistics 53* (4), pp. 467-480.

Behrman, J.R. and J.C. Knowles (1999). Household Income and Child Schooling in Vietnam. *The World Bank Economic Review 13* (2), pp. 211-256.

Bigsten, A., B. Kebede, A. Shimeles and M. Taddesse (2003). Growth and Poverty Reduction in Ethiopia: Evidence from Household Panel Surveys. *World Development 31* (1), pp. 87-106.

Björkman, M. (2006). Income Shocks and Gender Gaps in Education: Evidence from Uganda. *IIES Seminar Paper No. 744*, Institute for International Economic Studies, Stockholm University.

Broich, M., M. Hansen, P. Potapov, B. Adusei, E. Lindquist and S. Stehman (2010, in review). Time Series Analysis of multiresolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation.*

Cameron, L. A. and C. Worswick (2001). Education Expenditure Responses to Crop Loss in Indonesia: A Gender Bias. *Economic Development and Cultural Change 49* (2), pp. 351-363.

Carter, M.R., P.D. Little, T. Mogues, and W. Negatu (2007). Poverty Traps and Natural Disasters in Ethiopia and Honduras. *World Development 35* (5), pp. 835-856.

Casson, A., L. Tacconi, and K. Deddy (2007). Strategies to Reduce Carbon Emissions from the Oil Palm Sector in Indonesia. Indonesian Forest Carbon Alliance.

Chomitz, K.M. and C. Griffiths (1996). Deforestation, shifting cultivation, and tree crops in Indonesia: Nationwide patterns of smallholder agriculture at the forest frontier. *Mimeo, the World Bank.*

Cline, W.R. (2007). Global Warming and Agriculture: Impact Estimates by Country. Center for Global Development and Peterson Institute for International Economics, Washington D.C.

Cook, G.D. and R.G. Heerdegen (2001). Spatial Variation in the duration of the rainy season on monsoonal Australia. *International Journal of Climatology 21* (14), pp. 1723-1732.

De Datta, S.K. (1981). Principles and Practices of Rice Production. John Wiley and Sons, New York.

Deasy, G F. (1942). Localization of Sumatra's Oil Palm Industry. *Economic Geography 18* (2), pp. 153-158.

Dercon, S. (2004). Growth and Shocks: Evidence from Rural Ethiopia. *Journal of Development Economics 74*, pp. 309-329.

Departementen Pertanian; Sekretariat Direktorat Jenderal Perkebunan (2009). Statistik Perkebunan, Tree Crop Estate Statistik 2008-2010, Kelapa Sawit, Oil Palm. Jakarta 2009.

Duflo, E. and R. Pande (2007). Dams. *Quarterly Journal of Economics 122* (2), pp. 601-646.

Easterling, W.E., P.K. Aggarwal, P. Batima, K.M. Brander, L. Erda, S.M. Howden, A. Kirilenko, J. Morton, J.-F. Soussana, J. Schmidhuber, and F.N. Tubiello (2007). Food, fibre and forest products. Climate Change 2007: Impacts, Adaptation and Vulnerability. *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 273-313.

Falcon, W.P., R.L. Naylor, W.L. Smith, M.B. Burke E.B. McCullough (2004). Using Climate Models to Improve Indonesian Food Security. *Bulletin of Indonesian Economic Studies 40* (3), pp. 355-377.

Feintrenie, L., W.K. Chong and P. Levang (2010). Why do Farmers Prefer Oil Palm? Lessons Learnt from Bungo District. *Small-Scale Forestry 9.*

Fitzsimons, E. (2007). The Effects of Risk on Education in Indonesia. *Economic Development and Cultural Change 56*, pp. 1-25.

Food and Agriculture Organization (FAO) (2005). Global Forest Resources Assessment 2005. Progress towards sustainable forest management. *FAO Forestry Paper No 147.*

Foster, A. D. (1995). Prices, Credit Markets and Child Growth in Low-Income Rural Areas. *The Economic Journal 105* (430), pp. 551-570.

Frankenberg, E. and L. Karoly (1995). The 1993 Indonesian Family Life Survey: Overview and Field Report. November, 1995. RAND. DRU-1195/1-NICHD/AID.

Frankenberg, E., D. McKee, and D. Thomas (2005). Health Consequences of Forest Fires in Indonesia. *Demography* 42 (1), pp. 109-129.

Frankenberg, E., J.P. Smith, and D. Thomas (2003). Economic shocks, wealth and welfare. *Journal of Human Resources 38* (2), pp. 280– 321.

Frankenberg, E. and D. Thomas (2000). The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2. March, 2000. DRU-2238/1-NIA/NICHD.

Gillison, A. and N. Liswanti (1999). Impact of oil palm plantations on biodiversity in Jambi, Central Sumatra, Indonesia. Available at: http://www.asb.cgiar.org/data/dataset/IDA1AMZB.htm (Accessed 9.1.2011).

Glewwe, P. (1991). Investigating the determinants of household welfare in Côte d'Ivoire. *Journal of Development Economics 35* (2), pp. 307-337.

Godoy, R.A. (1992). Determinants of smallholder tree crops cultivation. *World Development 20* (5), pp. 713-725.

del Granado, F. J. A., W. Fengler, A. Ragatz, E. Yavuz (2007). Investing in Indonesia's Education: Allocation, Equity, and Efficiency of Public Expenditures. *Policy Research Working Paper No. 4329*, the World Bank.

Grimm, M. (2008). Food price inflation and children's schooling. *European Development Research Network Working Paper*, EUDN WP 2008 - 09.

Gujarati, D.N. (2003). Basic Econometrics, Fourth Edition, McGraw-Hill.

Henry, S.,B. Schoumaker, and C. Beauchemin (2004). The Impact of Rainfall on the First Out-Migration: A Multi-level Event-History Analysis in Burkina Faso. *Population and Environment 24* (5), pp. 423-460.

Houghton, R. A. (1993). The role of the world's forest in global warming. In Ramakrishna, K. and G.M. Woodwell (Eds): The World Forests for the Future. Yale University Press, New Haven.

Jacoby, H. and E. Skoufias (1997). Risk, Financial Markets, and Human Capital in a Developing Country. *Review of Economic Studies 64* (3), pp. 311-335.

Jalan, J. and M. Ravallion (1999). Are the poor less well insured? Evidence on the vulnerability to income risk in rural China. *Journal of Development Economics 58* (1), pp. 61-81.

Jayachandran, S. (2006). Selling Labor Low: Wage responses to Productivity Shocks in Developing Countries. *Journal of Political Economy 114* (3), pp. 538-575.

Jensen, R. (2000). Agricultural Volatility and Investments in Children. *American Economic Review, Papers and Proceedings 90* (2), pp. 399-404.

Kazianga, H. (2005). Income Risk and Schooling Outcomes in Burkina Faso. Available at SSRN: http: //ssrn.com/abstract=872575.

Kazianga, H. and C. Udry (2006). Consumption smoothing? Livestock, insurance and drought in rural Burkina Faso. *Journal of Development Economics 79*, pp. 413-446.

Kessler, J.J., T. Rood, T. Tekelenburg and M. Bakkenes (2007). Biodiversity and Socioeconomic Impacts of Selected Agro-Commodity Production Systems. *The Journal of Environment and Development 16*, pp. 131-160.

Kirono, D.G.C., N.J. Tapper, and J.L. McBride (1999). Documenting Indonesian Rainfall in the 1997-1998 El Niño Event. *Physical Geography 20*, pp. 422-435.

Kishore, K., A.R. Subbiah, T. Sribimawati, S. Diharto, S. Alimoeso, P. Rogers, and A. Setiana (2000). Indonesia Country Study. Asian Disaster Preparedness Center (ADPC). Available at: http://www.unu.edu/env/govern/ElNino/CountryReports/pdf/Indonesia.pdf (accessed 29.4.2008)

Kochar, A. (1999). Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India. *The Review of Economics and Statistics 81* (1), pp. 50-61.

Kruger, D.I. (2007). Coffee production effects on child labor and schooling in rural Brazil. *Journal of Development Economics 82*, pp. 448-462.

Korkeala, O., D. Newhouse, and M. Duarte (2009). Distributional Impact Analysis of Past Climate Variability in Rural Indonesia. *Policy Research Working Paper No. 5072*, the World Bank.

Lavy, V. (1996). School supply constraints and children's educational outcomes in rural Ghana. *Journal of Development Economics 51*, pp. 291-314.

Lobell, D.B. and C.B. Field (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters 2.*

Maccini, S., and D. Yang, D. (2009). Under the Weather: Health, Schooling, and Socioeconomic Consequences of Early-Life Rainfall. *American Economic Review 99* (3), pp. 1006-1026.

Maddox, T.M. (2007). Oil palm and mammal conservation. Paper to the International Conference on Oil Palm and Environment 2007, Bali, Indonesia.

Manning, C. (2000): The Economic Crisis and Child Labour in Indonesia. ILO IPEC (International Programme of the Elimination of Child Labour) Working Paper.

McCarthy, J. and Z. Zen (2010). Regulating the Oil Palm Boom: Assessing the Effectiveness of Environmental Governance Approaches to Agro-Industrial Pollution in Indonesia. *Law and Policy 32* (1), pp. 153-179.

McCulloch, N. (2008). Rice Prices and Poverty in Indonesia. *Bulletin of Indonesian Economic Studies 44* (1), pp. 45-63.

Miyamoto, M. (2006). Forest conversion to rubber around Sumatran villages in Indonesia: Comparing the impacts of road construction, transmigration projects and population. *Forest Policy and Economics 9*, pp. 1-12.

Montgomery, M., A. Kouamé, and R. Oliver (1995). The Tradeoff between Number of Children and Child Schooling: Evidence from Côte d'Ivoire and Ghana. *LSMS Working Paper No. 112*, The World Bank.

Moron, V., A. Robertson, and R. Boer (2009). Spatial Coherence and Seasonal Predictability of Monsoon onset over Indonesia. *Journal of Climate 22*, pp. 840-850.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U.S. Labour Market. *Quarterly Journal of Economics 18* (2), pp. 549-599.

Murray, M.P. (2006). Avoiding Invalid Instruments and Coping with Weak Instruments. *Journal of Economic Perspectives 20* (4), pp. 111-132.

Natawidjaja, R.S., D. Supyandi, C. Tulloh, A.C. Tridakusumah, E.M. Calford, and M. Ford (2009). Climate change, food security and income distribution: Adaptations of small scale farmers. Policy Brief No. 9, Australia Indonesia Governance Research Partnership.

National Research Council, Committee on Environmental Epidemiology (U.S.) (1991). Environmental Epidemiology: Public Health and Hazardous Waste. National Academy Press, Washington D.C.

Naylor, R.L., D.S. Battisti, W.P. Falcon, and M. Burke (2007a). Assessing Risks of climate variability and climate change for Indonesian rice agriculture. *PNAS 104*, (19), pp. 7752-7757.

Naylor, R. L., A.J. Liska, M.B. Burke, W.P. Falcon, J.C. Gaskell, S.D. Rozelle, and K. H. Cassman (2007b). The Ripple Effect: Biofuels, Food Security, and the Environment. *Environment: Science and Policy for Sustainable Development 49* (9), pp. 31-43.

Naylor, R.L.,W.P. Falcon, D. Rochberg, and N. Wada (2001). Using El Niño-Southern Oscillation Climate Data to Predict Rice Production in Indonesia. *Climatic Change 50* (3), pp. 255-265.

Newhouse, D. (2005). The Persistence of Income Shocks: Evidence from Rural Indonesia. *Review of Development Economics 9* (3), pp. 415-433.

O'Donnell, O., F.C. Rosati, and E. Van Doorslaer (2005). Health effects of child work: Evidence from rural Vietnam. *Journal of Population Economics, 18* (3), pp. 437-567.

Osterman, K. and M. Brauer (2001). Air quality during haze episodes and its impact on health. In P. Eaton and M. Radojevic (eds): Forest Fires and Regional Haze in Southeast Asia. Nova Science Publishers, New York.

Overpeck, J. T. and J.E. Cole (2007). Climate change: Lessons from a distant monsoon. *Nature 445*, pp. 270-271.

Paxon, C. (1992). Using Weather Variability to estimate the Response of Savings to Transitory Income in Thailand. *American Economic Review 82* (1), pp. 15-33.

Peskett, L., R. Slater, C. Stevens, and A. Dufey (2007). Biofuels, Agriculture and Poverty Reduction. Natural Resource Perspectives 107, Overseas Development Institute.

Pradhan, M. (1998). Enrolment and Delayed Enrolment of Secondary School age Children in Indonesia. *Oxford Bulletin of Economics and Statistics 60* (4), pp. 413-430.

Priyambada, A., A. Suryahadi, S. Sumarto (2005). What Happened to Child Labor in Indonesia during the Economic Crisis: The Trade-off between School and Work. SMERU Working Paper. Available at: http://www.smeru.or.id/report/workpaper/childlabor/whchildlaborrevisedeng.pdf. (Accessed 20th November 2011).

Purnamasari, R., O. Cacho and P. Simmons (1999). Management Strategies for Indonesian small-holder rubber production in South Sumatra: A bioeconomic analysis. University of New England, Working Paper Series in Agricultural and Resource Economics, No. 99-14.

Rist, L., L. Feintrenie and P. Levang (2010). The livelihood impacts of oil palm: Smallholders in Indonesia. *Biodiversity and Conservation 19* (4), pp. 1009-1024.

Ritung, S., F. Agus and H. Hidayat (2007). Land Suitability Evaluation: With a Case Map of Aceh Barat District. Indonesian Soil Research Institute and World Agroforestry Centre, Mimeo. Available at: http://www.worldagroforestry.org/downloads/publications/PDFs/mn15224.pdf (accessed 24 Nov 2010).

Rose, E. (2001). Ex ante and ex post labour supply response to risk in a low-income area. *Journal of Development Economics, 64*, pp. 371-388.

Rosenzweig, M. and H. Binswanger (1993). Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments. *Economic Journal 103*, pp. 56-78.

Rosenzweig, M. and K. Wolpin (2000). Natural "Natural Experiments" in Economics. *Journal of Economic Literature 38* (4), pp. 827-874.

Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of Urban Economics 61*, pp. 345-371.

Sanghi, A. and R. Mendelsohn (2008). The impacts of global warming on farmers in Brazil and India. *Global Environmental Change 18* (4), pp. 655-665.

Sheil, D., A. Casson, E. Meijaard, M. Van Noordwijk, J. Gaskell, J. Sunderland-Groves, K. Wertz, and M. Kanninen (2009). The impacts and opportunities of oil palm in Southeast Asia. What do we know and what do we need to know? CIFOR Occasional Paper No. 51.

Sirait, M.T. (2009). Indigenous Peoples and Oil Palm Plantation Expansion in West Kalimantan, Indonesia. Cordaid memisa, the Hague. Available at:

http://www.worldagroforestrycentre.org/sea/Publications/files/report/RP0247-09/RP0247-09-1.PDF (Accessed 9 June 2011).

Skoufias, E, B. Essama-Nssah, and R. Katayama, R. (2011). Too Little Too Late: Welfare Impacts of Rainfall Shocks in Rural Indonesia. *World Bank Policy Research Working Paper No. 5165,* The World Bank.

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. March 2004. WR144/1-NIA/NICHD.

Susila, W.R. (2004). Contribution of oil palm industry to economic growth and poverty alleviation in Indonesia. *Jurnal Litbang Pertanian 23,* pp. 107-114.

Thomas, D., K. Beegle. E. Frankenberg, B. Sikoki, J. Strauss, and G. Teruel (2004). Education in a crisis. *Journal of Development Economics 74,* pp. 53-85.

Townsend, Robert M. (1994). Risk and Insurance in Village India. *Econometrica, 62* (3), p. 539-591.

Tomich, T. P., M. Van Noordwijk, S. Budidarsono, A. Gilison, T. Kusumanto, D. Murdiyarso, F. Stolle and A.M. Fagi (1998). Alternatives to Slash-and-Burn in Indonesia, Summary Report and Synthesis of Phase II. ASB-Indonesia Report No. 8. Bogor, Indonesia. Available at: http://www.worldagroforestrycentre.org/sea/Publications/files/report/RP0002-04.PDF. (Accessed 4 October 2010).

United States Department of Agriculture (USDA) (2009). Oilseeds: World Markets and Trade. Available at http://www.fas.usda.gov/psdonline/circulars/oilseeds.pdf (accessed on 23rd September 2009).

Vermeulen, S. and N. Goad (2006). Towards better practise in smallholder palm oil production. *Natural Resource Issues Series No, 5.* International Institute for Environment and Development. London, UK.

Walker, T. and J. G. Ryan (1990). *Village and Household Economies in India's Semi-Arid Tropics.* Baltimore, Johns Hopkins University Press.

World Bank (1994). Indonesia: Environment and Development. The World Bank, Washington DC.

World Bank (2006a). Making the New Indonesia Work for the Poor. The World Bank, Washington DC.

World Bank (2006b). World Development Report 2007: Development and Next Generation. The World Bank, Washington DC.

World Bank (2008). Adapting to Climate Change: The Case of Rice in Indonesia A Study under the Rice Policy Dialogue AAA (P108646). Report No. 44434-ID.

World Bank (2010). Environmental, Economic and Social Impacts of Oil palm in Indonesia: A Synthesis of Opportunities and Challenges. Discussion Paper, May 2010. Available at: http://www.ifc.org/ifcext/agriconsultation.nsf/AttachmentsByTitle/WB+discussion+paper/$FILE/WB_Oil+Palm+SynthesisDiscussionPaperMay2010.pdf (Accessed 29 November 2011).

Yang, D. (2008). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *The Economic Journal 188*, pp. 591-630.

Zimmerman, F. J. and M. R. Carter (2003). Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints. *Journal of Development Economics 71* (2), pp. 233-260.

WWW-sites:

Agrometeorology.Org <http://www.agrometeorology.org/topics/accounts-of-operational-agrometeorology/climate-field-schools-in-indonesia-coping-with-climate-change-and-beyond>. Accessed 10 October 2009.

Badan Pusat Statistik <http://www.bps.go.id/tnmn_pgn.php?eng=0>. Accessed 5 February 2010.

Climate Prediction Centre (CPC) <http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/impacts/warm_impacts.shtml>. Accessed 14 October 2009.

FAOSTAT <http://faostat.fao.org/>

Istockanalyst <http://www.istockanalyst.com/article/viewiStockNews/articleid/3660667>. Accessed 18 September 2011.

Jakartaglobe 1 February 2010 <http://www.thejakartaglobe.com/business/indonesia-inflation-hits-seven-month-high-due-to-rising-food-prices/356040>. Accessed 21 June 2010

Jakarta Post 28 June 2010 (available at: <http://www.thejakartapost.com/news/2010/06/28/analysis-indonesia%E2%80%99s-12year-compulsory-education-program.html>. Accessed 14 January 2011).

Jakarta Post 12 February 2009 (available at: <http://www.thejakartapost.com/news/2009/12/02/palm-oil-economic-pillar-indonesia.html>. Accessed 15 October 2010).

USDA <http://www.pecad.fas.usda.gov/highlights/2007/12/Indonesia_palmoil/>. Accessed 15 September 2010

Appendix A

Appendix Chapter One

A1 Note to the daily rainfall data

The original rainfall data is the NOAA Global Summary or the day. As the rainfall data set contains some missing values I requested imputed values from CEREGE (*Centre Europeen de Recherche et d'Enseignement des Geosciences de l'Environnement)*. CEREGE has used the imputed values in their own research; see for example Moron et al., (2009). CEREGE provided me with two datasets, one with missing values and another one where missing values have been imputed using simple stochastic weather generator, considering wet-to-wet and dry-to-wet persistence and gamma distribution for wet days, computed on a monthly basis for each station. If a month is completely missing (<6% of station-months for SOND[1]), this method simulates a climatological daily sequence for the month in question. The data provided by CEREGE contains 58 weather stations but unfortunately the dataset does not cover Bali, West Nusa Tenggara, and South Sulawesi, which are IFLS provinces. In the data the share of missing (imputed values) is 0.204 for August-December (months used for calculating the monsoon onset in my research).

In order to complete missing data also for the stations without the imputed values I requested additional data from the Indonesia Meteorological Agency (BMG). Data provided by BMG helped to fulfill some gaps but complete daily rainfall data was not possible to obtain, however. The NOAA data completed with the BMG data contain six stations (one in Bali, three in West Nusa Tenggara and two in South Sulawesi), implying that in total I have data for 64 weather stations.

[1] SOND stands for September, October, November and December.

The issue of measurement error is topical considering that I have imputed/missing values in my main explanatory variable. I decided to identify stations that have lots of imputed values in order to decrease the problem of measurement error

I calculated the share of missing values for August-December using the data provided by CEREGE. I mainly considered IFLS data and provinces for the purpose of cleaning the rainfall data. Therefore I might have left some stations with high share of missing values that do not match with IFLS villages but do match SUSENAS households.

After dropping the 13 problematic stations the share of imputed (missing) values is 0.148 for August-December. Considering only the stations that match with the IFLS villages the share of imputed values is 0.121. These figures do not contain missing values for stations in South Sulawesi, Bali or West Nusa Tenggara (stations not included in the imputed data).

A2 Variable descriptions and summary statistics

Table A.1. Variable descriptions, chapter two.

Variable	Variable description
Monthly expenditure pc	Household monthly expenditure per capita
Log total monthly expenditure	Log of total household monthly expenditure
Monthly farm profits pc	Monthly farm profits per capita in rupiahs
Log household size	Log of household size
Lagged monsoon onset	Monsoon onset previous year to the survey
Monsoon onset 2 years ago	Monsoon onset 2 years ago to the survey
$Spline_1$	Lagged monsoon onset < -0.6 standard deviations
$Spline_2$	Lagged monsoon onset > -0.6 standard deviations
$Spline_{21}$	Lag monsoon onset 2 years ago < -0.6 standard deviations
$Spline_{22}$	Lag monsoon onset 2 years ago >-0.6 standard deviations
Head female	1 if household has female head
Age of the head	Set of dummy variables, reference category is aged 15-25 years
Head 15-24 years [a]	1 if head below 25 years
Head 25-34	1 if head 25-34 years
Head 35-49	1 if head 35-49 years
Head 50-64	1 if head 50-64 years
Head 65+	1 if head over 65 years
Education of the head	Set of dummy variables, reference category no formal education
Head no education [a]	1 if no formal education
Head primary school	1 if some primary or completed primary
Head Junior high school	1 if some or completed junior high school
Head Senior high school	1 if some or completed senior high school
Head university	1 if some or completed university education
Father some Jr	1 if some junior high school

Variable	Variable description
Province	Set of dummy variables, reference category West Java
Aceh	
North Sumatra	North Sumatra
West Sumatra	West Sumatra
Riau	Riau
Jambi	Jambi
South Sumatra	South Sumatra
Bengkulu	Bengkulu
Lampung	Lampung
Belitung	Belitung
Riau Islands	Riau Islands
West Java [a]	West Java
Central Java	Central Java
Yogyakarta	Yogyakarta
East Java	East Java
Banten	Banten
Bali	Bali
West Nusa Tenggara	Nusa Tengarra Barat
East Nusa Tenggara	East Nusa Tenggara
West Kalimantan	West Kalimantan
Central Kalimantan	Central Kalimanta
South Kalimantan	South Kalimantan
East Kalimantan	East Kalimantan
South Sulawesi	South Sulawesi

[a]The reference category, which is not included in the regression analysis

Table A.2. Sample Descriptive Statistics, chapter two.

	IFLS		SUSENAS	
Variable	Mean	s.d.[a]	Mean	s.d.[a]
Log monthly expenditure pc [a]	12.089	0.710	12.426	0.450
Log total monthly expenditure [b]	13.435	0.743	NA	NA
Monthly farm profits pc [b]	31711.84	56959.32	NA	NA
Log household size	1.346	0.509	NA	NA
Monsoon onset previous year	-0.563	0.631	-0.004	1.085
Monsoon onset 2 years ago	-0.280	1.16	0.081	1.112
$Spline_1$	-0.833	0.377	-0.775	0.349
$Spline_2$	0.271	0.361	0.772	0.886
$Spline_{21}$	-0.953	0.502	-0.769	0.341
$Spline_{22}$	0.672	0.789	0.850	0.912
Head female	NA	NA	0.121	0.326
Head age 15-24 years	NA	NA	0.030	0.169
Head age 25-34 years	NA	NA	0.217	0.412
Head age 35-49 years	NA	NA	0.390	0.488
Head age 50-64 years	NA	NA	0.255	0.436
Head age 65+ years	NA	NA	0.108	0.310
Head no schooling	NA	NA	0.158	0.365
Head primary education	NA	NA	0.633	0.482
Head junior high school	NA	NA	0.107	0.310
Head senior high school	NA	NA	0.085	0.279
Head university	NA	NA	0.017	0.129
Aceh	NA	NA	0.030	0.171
North Sumatra[a]	NA	NA	0.057	0.233
West Sumatra	NA	NA	0.030	0.171
Riau	NA	NA	0.031	0.173

Variable	IFLS Mean	IFLS s.d.[a]	SUSENAS Mean	SUSENAS s.d.[a]
Jambi	NA	NA	0.017	0.128
South Sumatra	NA	NA	0.057	0.232
Bengkulu	NA	NA	0.016	0.127
Lampung	NA	NA	0.065	0.246
Belitung	NA	NA	0.012	0.108
Riau Islands	NA	NA	0.004	0.065
West Java	NA	NA	0.162	0.368
Central Java	NA	NA	0.128	0.334
Yogyakarta	NA	NA	0.009	0.096
East Java	NA	NA	0.163	0.370
Banten	NA	NA	0.031	0.173
Bali	NA	NA	0.012	0.107
West Nusa Tenggara	NA	NA	0.023	0.150
East Nusa Tenggara	NA	NA	0.025	0.155
West Kalimantan	NA	NA	0.029	0.168
Central Kalimantan	NA	NA	0.021	0.143
South Kalimantan	NA	NA	0.018	0.134
East Kalimantan	NA	NA	0.015	0.123
South Sulawesi	NA	NA	0.037	0.190
Sample size	11333		97176	

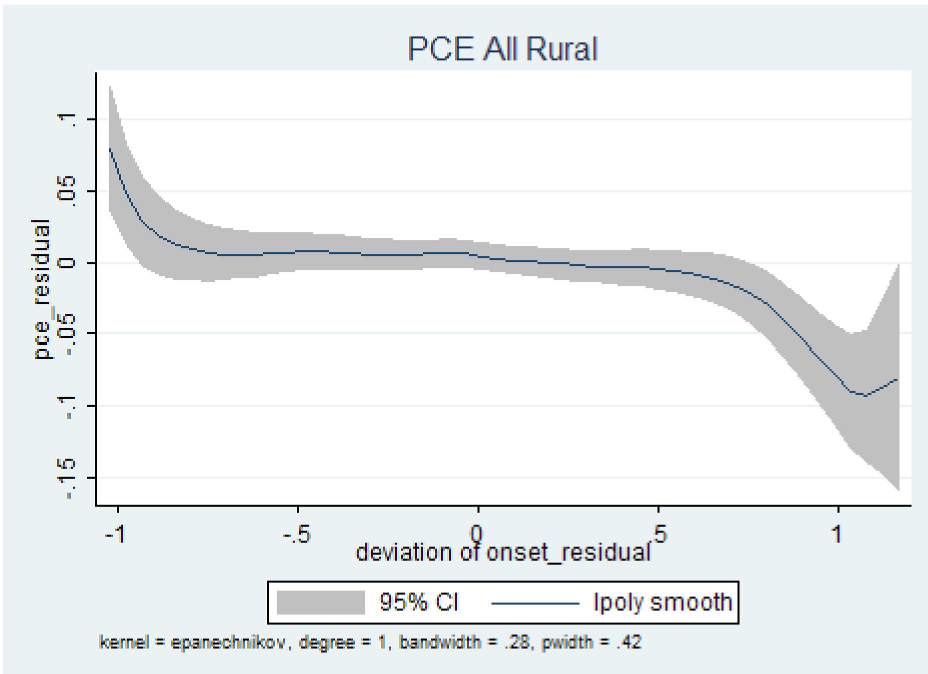
NA signifies not applicable.

[a] denotes for standard deviation

[b] Household expenditure and farm profits are expressed in December 2000 Jakarta prices in the IFLS data and 2007 prices in the SUSENAS data.

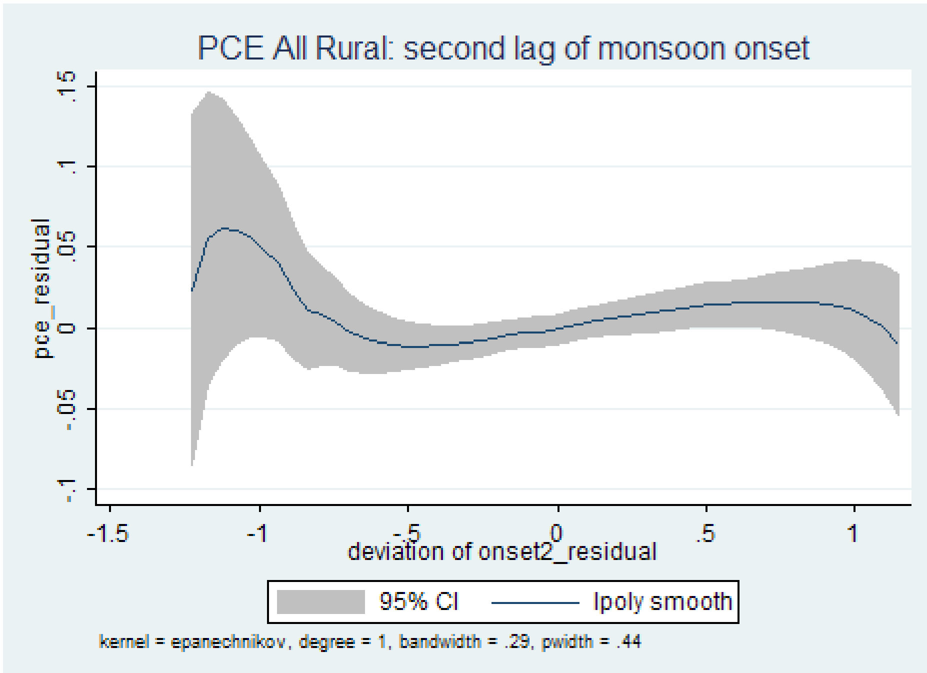
A3 Estimation results

Figure A.1. Local polynomial smoothing for per capita expenditure and the first lag of monsoon onset, all rural households.



Note: Local polynomial smooth, conditional on household and year fixed effects.

Figure A.2. Local polynomial smoothing for per capita expenditure and the second lag of monsoon onset.



Note: Local polynomial smooth, conditional on household and year fixed effects.

Table A.3. Prediction regression, dependent variable average household per capita expenditure (in rupiahs) over the panel (estimation results from equation 2.3).

Dependent variable	Average household per capita expenditure
Age	3692.6**
	(1824.4)
Age squared	-26.3
	(16.8)
Elementary	31152.1***
	(10979.2)
Junior high	84338.2***
	(18477.2)
Senior high	152855.9***
	(21413.9)
University	320751.9***
	(85819.3)
Other education	-161199.7
	(115028.0)
Worked three years ago	2921.1
	(10636.8)
Weeks worked	547.6*
	(305.5)
Hours worked	-486.3
	(435.5)
R-squared	0.106
Observations	3725
District fixed effects	Yes

Robust standard errors in parenthesis
In education the reference category is no formal education
District fixed effects refer to the district of the head at the age of 12
* p<0.1, ** p<0.05, *** p<0.01

Table A.4. Per capita expenditure, all rural households. Spline function approach, knot at -0.6, IFLS data.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.048	0.038	0.025	0.036
$Spline_2$	-0.116	0.057**	-0.069	0.056
Two years ago				
$Spline_1$			-0.106	0.049**
$Spline_2$			0.086	0.039**
Observations	11333		11333	
Households	3988		3988	
R^2 (within)	0.123		0.127	
Estimated Effects				
Early Onset				
Prior year	-0.048	0.038	-0.025	0.036
Two years ago			0.106	0.049**
Late Onset				
Prior Year	-0.116	0.057**	-0.069	0.056
Two years ago			0.086	0.039**

Monsoon onset is measured in standard deviations from historical mean. Estimated effects of early onset refer to effect on expenditure at -1 standard deviations, while late onset refers to 1 standard deviation. Additional controls include survey year effects and household-level fixed effects, with split-off households treated as new households. Standard errors are clustered on rain stations, and are robust to heteroscedasticity. Households are weighted by their mean sample weights over the panel.
* p<0.1, ** p<0.05, *** p<0.01

Table A.5. Per capita expenditure, all rural households. Quadratic approximation, IFLS data.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
Monsoon Onset	-0.097	0.054*	-0.061	0.049
Onset Squared	-0.055	0.030*	-0.035	0.029
Two years ago				
Monsoon onset			0.032	0.028
Onset Squared			0.036	0.017
Observations	11333		11333	
Households	3988		3988	
R^2 (within)	0.122		0.125	
Estimated Effects				
Early Onset				
Prior year	0.041	0.032	0.026	0.028
Two years ago			0.004	0.027
Late Onset				
Prior Year	-.152	.081*	-0.096	0.075
Two years ago			0.069	0.038*

See notes to table A4.

Table A.6. Farm profits per capita, all farm households. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Per capita farm profits			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	1861.29	2441.84	2289.77	2448.90
$Spline_2$	-7869.31	5099.19	-8939.33	4889.73*
Two years ago				
$Spline_1$			3027.43	3716.53
$Spline_2$			-469.20	2948.85
Observations	6745		6745	
Households	2439		2439	
R^2 (within)	0.01		0.01	
Estimated Effects				
Early Onset				
Prior year	-1861.29	2441.84	-2289.77	2448.90
Two years ago			-3027.43	3716.53
Late Onset				
Prior Year	-7869.31	5099.19	-8939.33	4889.73*
Two years ago			-469.20	2948.85

See notes to table A4.

Table A.7. Farm profits per capita. Quadratic function, IFLS data.

Dependent variable	Per capita farm profits			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
Monsoon Onset	-7,107.94	4,558.148	-7,857.44	4,330.97*
Onset Squared	-3,518.055	2,471.687	-3,909.88	2,348.44
Two years ago				
Monsoon onset			667.60	2,063.27
Onset Squared			-495.85	1,379.58
Observations	6745		6745	
Households	2439		2439	
R^2 (within)	0.01		0.01	
Estimated Effects				
Early Onset				
Prior year	3589.89	2315.43	3947.56	2240.88*
Two years ago			-1163.45	2038.33
Late Onset				
Prior Year	-10626	6957.78	-11767.32	6597.23*
Two years ago			171.76	2857.59

See notes to table A4.

Table A.8. Per capita expenditure, rural farm households. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.038	0.038	0.017	0.036
$Spline_2$	-0.107	0.072	-0.084	0.067
Two years ago				
$Spline_1$			-0.057	0.068
$Spline_2$			0.129	0.042***
Observations	7026		7026	
Households	2439		2439	
R^2 (within)	0.134		0.141	
Estimated Effects				
Early Onset				
Prior year	-0.038	0.038	-0.017	0.036
Two years ago			0.057	0.068
Late Onset				
Prior Year	-0.107	0.072	-0.084	0.067
Two years ago			0.129	0.042***

See notes to table A4.

Table A.9. Per capita expenditure, rural non-farm households. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.051	0.053	0.032	0.049
$Spline_2$	-0.117	0.047**	-0.049	0.047
Two years ago				
$Spline_1$			-0.177	0.035***
$Spline_2$			0.050	0.036
Observations	4290		4290	
Households	1542		1542	
R^2 (within)	0.116		0.123	
Estimated Effects				
Early Onset				
Prior year	-0.051	0.053	-0.032	0.049
Two years ago			0.177	0.035***
Late Onset				
Prior Year	-0.117	0.047**	-0.049	0.047
Two years ago			0.050	0.036

See notes to table A4.

Table A.10. Farm profits per capita, first expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Per capita farm profits			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	4260.95	2897.41	3754.64	3223.02
$Spline_2$	-7676.00	3585.93**	-7247.31	3735.16*
Two years ago				
$Spline_1$			-1354.10	3719.43
$Spline_2$			2683.25	2508.61
Observations	2437		2437	
Households	872		872	
R^2 (within)	0.031		0.033	
Estimated Effects				
Early Onset				
Prior year	-4260.95	2897.41	-3754.64	3223.02
Two years ago			1354.10	3719.43
Late Onset				
Prior Year	-7676.00	3585.93**	-7247.31	3735.16*
Two years ago			2683.25	2508.61

See notes to table A4.

Table A.11. Farm profits per capita, second expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Per capita farm profits			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	1856.59	2548.98	1746.72	2284.95
$Spline_2$	-13998.66	3949.28***	-13476.91	3632.56***
Two years ago				
$Spline_1$			-1419.61	2280.17
$Spline_2$			-559.61	4901.96
Observations	2239		2239	
Households	816		816	
R^2 (within)	0.022		0.022	
Estimated Effects				
Early Onset				
Prior year	-1856.59	2548.98	-1746.72	2284.95
Two years ago			1419.61	2280.17
Late Onset				
Prior Year	-13998.66	3949.28***	-13476.91	3632.56***
Two years ago			559.61	4901.96

See notes to table A4.

Table A.12. Farm profits per capita, third expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Per capita farm profits			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	-444.88	3936.31	2225.83	4110.85
$Spline_2$	-3298.93	12248.15	-8273.35	12095.48
Two years ago				
$Spline_1$			13236.23	9544.74
$Spline_2$			-5573.09	5156.90
Observations	2069		2069	
Households	751		751	
R^2 (within)	0.012		0.014	
Estimated Effects				
Early Onset				
Prior year	444.88	3936.31	-2225.83	4110.85
Two years ago			-13236.23	9544.74
Late Onset				
Prior Year	-3298.93	12248.15	-8273.35	12095.48
Two years ago			-5573.09	5156.90

See notes to table A4.

Table A.13. Per capita expenditure, first expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log of per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.096	0.026***	0.075	0.023***
$Spline_2$	-0.108	0.046**	-0.071	0.044
Two years ago				
$Spline_1$			-0.086	0.051
$Spline_2$			0.075	0.042*
Observations	3758		3758	
Households	1307		1307	
R^2 (within)	0.162		0.166	
Estimated Effects				
Early Onset				
Prior year	-0.096	0.026**	-0.075	0.023***
Two years ago			0.086	0.051
Late Onset				
Prior Year	-0.108	0.046**	-0.071	0.044
Two years ago			0.075	0.042*

See notes to table A4.

Table A.14. Per capita expenditure, second expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log of per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.020	0.035	0.006	0.031
$Spline_2$	-0.188	0.070**	-0.151	0.071**
Two years ago				
$Spline_1$			-0.078	0.054
$Spline_2$			0.050	0.047
Observations	3787		3787	
Households	1327		1327	
R^2 (within)	0.141		0.143	
Estimated Effects				
Early Onset				
Prior year	-0.020	0.035	-0.006	0.031
Two years ago			0.078	0.054
Late Onset				
Prior Year	-0.188	0.07**	-0.151	0.071**
Two years ago			0.050	0.047

See notes to table A4.

Table A.15. Per capita expenditure, third expenditure tercile. Spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log of per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.023	0.071	-0.007	0.072
$Spline_2$	-0.049	0.079	0.014	0.085
Two years ago				
$Spline_1$			-0.143	0.059**
$Spline_2$			0.117	0.047**
Observations	3788		3788	
Households	1354		1354	
R^2 (within)	0.105		0.110	
Estimated Effects				
Early Onset				
Prior year	-0.023	0.071	0.007	0.072
Two years ago			0.143	0.059**
Late Onset				
Prior Year	-0.049	0.079	0.014	0.085
Two years ago			0.117	0.047**

See notes to table A4.

Table A.16. The effect of monsoon onset on log household size, 2 lags, spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log household size							
	All households		**Poor households**		**Middle-class households**		**Rich households**	
Variable	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Prior year								
$spline_1$	-0.010	0.007	-0.029	0.014**	-0.021	0.014	0.019	0.015
$spline_2$	0.050	0.017***	0.059	0.019***	0.074	0.019***	0.021	0.026
Two years ago								
$spline_{21}$	-0.030	0.015**	-0.062	0.023**	-0.000	0.021	-0.022	0.018
$spline_{22}$	-0.001	0.010	0.011	0.014	0.009	0.016	-0.016	0.012
Observations	11333		3758		3787		3788	
Households	3988		1307		1327		1354	
R^2 (within)	0.00		0.01		0.01		0.01	
<u>Estimated Effects</u>								
Early Onset								
Prior year	0.010	0.007	0.029	0.014**	0.021	0.014	-0.019	0.015
Two years ago	0.030	0.015**	0.062	0.023**	0.000	0.021	0.022	0.018
Late Onset								
Prior Year	0.050	0.017***	0.059	0.019***	0.074	0.019***	0.021	0.026
Two years ago	-0.001	0.010	0.011	0.014	0.009	0.016	-0.016	0.012

See notes to table A4.

Table A.17. The effect of monsoon onset on log total household expenditure, 2 lags, spline function approach, knot at -0.6 standard deviations. IFLS data.

Dependent variable	Log total household expenditure							
	All households		**Poor households**		**Middle-class households**		**Rich households**	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Prior Year								
$spline_1$	0.014	0.036	0.045	0.025*	-0.015	0.028	0.011	0.071
$spline_2$	-0.019	0.054	-0.011	0.052	-0.077	0.064	0.030	0.076
Two years ago								
$spline_{21}$	-0.135	0.054**	-0.146	0.055**	-0.079	0.058	-0.164	0.064**
$spline_{22}$	0.086	0.039**	0.087	0.045*	0.058	0.039	0.102	0.049**
Observations	11333		3758		3787		3788	
Households	3988		1307		1327		1354	
R^2 (within)	0.12		0.16		0.13		0.10	
Estimated Effects								
Early Onset								
Prior year	-0.014	0.036	-0.045	0.025*	0.015	0.028	-0.011	0.071
Two years ago	0.135	0.054**	0.146	0.055**	0.079	0.058	0.164	0.064**
Late Onset								
Prior Year	-0.019	0.054	-0.011	0.052	-0.077	0.064	0.030	0.076
Two years ago	0.086	0.039**	0.087	0.045*	0.058	0.039	0.102	0.049**

See notes to table A4.

Table A.18. The effect of rainfall intensity on per capita expenditure, spline fuction approach. Knot at -0.6 standard deviations, two lags of monsoon onset. IFLS data.

Dependent variable	Log per capita expenditure							
	All households		**Poor households**		**Middle-class households**		**Rich households**	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Prior year								
$spline_1$	0.098	0.054*	-0.010	0.053	0.149	0.062**	0.108	0.078
$spline_2$	-0.011	0.026	-0.001	0.028	-0.003	0.021	-0.010	0.039
Two years ago								
$spline_{21}$	-0.025	0.088	0.030	0.092	-0.122	0.082	0.020	0.102
$spline_{22}$	0.041	0.027	0.032	0.030	0.055	0.027	0.030	0.033
Observations	11333		3758		3787		3788	
Households	3988		1307		1327		1354	
R^2 (within)	0.123		0.158		0.145		0.107	
<u>Estimated Effects</u>								
Negative shock								
Prior year	-0.098	0.054*	0.010	0.053	-0.149	0.062**	-0.108	0.078
Two years ago	0.025	0.088	-0.030	0.092	0.122	0.082	-0.020	0.102
Positive shock								
Prior Year	-0.011	0.026	-0.001	0.028	-0.003	0.021	-0.010	0.039
Two years ago	0.041	0.027	0.032	0.030	0.055	0.027	0.030	0.033

Rainfall intensity is measured in standard deviations from historical mean. Estimated effects of negative intensity shock refer to effect on expenditure at -1 standard deviations, while positive shock refers to 1 standard deviation. Additional controls include survey year effects and household-level fixed effects, with split-off households treated as new households. Standard errors are clustered on rain stations, and are robust to heteroscedasticity. Households are weighted by their mean sample weights.
* p<0.1, ** p<0.05, *** p<0.01

Table A.19. Timing of the rainfall vs intensity of rainfall, spline fuction approach. Knot at -0.6 standard deviations, two lags of monsoon onset. IFLS data.

Dependent variable	Log per capita expenditure							
	All households		**Poor households**		**Middle-class households**		**Rich households**	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Onset								
Prior year								
$spline_1$	0.068	0.037*	0.109	0.031***	0.080	0.036**	0.004	0.071
$spline_2$	-0.077	0.058	-0.086	0.051	-0.146	0.071*	0.015	0.091
Two years ago								
$spline_{21}$	-0.060	0.048	-0.048	0.045	-0.017	0.040	-0.117	0.074
$spline_{22}$	0.079	0.042*	0.081	0.040	0.040	0.047	0.106	0.053*
Intensity								
Prior year								
$spline_1$	0.036	0.081	-0.072	0.076	0.074	0.074	0.063	0.114
$spline_2$	0.005	0.026	0.026	0.029	0.006	0.028	-0.002	0.032
Two years ago								
$spline_{21}$	-0.064	0.060	-0.044	0.063	-0.131	0.068*	0.001	0.078
$spline_{22}$	0.042	0.027	0.039	0.025	0.061	0.025**	0.019	0.042
Observations	11333		3758		3787		3788	
Households	3988		1307		1327		1354	
R^2 (within)	0.129		0.169		0.150		0.111	
Estimated Effects								
Early onset								
Prior year	-0.068	0.037*	-0.109	0.031***	-0.080	0.036**	-0.004	0.071
Two years ago	0.060	0.048	0.048	0.045	0.017	0.040	0.117	0.074
Late onset								
Prior year	-0.077	0.058	-0.086	0.051	-0.146	0.071*	0.015	0.091
Two year ago	0.079	0.042*	0.081	0.040	0.040	0.047	0.106	0.053*

Dependent variable	Log per capita expenditure							
	All households		**Poor households**		**Middle-class households**		**Rich households**	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.
Intensity								
Negative shock								
Prior year	-0.036	0.081	0.072	0.076	-0.074	0.074	-0.063	0.114
Two years ago	0.064	0.060	0.044	0.063	0.131	0.068*	-0.001	0.078
Positive shock								
Prior Year	0.005	0.026	0.026	0.029	0.006	0.028	-0.002	0.032
Two years ago	0.042	0.027	0.039	0.025	0.061	0.025**	0.019	0.042

Monsoon onset is measured in standard deviations from historical mean. Estimated effects of early onset refer to effect on expenditure at -1 standard deviations, while late onset refers to 1 standard deviation. . Estimated effects of negative intensity shock refer to effect on expenditure at -1 standard deviations, while positive shock refers to 1 standard deviation. Additional controls include survey year effects and household-level fixed effects, with split-off households treated as new households. Standard errors are clustered on rain stations, and are robust to heteroscedasticity. Households are weighted by their mean sample weights.

* p<0.1, ** p<0.05, *** p<0.01

Table A.20. Real community rice price (log) per KG, community fixed effect regression.

Dependent variable	Real community rice price (log) per kg
Lagged monsoon onset	0.062
	(0.021)***
Observations	228
Number of communities	128
R-squared (within)	0.71

Additional controls include survey year effects and community fixed effects. Standard errors are clustered on rain stations, and are robust to heteroscedasticity.
* p<0.1, ** p<0.05, *** p<0.01

Table A.21. Real community rice price (logs) per KG, pooled cross section with province fixed effects.

Dependent variable	Real community rice price (logs) per KG
Lagged monsoon onset	0.035
	(0.015)**
Observations	703
R-squared	0.49

Additional controls include survey year effects and province dummies indicating province fixed effects. Standard errors are clustered on rain stations, and are robust to heteroscedasticity.
* p<0.1, ** p<0.05, *** p<0.01

Table A.22. Per capita expenditure, spline function approach, knot at -0.6 standard deviations. SUSENAS data.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	0.011	0.013	0.012	0.015
$Spline_2$	0.002	0.006	0.001	0.006
Two years ago				
$Spline_1$			0.011	0.017
$Spline_2$			-0.002	0.005
Observations	97176		97176	
R^2	0.180		0.180	
<u>Estimated Effects</u>				
Early Onset				
Prior year	-0.011	0.013	-0.012	0.015
Two years ago			-0.011	0.017
Late Onset				
Prior Year	0.002	0.006	0.001	0.006
Two years ago			-0.002	0.005

Onset is measured in standard deviations from historical mean. Data on onset are taken from 46 rain stations. Additional controls include survey year effects and province-level fixed effects and time-invariant household controls, including gender, age and education of the head. Standard errors are clustered on rain stations, and are robust to heteroscedasticity. Households are weighted by their sample weights.

* p<0.1, ** p<0.05, *** p<0.01

Table A.23. Per capita expenditure, SUSENAS data, only IFLS years and provinces.

Dependent variable	Log per capita expenditure			
	One lag		**Two lags**	
	Coefficient	S.E.	Coefficient	S.E.
Prior Year				
$Spline_1$	-0.001	0.019	0.009	0.020
$Spline_2$	0.022	0.026	0.009	0.026
Two years ago				
$Spline_1$			0.053	0.029*
$Spline_2$			-0.011	0.021
Observations	17058		17058	
R^2	0.210		0.211	
<u>Estimated Effects</u>				
Early Onset				
Prior year	0.001	0.019	-0.009	0.020
Two years ago			-0.053	0.029*
Late Onset				
Prior Year	0.022	0.026	0.009	0.026
Two years ago			-0.011	0.021

Onset is measured in standard deviations from historical mean. Data on onset are taken from 46 rain stations. Additional controls include survey year effects and province-level fixed effects and time-invariant household controls, including gender, age and education of the head. Standard errors are clustered on rain stations, and are robust to heteroscedasticity. Households are weighted by their sample weights.

* p<0.1, ** p<0.05, *** p<0.01

Appendix B

Appendix Chapter Three

B1 Variable descriptions and summary statistics

Table B.1. Variable descriptions, chapter three.

Variable	Variable description
Attend	Indicator variable, 1 if child is attending school
Ever attend	Indicator variable, 1 if child has ever attended school
Work2	Indicator variable, 1 if child worked during past 12 months
Work3[c]	Indicator variable, 1 if child worked during past 4 weeks, only year 2000
Wage work	Indicator variable, 1 if child worked for wages during past 4 weeks, only year 2000
Family work	Indicator variable, 1 if child worked on family business during past 4 weeks, only year 2000
Spline1	Lagged monsoon onset < -0.5 standard deviations
Spline2	-0.5 < Lagged monsoon onset < 0.5 standard deviations
Spline3	Lagged monsoon onset > 0.5 standard deviations
Lagged monsoon onset	Monsoon onset previous year to the IFLS survey, in linear from (deviation from the mean)
Monsoon onset in transition year	Monsoon onset in transition year from primary to secondary (deviation from the mean)
Weather risk	The coefficient of variation of the monsoon onset
PCE	The log of household per capita expenditure, excluding the education expenditure
Value of land	Real value of household’s land (in logs)
Share of food	Share of food in the household budget

Variable	Variable description
Age	Age of the child
Age2	Age of the child squared
Female	1 if child female
Head female	1 if household has female head
Religion of the head	Set of dummy variables, reference category is Muslim
Muslim[a]	1 if Muslim
Christian	1 if Christian
Hindu	1 if Hindu
Other religion	1 if other religion
Age of the head	Set of dummy variables, reference category
Head 15-24 years	1 if head below 25 years
Head 25-34	1 if head 25-34 years
Head 35-49	1 if head 35-49 years
Head 50-64	1 if head 50-64 years
Head 65+	1 if head over 65 years
Education of the mother	Set of dummy variables, reference category no education
Mother no education[a]	1 if no formal education
Mother some primary	1 if some primary
Mother primary	1 if completed primary
Mother some Jr	1 if some junior high school
Mother Jr high	1 if completed junior high school
Mother Sr high	1 if some or completed senior high school
Mother university	1 if some university education
Mother edu missing	1 if informal or education information missing
Education of the farther	Set of dummy variables, reference category no education
Father no education[a]	1 if no formal education
Father some primary	1 if some primary

Variable	Variable description
Father primary	1 if completed primary
Father some Jr	1 if some junior high school
Father Jr high	1 if completed junior high school
Father Sr high	1 if some or completed senior high school
Father university	1 if some university education
Father edu missing	1 if informal or education information missing
Maternal orphan	1 if mother has died
Paternal orphan	1 if father has died
Time	Community averaged travel time to school in minutes
Education cost	The log of community average cost of schooling
#Young	Number of children below 6 years in the household
#Schoolage	Number of children 6-16 years in the household
#Adult	Number of adults in the household
#Old	Number of old people in the household
HH farm	1 if household owns a farm business
HH nonfarm biz	1 if household owns a non-farm business
Province	Set of dummy variables, reference category North Sumatra
North Sumatra[a]	North Sumatra
West Sumatra	West Sumatra
South Sumatra	South Sumatra
Lampung	Lampung
Jakarta	Jakarta
West Java	West Java
Central Java	Central Java
Yogya	Yogya
East Java	East Java

Variable	Variable description
Bali	Bali
West Nusa Tenggara	West Nusa Tenggara
South Kalimantan	South Kalimantan
South Sulawesi	South Sulawesi
1993[a]	Year 1993
1997	Year 1997
2000	Year 2000

NA signifies not applicable.

[a]The reference category, which is not included in the regression analysis

[c] Only year 2000

Table B.2. Sample descriptive statistics. School attendance specification for children aged 6-16 years who have not yet completed grade 9 unless stated otherwise.

	Mean	Sd	Min	Max
Attend	0.8037818	0.3971504	0	1
Ever attend	0.9370979	0.2427976	0	1
Work2[b]	0.0440044	0.205121	0	1
Spline1	-0.7653727	0.3915798	-1.983609	-0.5
Spline2	0.2166674	0.3065688	0	1
Spline3	0.013746	0.0694708	0	0.5829886
Monsoon onset in transition year[c]	0.0643102	1.111159	-2.866436	3.799383
Weather risk	0.3257536	0.15782	0.1217305	0.7442609
Share of food	0.6277523	0.1558563	0.0666682	0.9028614
Pce	11.8621	0.6913849	9.277908	16.10229
Value of land (logs)	5.820885	10.70796	-4.60517	20.74421
Age	10.65334	2.967613	6	16

	Mean	Sd	Min	Max
Age2	122.2996	64.26946	36	256
Female	0.4977114	0.5000135	0	1
Head female	0.1154799	0.3196121	0	1
Muslim[a]	0.8932243	0.3088395	0	1
Christian	0.0556764	0.2293044	0	1
Hindu	0.0455466	0.2085076	0	1
Other religion	.0055526	.0743116	0	1
Head 15-25[a]	0.0165829	0.1277071	0	1
Head 25-35	0.1968935	0.3976661	0	1
Head 35-50	0.5175959	0.499709	0	1
Head 51-65	0.2124259	0.40904	0	1
Head 65+	0.0565018	0.2308969	0	1
Mother no education[a]	0.1654536	0.3716034	0	1
Mother some primary	0.3487657	0.4765975	0	1
Mother primary	0.2959406	0.4564815	0	1
Mother some Jr	0.0204097	0.1414024	0	1
Mother Jr High	0.0640054	0.2447717	0	1
Mother Senior High	0.0535754	0.2251864	0	1
Mother university	0.0165829	0.1277071	0	1
Mother edu missing	0.0352668	0.1844602	0	1
Father no education[a]	0.1130037	0.3166092	0	1
Father some primary	0.290613	0.4540623	0	1
Father primary	0.3024687	0.4593443	0	1

	Mean	Sd	Min	Max
Father some Jr	0.0305395	0.1720729	0	1
Father Jr High	0.0830645	0.27599	0	1
Father Senior High	0.0927441	0.2900843	0	1
Father university	0.0321903	0.1765118	0	1
Father edu missing	.0553763	0.2287219	0	1
Maternal orphan	0.029564	0.1693876	0	1
Paternal orphan	0.0492234	0.2163422	0	1
Time	15.27924	7.422775	1	76.33334
Education cost	10.43239	.5799066	7.36781	12.93227
#Young	0.6478577	0.7913116	0	5
#School age	2.355369	1.141398	1	11
#Adult	2.451414	1.124686	0	12
#Old	0.2658513	0.5280009	0	3
HH farm	0.610565	0.4876405	0	1
HH nonfarm biz	0.3685751	0.4824365	0	1
North Sumatra[a]	0.0664816	0.2491314	0	1
West Sumatra	0.0645307	0.2457051	0	1
South Sumatra	.0696331	0.2545372	0	1
Lampung	0.0709837	0.2568073	0	1
West Java	0.1687552	0.3745496	0	1
Central Java	0.1307121	0.3370979	0	1
Yogya	0.0312148	0.1739044	0	1
East Java	0.1374653	0.3443508	0	1
Bali	0.0435957	0.2042015	0	1
West Nusa Tenggara	0.1096271	0.3124361	0	1

	Mean	Sd	Min	Max
South Kalimantan	0.0460719	0.2096487	0	1
South Sulawesi	0.0592031	0.2360133	0	1
1993[a]	0.3201771	0.4665619	0	1
1997	0.3271554	0.4691921	0	1
2000	0.3526675	0.4778182	0	1
N	13327			

[a] The reference category, which is not included in the regression analysis

[b] Only children aged 10-14 years

[c] Only children aged 6-19 years who have completed primary school

Table B.3. Sample descriptive statistics. Child labour specification, year 2000.

	Mean	Sd	Min	Max
Work3	0.1706119	0.3762535	0	1
Wage work	0.0397499	0.1954147	0	1
Family work	0.1424743	0.349614	0	1
Lagged monsoon onset	-0.2116888	0.5732069	-1.756325	1.082989
Share of food	0.6329645	0.1476538	.1274308	.8925965
Pce	11.9879	0.6463436	10.10725	15.14419
Value of land	5.70234	10.44994	-4.60517	20.02765
Age	11.99285	1.419242	10	14
Age2	145.8419	34.10594	100	196
Female	0.4917374	0.5000434	0	1
Head female	0.1281822	0.3343673	0	1
Muslim	0.9030817	0.2959125	0	1
Christian	0.0518088	0.2216904	0	1

Hindu	0.0446628	0.2066086	0	1
Other religion	0.0004466	0.0211336	0	1
Head 15-25	0.0107191	0.1029996	0	1
Head 25-35	0.1406878	0.3477769	0	1
Head 35-50	0.5815096	0.4934216	0	1
Head 51-65	0.1983028	0.3988105	0	1
Head 65+	0.0687807	0.2531374	0	1
Mother no education	0.1549799	0.3619664	0	1
Mother some primary	0.3479232	0.4764179	0	1
Mother primary	0.2979008	0.4574379	0	1
Mother some Jr	0.026351	0.1602128	0	1
Mother Jr Hig	0.073247	0.2605997	0	1
Mother Senior High	0.0602948	0.2380853	0	1
Mother university	0.0200983	0.1403678	0	1
Mother edu missing	0.019205	0.1372756	0	1
Father no education	0.1089772	0.3116802	0	1
Father some primary	0.3010272	0.458807	0	1
Father primary	0.2974542	0.4572402	0	1
Father some Jr	0.0317106	0.1752676	0	1
Father Jr High	0.0884323	0.2839861	0	1
Father Senior High	0.1067441	0.3088566	0	1
Father university	0.0299241	0.1704159	0	1
Father edu missing	0.0357302	0.1856582	0	1
Maternal orphan	0.0276909	0.1641224	0	1

Paternal orphan	0.041983	0.2005952	0	1
Time	16.07606	6.947539	1	50
Education cost	10.38079	0.505855	7.469246	12.52205
#Young	0.5355069	0.6865643	0	5
#School age	2.317552	1.087179	1	7
#Adult	2.376061	1.072309	0	8
#Old	0.2764627	0.542999	0	3
HH farm	0.6150067	0.4867024	0	1
HH nonfarm biz	0.430996	0.4953262	0	1
West Sumatra	0.0625279	0.242166	0	1
South Sumatra	0.0781599	0.268483	0	1
Lampung	0.0710138	0.2569053	0	1
West Java	0.1710585	0.3766442	0	1
Central Java	0.1255025	0.3313618	0	1
Yogya	0.031264	0.1740691	0	1
East Java	0.1357749	0.3426259	0	1
Bali	0.0473426	0.212418	0	1
West Nusa Tenggara	0.1170165	0.3215118	0	1
South Kalimantan	0.0451094	0.2075905	0	1
South Sulawesi	0.0468959	0.2114632	0	1
N	2239			

B2 Estimation results

Table B.4. School attendance for all children aged 6-16 years.

Dependent variable	School attendance		
	Pooled probit	LPM	IVREG; LPM
Spline1	0.003	-0.004	0.000
	(0.017)	(0.016)	(0.022)
Spline2	0.029	0.039	0.054
	(0.028)	(0.026)	(0.036)
Spline3	-0.063	-0.105	-0.120
	(0.140)	(0.109)	(0.096)
Pce		0.034***	0.335**
		(0.009)	(0.138)
Share of food	-0.271***		
	(0.024)		
Age	0.331***	0.385***	0.378***
	(0.013)	(0.016)	(0.017)
Age2	-0.016***	-0.019***	-0.019***
	(0.001)	(0.001)	(0.001)
Female	-0.001	-0.002	-0.006
	(0.008)	(0.007)	(0.007)
Head female	-0.013	-0.004	0.028
	(0.019)	(0.018)	(0.022)
Christian	0.054*	0.056	0.078*
	(0.028)	(0.035)	(0.043)
Hindu	-0.030	-0.016	-0.049
	(0.041)	(0.039)	(0.042)
Other religion	-0.147*	-0.141**	-0.190***
	(0.085)	(0.066)	(0.064)
Head 25-35	-0.003	0.008	0.008
	(0.026)	(0.032)	(0.040)
Head 35-50	0.018	0.026	0.035
	(0.029)	(0.035)	(0.042)
Head 51-65	0.008	0.020	0.044
	(0.028)	(0.034)	(0.041)
Head 65+	-0.036	-0.023	0.005
	(0.045)	(0.044)	(0.051)
Mother some primary	0.049***	0.073***	0.052***
	(0.010)	(0.013)	(0.019)
Mother primary	0.094***	0.123***	0.095***
	(0.010)	(0.014)	(0.022)
Mother some Jr	0.079***	0.115***	0.045
	(0.019)	(0.024)	(0.041)
Mother Jr high	0.096***	0.127***	0.070**
	(0.012)	(0.017)	(0.028)
Mother Sr high	0.074***	0.094***	-0.014
	(0.014)	(0.014)	(0.046)
Mother university	0.083***	0.096***	-0.075

Dependent variable	School attendance		
	Pooled probit	LPM	IVREG; LPM
	(0.030)	(0.017)	(0.073)
Mother edu missing	0.009	0.023	-0.008
	(0.019)	(0.025)	(0.029)
Father some primary	0.018	0.034	0.035*
	(0.016)	(0.021)	(0.020)
Father primary	0.052***	0.070***	0.045*
	(0.014)	(0.018)	(0.026)
Father some Jr	0.071***	0.090***	0.040
	(0.015)	(0.023)	(0.040)
Father Jr high	0.077***	0.095***	0.038
	(0.011)	(0.019)	(0.041)
Father Sr high	0.115***	0.130***	0.057
	(0.009)	(0.016)	(0.047)
Father university	0.088***	0.109***	-0.030
	(0.024)	(0.023)	(0.081)
Father edu missing	0.038**	0.059**	0.036
	(0.018)	(0.028)	(0.039)
Maternal orphan	-0.037*	-0.032	-0.033
	(0.021)	(0.022)	(0.025)
Paternal orphan	-0.059**	-0.064**	-0.052**
	(0.026)	(0.024)	(0.025)
Time	-0.002***	-0.003**	-0.002**
	(0.001)	(0.001)	(0.001)
Education cost	0.020	0.024	-0.026
	(0.015)	(0.014)	(0.036)
#Young	-0.008**	-0.009*	0.037*
	(0.004)	(0.005)	(0.020)
#Schoolage	-0.016***	-0.011**	0.017
	(0.004)	(0.004)	(0.014)
#Adult	0.004	0.007	0.021***
	(0.004)	(0.004)	(0.007)
#Old	0.014	0.016	0.043***
	(0.013)	(0.013)	(0.011)
HH farm	0.005	0.004	-0.006
	(0.012)	(0.011)	(0.014)
HH nonfarm biz	0.008	-0.000	-0.070**
	(0.009)	(0.007)	(0.031)
Observations	13327	13327	13322
F-test 1stage			39.72***

Robust standard errors in parenthesis

Spline function, knots are located at -0.5 and 0.5 standard deviations of monsoon onset. Specifications also include province and year fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.5. School attendance disaggregated by age.

Dependent variable	School attendance			
	Young children aged 6-10 years		Old children aged 11-16 years	
	Pooled probit	IVREG; LPM	Pooled probit	IVREG; LPM
Spline1	0.001	-0.004	0.005	0.009
	(0.017)	(0.015)	(0.022)	(0.043)
Spline2	-0.006	-0.000	0.068*	0.106*
	(0.023)	(0.029)	(0.041)	(0.056)
Spline3	-0.045	-0.077	-0.033	-0.109
	(0.119)	(0.114)	(0.151)	(0.139)
Pce		0.143		0.524***
		(0.107)		(0.198)
Share of food	-0.087***		-0.445***	
	(0.029)		(0.043)	
Age	0.483***	0.820***	-0.058	0.135
	(0.031)	(0.050)	(0.089)	(0.112)
Age2	-0.027***	-0.046***	-0.002	-0.009**
	(0.002)	(0.003)	(0.003)	(0.004)
Female	0.016**	0.016**	-0.020	-0.028*
	(0.007)	(0.008)	(0.013)	(0.015)
Head female	-0.011	0.009	-0.001	0.045*
	(0.017)	(0.022)	(0.023)	(0.024)
Christian	0.016	0.030	0.106***	0.137***
	(0.027)	(0.041)	(0.026)	(0.045)
Hindu	-0.102	-0.090	0.009	0.011
	(0.078)	(0.060)	(0.049)	(0.091)
Other religion	0.029	0.001	-0.287***	-0.284***
	(0.069)	(0.092)	(0.085)	(0.060)
Head 25-35	-0.015	-0.025	-0.013	0.004
	(0.030)	(0.041)	(0.048)	(0.072)
Head 35-50	0.006	-0.003	0.021	0.059
	(0.032)	(0.047)	(0.055)	(0.072)
Head 51-65	-0.035	-0.039	0.030	0.093
	(0.039)	(0.047)	(0.052)	(0.073)
Head 65+	-0.011	-0.013	-0.069	0.008
	(0.034)	(0.042)	(0.080)	(0.088)
Mother some primary	0.027***	0.042**	0.060***	0.047
	(0.010)	(0.020)	(0.015)	(0.029)
Mother primary	0.044***	0.058***	0.144***	0.124***
	(0.007)	(0.018)	(0.017)	(0.036)
Mother some Jr	0.028	0.029	0.144***	0.068
	(0.018)	(0.033)	(0.023)	(0.062)
Mother Jr high	0.056***	0.077***	0.126***	0.064
	(0.007)	(0.023)	(0.026)	(0.047)
Mother Sr high	0.047***	0.042	0.118***	-0.094
	(0.010)	(0.038)	(0.031)	(0.090)
Mother university	0.052***	0.012	0.099	-0.170
	(0.011)	(0.067)	(0.092)	(0.106)

Dependent variable	School attendance			
	Young children aged 6-10 years		Old children aged 11-16 years	
	Pooled probit	IVREG; LPM	Pooled probit	IVREG; LPM
Mother edu missing	-0.003	-0.005	0.027	-0.001
	(0.023)	(0.034)	(0.024)	(0.037)
Father some primary	0.022	0.041*	0.007	0.037*
	(0.013)	(0.025)	(0.023)	(0.022)
Father primary	0.036**	0.049*	0.058***	0.043
	(0.015)	(0.027)	(0.021)	(0.035)
Father some Jr	0.041***	0.047	0.105***	0.057
	(0.008)	(0.031)	(0.031)	(0.057)
Father Jr high	0.040***	0.047	0.126***	0.039
	(0.012)	(0.036)	(0.018)	(0.058)
Father Sr high	0.066***	0.077**	0.159***	0.052
	(0.007)	(0.034)	(0.015)	(0.063)
Father university	0.043***	0.017	0.158***	-0.050
	(0.016)	(0.058)	(0.025)	(0.118)
Father edu missing	0.041***	0.090**	0.034	-0.018
	(0.015)	(0.044)	(0.023)	(0.054)
Maternal orphan	-0.030	-0.021	-0.051**	-0.056
	(0.032)	(0.026)	(0.024)	(0.036)
Paternal orphan	-0.060*	-0.066**	-0.064*	-0.022
	(0.032)	(0.030)	(0.034)	(0.033)
Time	-0.002***	-0.003***	-0.002*	-0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Education cost	0.007	-0.008	0.028	-0.047
	(0.008)	(0.024)	(0.025)	(0.058)
#Young	-0.013***	0.002	-0.004	0.070**
	(0.004)	(0.016)	(0.007)	(0.030)
#Schoolage	-0.016***	-0.007	-0.010*	0.044**
	(0.004)	(0.013)	(0.006)	(0.020)
#Adult	-0.000	0.007	0.011**	0.034***
	(0.004)	(0.007)	(0.005)	(0.008)
#Old	0.002	0.012	0.031*	0.080***
	(0.012)	(0.013)	(0.017)	(0.020)
HH farm	-0.008	-0.009	0.031	0.002
	(0.006)	(0.008)	(0.024)	(0.021)
HH nonfarm biz	0.011	-0.023	0.008	-0.117***
	(0.007)	(0.025)	(0.016)	(0.043)
Observations	6520	6519	6807	6803
F-test 1st stage		24.79***		40.49***

Robust standard errors in parenthesis

Spline function, knots are located at -0.5 and 0.5 standard deviations of monsoon onset. Specifications also include province and year fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors are clustered on rain stations. * p<0.1, ** p<0.05, *** p<0.01

Table B.6. School attendance for children aged 6-19 years who have completed primary school, conditioning on monsoon onset in the transition year from primary to secondary school.

Dependent variable	School attendance after completion of the primary school					
	Start year based on birth cohort			Start year based on reported school starting age		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
Monsoon onset in transition year	-0.028***	-0.023***	-0.021***	-0.019***	-0.013**	-0.010*
	(0.010)	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)
PCE		0.037**	0.291		0.042***	0.325*
		(0.014)	(0.183)		(0.014)	(0.195)
Share of food	-0.649***			-0.679***		
	(0.077)			(0.065)		
Age	0.151**	0.150***	0.176***	0.125*	0.113**	0.131***
	(0.069)	(0.054)	(0.053)	(0.074)	(0.049)	(0.050)
Age2	-0.007***	-0.006***	-0.007***	-0.008***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female	-0.077***	-0.050***	-0.053***	-0.072***	-0.045***	-0.050***
	(0.021)	(0.016)	(0.017)	(0.018)	(0.014)	(0.014)
Head female	-0.010	0.005	0.023	0.018	0.018	0.032
	(0.026)	(0.017)	(0.027)	(0.026)	(0.016)	(0.024)
Christian	0.102	0.073	0.100*	0.151**	0.100*	0.146**
	(0.076)	(0.050)	(0.058)	(0.069)	(0.051)	(0.064)
Hindu	-0.071	-0.039	-0.044	-0.021	-0.009	-0.031
	(0.050)	(0.041)	(0.050)	(0.052)	(0.048)	(0.058)
Other religion	-0.241*	-0.183*	-0.209***	-0.244*	-0.181*	-0.210***
	(0.145)	(0.094)	(0.066)	(0.144)	(0.098)	(0.070)
Head 25-35	-0.054	-0.040	-0.069*	0.003	-0.003	-0.022
	(0.037)	(0.027)	(0.040)	(0.039)	(0.027)	(0.039)
Head 35-50	0.016	0.012	0.014	0.078**	0.051**	0.056*
	(0.034)	(0.023)	(0.031)	(0.037)	(0.023)	(0.029)
Head 51-65	0.022	0.014	0.018	0.076	0.050	0.054
	(0.043)	(0.030)	(0.036)	(0.052)	(0.035)	(0.040)
Head 65+	-0.042	-0.035	-0.025	0.021	0.005	0.024
	(0.055)	(0.035)	(0.042)	(0.059)	(0.038)	(0.044)
Mother some primary	0.109***	0.095***	0.075***	0.138***	0.113***	0.085***
	(0.024)	(0.019)	(0.024)	(0.025)	(0.019)	(0.026)
Mother primary	0.219***	0.187***	0.156***	0.227***	0.189***	0.148***
	(0.029)	(0.022)	(0.027)	(0.031)	(0.023)	(0.031)
Mother	0.323***	0.277***	0.214***	0.344***	0.292***	0.207***

Dependent variable	School attendance after completion of the primary school					
	Start year based on birth cohort			Start year based on reported school starting age		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
some Jr						
	(0.056)	(0.047)	(0.062)	(0.050)	(0.039)	(0.066)
Mother Jr high	0.266***	0.227***	0.173***	0.299***	0.248***	0.180***
	(0.033)	(0.030)	(0.043)	(0.030)	(0.029)	(0.049)
Mother Sr high	0.190***	0.160***	0.051	0.171***	0.144***	0.025
	(0.040)	(0.030)	(0.062)	(0.049)	(0.035)	(0.062)
Mother university	0.324***	0.243***	0.139	0.282***	0.238***	0.099
	(0.044)	(0.047)	(0.086)	(0.072)	(0.050)	(0.097)
Mother edu missing	0.110***	0.103***	0.074*	0.126**	0.113**	0.081
	(0.040)	(0.033)	(0.045)	(0.054)	(0.046)	(0.066)
Father some primary	0.040	0.038	0.043	0.032	0.031	0.038
	(0.038)	(0.028)	(0.027)	(0.041)	(0.030)	(0.030)
Father primary	0.101**	0.089**	0.071*	0.092*	0.081**	0.063
	(0.049)	(0.038)	(0.043)	(0.047)	(0.037)	(0.044)
Father some Jr	0.169***	0.152***	0.109**	0.152**	0.141***	0.087
	(0.057)	(0.044)	(0.049)	(0.062)	(0.044)	(0.061)
Father Jr high	0.219***	0.186***	0.141**	0.209***	0.176***	0.139**
	(0.037)	(0.035)	(0.062)	(0.039)	(0.035)	(0.059)
Father Sr high	0.301***	0.252***	0.192***	0.306***	0.245***	0.179***
	(0.031)	(0.032)	(0.063)	(0.031)	(0.033)	(0.068)
Father university	0.316***	0.257***	0.142	0.293***	0.233***	0.106
	(0.043)	(0.048)	(0.116)	(0.050)	(0.050)	(0.129)
Father edu missing	0.121**	0.105**	0.069	0.083	0.066	0.018
	(0.052)	(0.041)	(0.050)	(0.055)	(0.045)	(0.063)
Maternal orphan	-0.078**	-0.046*	-0.048*	-0.068	-0.036	-0.025
	(0.038)	(0.026)	(0.027)	(0.044)	(0.029)	(0.032)
Paternal orphan	-0.010	-0.013	-0.001	-0.015	-0.013	-0.001
	(0.032)	(0.019)	(0.018)	(0.036)	(0.022)	(0.021)
Time	-0.003**	-0.001	-0.001	-0.002	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Education cost	0.085***	0.076***	0.030	0.091**	0.078***	0.031

Dependent variable	School attendance after completion of the primary school					
	Start year based on birth cohort			Start year based on reported school starting age		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
	(0.030)	(0.023)	(0.048)	(0.039)	(0.027)	(0.051)
#Young	-0.007	-0.011	0.028	-0.012	-0.013	0.032
	(0.014)	(0.011)	(0.032)	(0.013)	(0.011)	(0.036)
#Schoolage	0.004	0.009	0.033*	0.001	0.008	0.035*
	(0.007)	(0.006)	(0.019)	(0.008)	(0.006)	(0.020)
#Adult	0.010	0.008	0.020**	0.014	0.011	0.028**
	(0.008)	(0.006)	(0.010)	(0.009)	(0.007)	(0.012)
#Old	0.012	0.013	0.039*	0.013	0.013	0.039**
	(0.019)	(0.013)	(0.021)	(0.016)	(0.012)	(0.015)
HH farm	0.035	0.019	0.012	0.035	0.016	0.008
	(0.023)	(0.015)	(0.016)	(0.025)	(0.016)	(0.016)
HH nonfarm biz	-0.006	-0.001	-0.055	-0.002	0.001	-0.056*
	(0.014)	(0.009)	(0.035)	(0.018)	(0.012)	(0.034)
Observations	6786	6786	6782	6181	6181	6177
F-test 1st stage			29.52***			31.46***

Robust standard errors in parenthesis

Specifications also include province and year fixed effects and the last two regressions also birth cohort fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.7. The impact of weather risk (coefficient of variation of monsoon onset) on the probability of a child ever attending school.

Dependent variable	Ever attended school					
	All children aged 6-16 years			Young children aged 6-10 years		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
Weather risk	-0.022	-0.060	-0.062	-0.092**	-0.140**	-0.138**
	(0.019)	(0.042)	(0.042)	(0.038)	(0.068)	(0.067)
Pce		0.017***	0.052		0.027***	0.116
		(0.005)	(0.069)		(0.006)	(0.115)
Share of food	-0.038***			-0.083***		
	(0.011)			(0.024)		
Age	0.062***	0.173***	0.173***	0.315***	0.697***	0.699***
	(0.004)	(0.016)	(0.016)	(0.024)	(0.062)	(0.063)
Age2	-0.003***	-0.007***	-0.007***	-0.017***	-0.039***	-0.039***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)
Female	0.004	0.007	0.007	0.014**	0.017*	0.017*
	(0.004)	(0.006)	(0.006)	(0.007)	(0.009)	(0.009)
Head female	-0.000	0.001	0.005	0.003	0.008	0.020
	(0.005)	(0.009)	(0.011)	(0.011)	(0.014)	(0.024)
Christian	0.010	0.018	0.019	0.020	0.025	0.030
	(0.006)	(0.020)	(0.021)	(0.016)	(0.030)	(0.033)
Hindu	0.008	0.027	0.025	0.020	0.027	0.017
	(0.006)	(0.021)	(0.021)	(0.022)	(0.031)	(0.034)
Other religion	-0.013	-0.010	-0.015	-0.016	-0.006	-0.041
	(0.034)	(0.040)	(0.032)	(0.094)	(0.076)	(0.072)
Head 25-35	0.008***	0.034***	0.034***	0.007	0.042	0.039
	(0.003)	(0.011)	(0.011)	(0.026)	(0.052)	(0.048)
Head 35-50	0.018***	0.043***	0.045***	0.026	0.062	0.059
	(0.006)	(0.015)	(0.014)	(0.027)	(0.054)	(0.052)
Head 51-65	0.009**	0.032**	0.034**	0.003	0.032	0.032
	(0.004)	(0.014)	(0.014)	(0.023)	(0.047)	(0.043)
Head 65+	0.007	0.027	0.030	-0.005	0.027	0.034
	(0.005)	(0.018)	(0.019)	(0.027)	(0.047)	(0.043)
Mother some primary	0.008***	0.026***	0.024**	0.016*	0.041**	0.034
	(0.003)	(0.009)	(0.012)	(0.008)	(0.016)	(0.022)
Mother primary	0.017***	0.043***	0.040***	0.033***	0.067***	0.060***
	(0.003)	(0.009)	(0.012)	(0.006)	(0.014)	(0.018)
Mother some Jr	0.008	0.030*	0.022	0.020	0.059*	0.039
	(0.007)	(0.017)	(0.021)	(0.015)	(0.032)	(0.040)
Mother Jr high	0.015***	0.051***	0.044***	0.039***	0.095***	0.078***
	(0.003)	(0.010)	(0.017)	(0.006)	(0.015)	(0.025)

Dependent variable	Ever attended school					
	All children aged 6-16 years			Young children aged 6-10 years		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
Mother Sr high	0.016***	0.053***	0.042*	0.040***	0.096***	0.070*
	(0.003)	(0.010)	(0.024)	(0.007)	(0.019)	(0.037)
Mother university	0.016***	0.042***	0.023	0.036***	0.076***	0.024
	(0.003)	(0.008)	(0.038)	(0.008)	(0.018)	(0.071)
Mother edu missing	-0.010	-0.018	-0.021	-0.032	-0.032	-0.040
	(0.012)	(0.021)	(0.018)	(0.027)	(0.038)	(0.037)
Father some primary	0.005	0.021	0.020	0.009	0.026	0.022
	(0.005)	(0.018)	(0.018)	(0.011)	(0.027)	(0.026)
Father primary	0.009**	0.029*	0.025	0.019	0.039	0.026
	(0.004)	(0.016)	(0.019)	(0.012)	(0.027)	(0.031)
Father some Jr	0.012***	0.039*	0.033	0.025***	0.051	0.033
	(0.003)	(0.022)	(0.028)	(0.009)	(0.033)	(0.043)
Father Jr high	0.012***	0.037*	0.030	0.025**	0.052	0.032
	(0.004)	(0.018)	(0.026)	(0.011)	(0.033)	(0.043)
Father Sr high	0.017***	0.048***	0.039	0.036***	0.064**	0.036
	(0.002)	(0.014)	(0.029)	(0.007)	(0.024)	(0.048)
Father university	0.011**	0.037**	0.019	0.018	0.039	-0.008
	(0.005)	(0.017)	(0.044)	(0.016)	(0.029)	(0.070)
Father edu missing	0.013***	0.056**	0.054**	0.035***	0.097**	0.094**
	(0.003)	(0.024)	(0.026)	(0.007)	(0.043)	(0.045)
Maternal orphan	-0.005	-0.010	-0.009	-0.021	-0.031	-0.023
	(0.007)	(0.010)	(0.009)	(0.025)	(0.030)	(0.026)
Paternal orphan	-0.023**	-0.031*	-0.028*	-0.037	-0.043	-0.039
	(0.012)	(0.016)	(0.017)	(0.031)	(0.034)	(0.032)
Time	-0.000***	-0.002***	-0.001***	-0.001***	-0.002***	-0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Education cost	0.011***	0.021**	0.016	0.017*	0.025*	0.011
	(0.004)	(0.009)	(0.019)	(0.010)	(0.014)	(0.029)
#Young	-0.003**	-0.004	0.002	-0.009***	-0.012**	0.002
	(0.002)	(0.003)	(0.014)	(0.003)	(0.005)	(0.020)
#Schoolage	-0.005***	-0.011**	-0.007	-0.014***	-0.019***	-0.011
	(0.002)	(0.004)	(0.010)	(0.003)	(0.006)	(0.014)
#Adult	0.000	0.001	0.003	0.001	0.004	0.009

Dependent variable	Ever attended school					
	All children aged 6-16 years			Young children aged 6-10 years		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
	(0.002)	(0.003)	(0.005)	(0.003)	(0.004)	(0.009)
#Old	0.004	0.007	0.011*	0.004	0.007	0.015
	(0.004)	(0.007)	(0.006)	(0.010)	(0.014)	(0.011)
HH farm	-0.002	-0.002	-0.003	-0.004	-0.001	-0.003
	(0.003)	(0.006)	(0.006)	(0.007)	(0.010)	(0.010)
HH nonfarm biz	0.005*	0.002	-0.005	0.013**	0.008	-0.011
	(0.003)	(0.006)	(0.014)	(0.005)	(0.007)	(0.025)
Observations	11899	11895	11890	5684	5684	5683
F-test 1st stage			59.36***			38.81***

Robust standard errors in parenthesis

Coefficient of variation of monsoon onset is calculated for the period of 1979-2003. Specifications also include province and year fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.8. Child labour participation (if child worked in the past month), year 2000.

Dependent variable	Child labour (only year 2000)					
	All children aged 6-14 year			Old children aged 10-14 years		
	Pooled probit	LPM	IVREG; LMP	Pooled probit	LPM	IVREG; LPM
Lagged monsoon onset	0.021**	0.027*	0.027**	0.058**	0.051*	0.050**
	(0.011)	(0.014)	(0.014)	(0.027)	(0.026)	(0.025)
Pce		0.023*	0.058		0.036*	0.001
		(0.012)	(0.118)		(0.019)	(0.178)
Share of food	-0.020			0.005		
	(0.022)			(0.049)		
Age	0.049***	-0.050**	-0.048**	-0.038	-0.151	-0.161
	(0.018)	(0.019)	(0.022)	(0.144)	(0.143)	(0.152)
Age2	-0.001	0.004***	0.004***	0.003	0.008	0.009
	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)	(0.006)
Female	-0.004	-0.006	-0.007	-0.012	-0.014	-0.012
	(0.006)	(0.010)	(0.012)	(0.016)	(0.018)	(0.022)
Head female	0.031**	0.042**	0.045***	0.065**	0.067**	0.065**
	(0.015)	(0.019)	(0.016)	(0.032)	(0.029)	(0.027)
Christian	0.022	0.034	0.034	0.063	0.067	0.070
	(0.032)	(0.043)	(0.042)	(0.072)	(0.069)	(0.069)
Hindu	0.014	0.002	-0.002	0.029	0.004	0.007
	(0.073)	(0.094)	(0.094)	(0.157)	(0.150)	(0.145)
Other religion		-0.060	-0.080		-0.260***	-0.259***
		(0.065)	(0.064)		(0.087)	(0.083)
Head 25-35	0.042	0.064	0.064	0.055	0.067	0.065
	(0.045)	(0.040)	(0.040)	(0.081)	(0.062)	(0.057)
Head 35-50	0.024	0.049	0.050	0.042	0.057	0.054
	(0.031)	(0.041)	(0.039)	(0.062)	(0.056)	(0.050)
Head 51-65	0.022	0.042	0.043	0.025	0.041	0.037
	(0.045)	(0.048)	(0.046)	(0.082)	(0.072)	(0.062)
Head 65+	0.061	0.084	0.087*	0.079	0.098	0.093
	(0.071)	(0.059)	(0.051)	(0.108)	(0.093)	(0.074)
Mother some primary	-0.001	0.000	0.000	-0.008	-0.008	-0.007
	(0.014)	(0.023)	(0.022)	(0.030)	(0.034)	(0.033)
Mother primary	-0.023*	-0.027	-0.027	-0.048*	-0.047	-0.046
	(0.013)	(0.022)	(0.022)	(0.028)	(0.033)	(0.032)
Mother some Jr	-0.016	-0.028	-0.027	-0.018	-0.024	-0.016
	(0.025)	(0.045)	(0.044)	(0.071)	(0.079)	(0.076)
Mother Jr high	-0.012	-0.019	-0.024	-0.019	-0.020	-0.014

Dependent variable	Child labour (only year 2000)					
	All children aged 6-14 year			Old children aged 10-14 years		
	Pooled probit	LPM	IVREG; LMP	Pooled probit	LPM	IVREG; LPM
	(0.012)	(0.022)	(0.027)	(0.027)	(0.033)	(0.042)
Mother Sr high	-0.026**	-0.046*	-0.052	-0.074***	-0.096***	-0.086
	(0.012)	(0.025)	(0.032)	(0.022)	(0.034)	(0.056)
Mother university	-0.043***	-0.071*	-0.084	-0.099***	-0.096	-0.083
	(0.011)	(0.035)	(0.059)	(0.037)	(0.064)	(0.094)
Mother edu missing	-0.044***	-0.077**	-0.078**	-0.110***	-0.137***	-0.134***
	(0.010)	(0.032)	(0.031)	(0.028)	(0.049)	(0.042)
Father some primary	-0.024***	-0.043***	-0.043***	-0.044**	-0.053**	-0.055**
	(0.007)	(0.012)	(0.012)	(0.019)	(0.024)	(0.027)
Father primary	-0.013	-0.031	-0.033	-0.024	-0.037	-0.037
	(0.009)	(0.018)	(0.021)	(0.025)	(0.032)	(0.032)
Father some Jr	-0.043***	-0.081***	-0.083***	-0.092***	-0.125***	-0.117***
	(0.007)	(0.020)	(0.021)	(0.022)	(0.044)	(0.042)
Father Jr high	-0.014	-0.032	-0.036	-0.033	-0.043	-0.040
	(0.010)	(0.020)	(0.029)	(0.026)	(0.036)	(0.045)
Father Sr high	-0.031***	-0.057**	-0.065**	-0.066**	-0.077*	-0.074**
	(0.010)	(0.023)	(0.030)	(0.027)	(0.041)	(0.037)
Father university	-0.002	-0.023	-0.043	-0.014	-0.047	-0.026
	(0.021)	(0.028)	(0.066)	(0.040)	(0.057)	(0.105)
Father edu missing	-0.015	-0.039	-0.039	-0.007	-0.028	-0.027
	(0.014)	(0.024)	(0.025)	(0.046)	(0.054)	(0.051)
Maternal orphan	-0.002	-0.018	-0.017	-0.014	-0.014	-0.015
	(0.026)	(0.039)	(0.038)	(0.051)	(0.046)	(0.044)
Paternal orphan	-0.004	-0.000	0.001	-0.008	-0.004	-0.007
	(0.020)	(0.037)	(0.038)	(0.048)	(0.055)	(0.058)
Time	0.001	0.001	0.001	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Education cost	-0.005	-0.012	-0.019	-0.014	-0.023	-0.015
	(0.013)	(0.016)	(0.032)	(0.033)	(0.031)	(0.062)
#Young	-0.001	0.002	0.007	0.000	0.007	0.001
	(0.005)	(0.006)	(0.019)	(0.010)	(0.009)	(0.028)
#Schoolage	0.007	0.010	0.013	0.016*	0.020*	0.017
	(0.004)	(0.007)	(0.011)	(0.009)	(0.011)	(0.016)

Dependent variable	Child labour (only year 2000)					
	All children aged 6-14 year			Old children aged 10-14 years		
	Pooled probit	LPM	IVREG; LMP	Pooled probit	LPM	IVREG; LPM
#Adult	-0.003	-0.003	0.000	-0.005	-0.002	-0.005
	(0.004)	(0.005)	(0.009)	(0.010)	(0.009)	(0.012)
#Old	0.002	0.003	0.005	0.016	0.019	0.016
	(0.009)	(0.014)	(0.021)	(0.019)	(0.023)	(0.033)
HH farm	0.029***	0.036**	0.035**	0.064***	0.060**	0.062***
	(0.010)	(0.016)	(0.014)	(0.022)	(0.025)	(0.021)
HH nonfarm biz	0.038***	0.043***	0.036	0.080***	0.067***	0.075
	(0.010)	(0.012)	(0.032)	(0.021)	(0.021)	(0.054)
Observations	3949	3951	3948	2240	2239	2237
F-test 1st stage			16.02***			33.80***

Robust standard errors in parenthesis

Lagged monsoon onset in linear form. Specifications also include province fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.9. Family work vs. wage work, year 2000, children aged 10-14 years.

Dependent variable	Child labour (family work or wage work)					
	Family Work			Wage Work		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
Lagged monsoon onset	0.036**	0.030*	0.029*	0.032***	0.034**	0.034**
	(0.018)	(0.017)	(0.017)	(0.011)	(0.016)	(0.015)
Pce		0.037**	0.089		0.004	-0.077
		(0.014)	(0.164)		(0.013)	(0.078)
Share of food	-0.020			0.016		
	(0.041)			(0.023)		
Age	0.111	0.033	0.041	-0.119***	-0.214***	-0.230***
	(0.138)	(0.151)	(0.153)	(0.042)	(0.059)	(0.065)
Age2	-0.003	-0.000	-0.000	0.005***	0.010***	0.010***
	(0.006)	(0.006)	(0.006)	(0.002)	(0.003)	(0.003)
Female	-0.013	-0.016	-0.017	0.000	0.000	0.003
	(0.012)	(0.013)	(0.016)	(0.007)	(0.010)	(0.011)
Head female	0.021	0.025	0.028	0.045***	0.053***	0.048***
	(0.032)	(0.032)	(0.025)	(0.015)	(0.017)	(0.014)
Christian	0.085	0.092	0.094	-0.015	-0.014	-0.014
	(0.068)	(0.062)	(0.060)	(0.010)	(0.024)	(0.025)
Hindu	0.056	0.026	0.022	-0.009	-0.015	-0.009
	(0.171)	(0.152)	(0.148)	(0.006)	(0.016)	(0.023)
Other religion		-0.139**	-0.139**		-0.123**	-0.121**
		(0.064)	(0.063)		(0.046)	(0.048)
Head 25-35	0.045	0.049	0.051	0.028	0.042	0.039
	(0.097)	(0.075)	(0.072)	(0.044)	(0.041)	(0.040)
Head 35-50	0.024	0.031	0.036	0.027	0.050	0.043
	(0.071)	(0.062)	(0.056)	(0.026)	(0.043)	(0.044)
Head 51-65	0.005	0.010	0.014	0.042	0.056	0.049
	(0.091)	(0.084)	(0.076)	(0.046)	(0.042)	(0.042)
Head 65+	0.037	0.049	0.056	0.042	0.048	0.037
	(0.113)	(0.102)	(0.087)	(0.054)	(0.043)	(0.043)
Mother some primary	-0.011	-0.014	-0.014	0.005	0.010	0.010
	(0.028)	(0.034)	(0.032)	(0.008)	(0.010)	(0.011)
Mother primary	-0.036	-0.035	-0.035	-0.003	-0.003	-0.002
	(0.028)	(0.036)	(0.035)	(0.006)	(0.010)	(0.011)
Mother some Jr	0.005	0.002	0.005		-0.028**	-0.025*
	(0.071)	(0.079)	(0.081)		(0.011)	(0.013)
Mother Jr high	-0.011	-0.005	-0.012	-0.009	-0.014	-0.002

Dependent variable	Child labour (family work or wage work)					
		Family Work			Wage Work	
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
	(0.027)	(0.035)	(0.042)	(0.007)	(0.011)	(0.017)
Mother Sr high	-0.080***	-0.115***	-0.130***	0.020	0.025	0.049
	(0.018)	(0.037)	(0.048)	(0.032)	(0.026)	(0.042)
Mother university	-0.082***	-0.087	-0.107		-0.009	0.022
	(0.029)	(0.062)	(0.086)		(0.018)	(0.038)
Mother edu missing	-0.086***	-0.109**	-0.114***	-0.021***	-0.031	-0.025
	(0.026)	(0.047)	(0.042)	(0.006)	(0.022)	(0.023)
Father some primary	-0.048**	-0.062**	-0.059*	-0.002	-0.000	-0.004
	(0.020)	(0.028)	(0.030)	(0.007)	(0.013)	(0.014)
Father primary	-0.021	-0.035	-0.037	-0.008	-0.013	-0.010
	(0.027)	(0.036)	(0.037)	(0.008)	(0.017)	(0.015)
Father some Jr	-0.096***	-0.153***	-0.155***	0.001	0.002	0.011
	(0.017)	(0.045)	(0.043)	(0.014)	(0.023)	(0.026)
Father Jr high	-0.026	-0.042	-0.048	-0.007	-0.009	-0.001
	(0.028)	(0.041)	(0.049)	(0.010)	(0.019)	(0.021)
Father Sr high	-0.050*	-0.065	-0.070	-0.017**	-0.030	-0.022
	(0.029)	(0.046)	(0.044)	(0.008)	(0.021)	(0.018)
Father university	0.008	-0.028	-0.060		-0.043	0.006
	(0.040)	(0.052)	(0.107)		(0.028)	(0.040)
Father edu missing	-0.029	-0.058	-0.059	0.010	0.012	0.014
	(0.038)	(0.050)	(0.048)	(0.026)	(0.037)	(0.035)
Maternal orphan	-0.030	-0.026	-0.026	0.009	0.008	0.007
	(0.045)	(0.036)	(0.035)	(0.025)	(0.029)	(0.029)
Paternal orphan	-0.005	-0.002	0.002	-0.011	-0.022	-0.028
	(0.048)	(0.056)	(0.055)	(0.010)	(0.029)	(0.029)
Time	0.002*	0.002	0.002	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)
Education cost	-0.027	-0.035	-0.049	0.011	0.012	0.032
	(0.029)	(0.028)	(0.056)	(0.008)	(0.012)	(0.027)
#Young	-0.004	0.001	0.009	0.004	0.007	-0.006
	(0.011)	(0.010)	(0.028)	(0.003)	(0.006)	(0.010)
#Schoolage	0.004	0.009	0.014	0.010***	0.014**	0.007
	(0.008)	(0.010)	(0.016)	(0.003)	(0.006)	(0.010)

Dependent variable	Child labour (family work or wage work)					
	Family Work			Wage Work		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LMP
#Adult	-0.003	-0.000	0.004	-0.006*	-0.008*	-0.015**
	(0.009)	(0.009)	(0.013)	(0.003)	(0.005)	(0.007)
#Old	0.018	0.022	0.027	0.002	0.004	-0.003
	(0.017)	(0.021)	(0.033)	(0.008)	(0.015)	(0.018)
HH farm	0.077***	0.076***	0.073***	-0.005	-0.006	-0.002
	(0.021)	(0.024)	(0.021)	(0.007)	(0.009)	(0.009)
HH nonfarm biz	0.083***	0.073***	0.062	0.000	-0.001	0.016
	(0.022)	(0.024)	(0.052)	(0.005)	(0.008)	(0.020)
Observations	2240	2239	2237	2096	2239	2237
F-test 1st stage			33.80***			33.80***

Robust standard errors in parenthesis

Lagged monsoon onset in linar form. Specifications also include province fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.10. Child labour specification, lagged monsoon onset interacted with female dummy.

Dependent variable	Child labour		
	Both family and wage	Family Work	Wage work
Lagged monsoon onset	0.051**	0.027	0.035***
	(0.026)	(0.018)	(0.011)
Lagged monsoon*female	0.016	0.023	-0.007
	(0.025)	(0.024)	(0.009)
Share of food	0.005	-0.020	0.016
	(0.050)	(0.041)	(0.023)
Observations	2240	2240	2096

Robust standard errors in parenthesis

Only survey year 2000, pooled probit specification. Specifications also includes same individual, household and community characteristics as in previous tables. Also province fixed effects included. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Table B.11. Child labour participation 1993-2000, children aged 10-14 years.

Dependent variable	Child labour					
	Survey years 1993, 1997 and 2000			Survey years 1997 and 2000		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
Spline1	-0.005*	-0.009	-0.008	0.009	0.019	0.017
	(0.003)	(0.007)	(0.008)	(0.008)	(0.015)	(0.017)
Spline2	-0.015*	-0.027*	-0.037*	-0.024**	-0.047**	-0.061**
	(0.009)	(0.014)	(0.019)	(0.011)	(0.020)	(0.025)
Spline3	0.095***	0.173***	0.186***	0.116***	0.233***	0.267***
	(0.036)	(0.054)	(0.065)	(0.028)	(0.043)	(0.058)
Pce		0.009	-0.093		0.020*	-0.119
		(0.010)	(0.069)		(0.010)	(0.087)
Share of food	0.033***			0.036***		
	(0.009)			(0.009)		
Age	0.009	-0.127*	-0.137*	-0.007	-0.159***	-0.168***
	(0.037)	(0.073)	(0.071)	(0.021)	(0.050)	(0.050)
Age2	0.000	0.006*	0.007**	0.001	0.007***	0.008***
	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Female	-0.005	-0.010	-0.006	-0.004	-0.009	-0.002
	(0.005)	(0.007)	(0.007)	(0.004)	(0.006)	(0.007)
Head female	0.015*	0.025*	0.018	0.013	0.025	0.013
	(0.008)	(0.012)	(0.015)	(0.009)	(0.016)	(0.022)
Christian	0.009	0.015	0.006	0.018	0.033	0.024
	(0.016)	(0.021)	(0.025)	(0.023)	(0.028)	(0.032)
Hindu	-0.015**	-0.030	-0.018	-0.003	-0.008	-0.015
	(0.006)	(0.024)	(0.017)	(0.014)	(0.025)	(0.031)
Other religion	0.174**	0.251***	0.248***		-0.042*	-0.006
	(0.087)	(0.089)	(0.084)		(0.024)	(0.033)
Head 25-35	0.007	0.015	0.021	0.007	0.014	0.019
	(0.018)	(0.026)	(0.028)	(0.014)	(0.029)	(0.032)
Head 35-50	0.002	0.007	0.008	-0.001	0.002	-0.001
	(0.013)	(0.023)	(0.026)	(0.011)	(0.027)	(0.030)
Head 51-65	0.004	0.010	0.006	0.006	0.013	0.009
	(0.015)	(0.023)	(0.024)	(0.012)	(0.026)	(0.026)
Head 65+	0.021	0.034	0.030	0.015	0.028	0.024
	(0.022)	(0.021)	(0.022)	(0.019)	(0.030)	(0.029)
Mother some primary	-0.006*	-0.015*	-0.005	-0.012***	-0.034***	-0.022
	(0.004)	(0.009)	(0.011)	(0.004)	(0.011)	(0.015)
Mother primary	-0.011***	-0.025***	-0.012	-0.011***	-0.034***	-0.019
	(0.003)	(0.007)	(0.013)	(0.003)	(0.010)	(0.017)
Mother some Jr	-0.020***	-0.055***	-0.030		-0.063***	-0.040**
	(0.002)	(0.010)	(0.021)		(0.009)	(0.019)

Dependent variable	Child labour					
	Survey years 1993, 1997 and 2000			Survey years 1997 and 2000		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
Mother Jr high	-0.012**	-0.022*	-0.002	-0.010**	-0.033**	-0.006
	(0.005)	(0.011)	(0.020)	(0.005)	(0.015)	(0.028)
Mother Sr high	-0.008	-0.025*	0.014	-0.011***	-0.046***	0.004
	(0.007)	(0.013)	(0.028)	(0.004)	(0.014)	(0.037)
Mother university		-0.039***	0.018		-0.047***	0.028
		(0.013)	(0.041)		(0.012)	(0.052)
Mother edu missing	0.009	0.017	0.035	0.003	0.004	0.034
	(0.014)	(0.026)	(0.024)	(0.012)	(0.031)	(0.029)
Father some primary	0.000	-0.005	-0.007	0.005	0.009	0.003
	(0.003)	(0.007)	(0.006)	(0.004)	(0.009)	(0.010)
Father primary	-0.014***	-0.029***	-0.023*	-0.009***	-0.019*	-0.014
	(0.004)	(0.010)	(0.013)	(0.003)	(0.010)	(0.012)
Father some Jr	-0.012*	-0.032*	-0.013	-0.012***	-0.032**	-0.015
	(0.007)	(0.017)	(0.020)	(0.004)	(0.013)	(0.015)
Father Jr high	-0.011**	-0.030*	-0.013	-0.007*	-0.016	0.002
	(0.005)	(0.015)	(0.019)	(0.004)	(0.012)	(0.019)
Father Sr high	-0.018***	-0.040**	-0.017	-0.010**	-0.023	-0.003
	(0.005)	(0.015)	(0.022)	(0.005)	(0.014)	(0.019)
Father university	-0.001	-0.024*	0.031	-0.012***	-0.036**	0.034
	(0.012)	(0.013)	(0.037)	(0.004)	(0.015)	(0.044)
Father edu missing	-0.011**	-0.031*	-0.024	-0.005	-0.009	-0.006
	(0.005)	(0.016)	(0.020)	(0.006)	(0.018)	(0.020)
Maternal orphan	0.018	0.028	0.031	0.016*	0.030	0.033
	(0.012)	(0.019)	(0.020)	(0.010)	(0.020)	(0.023)
Paternal orphan	0.016	0.032	0.026	0.005	0.009	0.001
	(0.011)	(0.020)	(0.021)	(0.011)	(0.025)	(0.026)
Time	0.000	0.000	0.000	-0.000	-0.000	-0.001
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Education cost	-0.000	-0.005	0.014	-0.000	-0.010	0.025
	(0.004)	(0.008)	(0.016)	(0.005)	(0.008)	(0.024)
#Young	-0.003	-0.002	-0.018	-0.002	0.001	-0.021
	(0.003)	(0.003)	(0.011)	(0.002)	(0.004)	(0.014)

Dependent variable	Child labour					
	Survey years 1993, 1997 and 2000			Survey years 1997 and 2000		
	Pooled Probit	LPM	IVREG; LPM	Pooled Probit	LPM	IVREG; LPM
#Schoolage	0.004	0.007	-0.003	0.005***	0.012*	0.000
	(0.002)	(0.006)	(0.009)	(0.002)	(0.006)	(0.011)
#Adult	-0.000	0.001	-0.003	-0.000	0.002	-0.005
	(0.002)	(0.004)	(0.006)	(0.002)	(0.004)	(0.006)
#Old	-0.000	0.000	-0.010	-0.005*	-0.006	-0.023*
	(0.003)	(0.005)	(0.008)	(0.003)	(0.005)	(0.012)
HH farm	0.004	0.006	0.007	0.003	0.005	0.006
	(0.004)	(0.008)	(0.009)	(0.004)	(0.009)	(0.010)
HH nonfarm biz	0.002	0.002	0.026	-0.000	-0.003	0.028
	(0.004)	(0.006)	(0.017)	(0.003)	(0.006)	(0.021)
Observations	6321	6406	6404	4455	4652	4650
F-test 1st stage			22.47***			15.87***

Robust standard errors in parenthesis

Child labour refers to the question if child worked in the past 12 months). Spline function approach, knots are located at -0.5, and 0.5. Specifications also include province and year fixed effects. The probit estimates are transformed into marginal effects for continuous variables and impact effects for binary variables, both evaluated at the mean of the explanatory variables. Observations are weighted according to their sampling weights. Standard errors clustered on rain stations.

* p<0.1, ** p<0.05, *** p<0.01

Appendix C

Appendix to Chapter Four

C1 Summary statistics

Table C.1. Sample descriptive statistics for household expenditure and smallholder specifications for Indonesia as a whole.

	Mean	sd	Min	max
Smallholder area	5744.746	16746.58	0	139195
Smallholder production	11142.61	35053.14	0	398553
LOG PCE	12.5744	0.5811897	7.805269	18.49094
Rural	0.6175889	0.4859765	0	1
HHsize	4.058301	1.743745	1	24
Female head	0.1237585	0.329306	0	1
Age	45.25697	13.59616	15	98
Age squared	2233.048	1349	225	9604
No school	0.0951212	0.2933824	0	1
Primary	0.4809648	0.4996378	0	1
Junior high	0.1615586	0.3680456	0	1
Senior high	0.1985837	0.3989341	0	1
University	0.0637717	0.2443459	0	1
Agriculture	0.4784527	0.4995358	0	1
Mining	0.0146229	0.1200377	0	1
Manufacturing	0.0668958	0.2498416	0	1
Electricity	0.0022967	0.0478684	0	1
Construction	0.0451354	0.2076011	0	1
Wholesale	0.1205522	0.3256064	0	1

	Mean	sd	Min	max
Transportation	0.0573389	0.2324892	0	1
Finance	0.0101224	0.1000998	0	1
Public service	0.0991597	0.2988765	0	1
Other or no work	0.1054234	0.3070984	0	1
Employer	0.044471	0.2061392	0	1
Employee	0.2445251	0.4298055	0	1
Casual worker	0.0748371	0.2631284	0	1
Family worker	0.1110323	0.3141723	0	1
2004	0.300265	0.4583734	0	1
2005	0.3064037	0.4609997	0	1
2006	0.3246424	0.4682414	0	1
2007	0.0686888	0.2529244	0	1
N	839918			

Table C.2. Sample descriptive statistics for household expenditure and smallholder specification, only Sumatra and Kalimantan.

	Mean	Sd	Min	Max
Smallholder area	14583.63	24469.31	0	139195
Smallholder production	28309.7	51956.84	0	398553
LOG PCE	12.56563	0.5533115	7.805269	17.51766
Rural	0.6632705	0.4725922	0	1
HHsize	4.176108	1.758919	1	18
Female head	0.112694	0.3162188	0	1
Age	44.19838	13.1966	15	98
Age squared	2127.646	1289.628	225	9604
No school	0.057537	0.2328661	0	1
Primary	0.4776209	0.4994997	0	1
Junior high	0.1885774	0.3911732	0	1
Senior high	0.2181645	0.4130003	0	1
University	0.0581002	0.2339333	0	1
Agriculture	0.5235613	0.4994453	0	1
Mining	0.0256071	0.1579602	0	1
Manufacturing	0.0471691	0.2120008	0	1
Electricity	0.0023676	0.0486003	0	1
Construction	0.041001	0.1982928	0	1
Wholesale	0.1135482	0.3172622	0	1
Transportation	0.0552994	0.228564	0	1
Finance	0.0078981	0.08852	0	1
Public service	0.0954431	0.2938264	0	1
Other or no work	0.0881051	0.2834482	0	1
Employer	0.045764	0.2089733	0	1
Employee	0.2513881	0.4338118	0	1
Casual worker	0.0548166	0.2276222	0	1
Family worker	0.0937471	0.291477	0	1
2004	0.2964774	0.4567047	0	1
2005	0.2992876	0.4579467	0	1

	Mean	Sd	Min	Max
2006	0.3455437	0.4755459	0	1
2007	00.0586914	.2350465	0	1
N	323114			

Table C.3. Sample descriptive statistics for health and smallholder specifications, Indonesia as a whole.

	Mean	sd	min	Max
Smallholder area	5836.318	16932.5	0	139195
Smallholder production	11384.26	35619.66	0	398553
Asthma	0.0178921	0.1325595	0	1
Rural	.6097884	0.4877979	0	1
Female	.499128	0.4999995	0	1
Age	33.40722	16.84675	10	98
Age squared	1399.855	1355.824	100	9604
No school	0.0686751	0.252901	0	1
Primary	0.4510991	0.4976033	0	1
Junior high	0.211398	0.4083003	0	1
Senior high	0.2130318	0.4094502	0	1
University	0.0557959	0.2295273	0	1
Agriculture	0.2528165	0.4346269	0	1
Mining	0.0073879	0.0856351	0	1
Manufacturing	0.0501011	0.2181538	0	1
Electricity	0.0012996	0.0360259	0	1
Construction	0.0221905	0.1473027	0	1
Wholesale	0.0890767	0.2848547	0	1
Transportation	0.0273112	0.1629886	0	1
Finance	0.0064227	0.0798838	0	1
Public service	0.066506	0.2491646	0	1

	Mean	sd	min	Max
Other or no work	0.4768879	0.4994658	0	1
Own toilet	0.620983	0.4851427	0	1
Tap water	0.202392	0.4017831	0	1
Own house	0.8406882	0.3659668	0	1
2004	0.2975256	0.45717	0	1
2005	0.3049548	0.4603885	0	1
2006	0.3244626	0.468174	0	1
2007	0.073057	0.2602303	0	1
N	825670			

Table C.4. Sample descriptive statistics for household expenditure and total area and production specifications in selected provinces in Kalimantan.

	Mean	Sd	min	Max
Total area	35411.81	53769.43	0	173172
Total production	64717.16	99866.58	0	324053
LOG PCE	12.62874	0.5178772	11.16622	16.13105
Rural	0.7311886	0.4433508	0	1
HHsize	4.186834	1.753495	1	18
Female head	0.1058066	0.3075962	0	1
Age	44.17188	12.93793	15	98
Age squared	2118.538	1252.879	225	9604
Primary	0.5188734	0.499654	0	1
Junior high	0.154993	0.3619055	0	1
Senior high	0.1668869	0.3728826	0	1
University	0.0491864	0.2162615	0	1
Mining	0.0247791	0.1554543	0	1
Manufacturing	0.0396465	0.1951313	0	1
Electricity	0.0023127	0.048036	0	1
Construction	0.0375403	0.1900855	0	1
Wholesale	0.0986619	0.2982137	0	1
Transportaion	0.0336582	0.1803516	0	1
Finance	0.0066077	0.0810207	0	1
Public service	0.0840423	0.277457	0	1
Other or no work	0.0791278	0.2699437	0	1
Employer	0.0396878	0.1952287	0	1
Employee	0.2161972	0.4116588	0	1
Casual worker	0.0388618	0.1932695	0	1
Family worker	0.0843727	0.2779516	0	1
2005	0.2923928	0.4548712	0	1
2006	0.577228	0.49401	0	1
2007	0.1303791	0.336727	0	1
N	24214			

Table C.5. Sample descriptive statistics for health and total area and production in selected provinces in Kalimantan.

	Mean	Sd	min	Max
Total area	39642.2	58181.14	0	187511
Total production	71550.65	109841.5	0	368637
Asthma	0.0230755	0.1501439	0	1
Rural	0.722321	0.4478559	0	1
Female	0.4948206	0.4999751	0	1
Age	32.63823	16.21265	10	98
Age squared	1328.102	1270.584	100	9604
No school	0.0786036	0.2691201	0	1
Primary	0.4936919	0.4999621	0	1
Junior high	0.2061566	0.4045458	0	1
Senior high	0.1762396	0.3810253	0	1
University	0.0453084	0.2079804	0	1
Agriculture	0.354922	0.4784915	0	1
Mining	0.0173627	0.1306191	0	1
Manufacturing	0.0294145	0.168966	0	1
Electricity	0.001577	0.0396805	0	1
Construction	0.0200374	0.1401288	0	1
Wholesale	0.0870762	0.2819477	0	1
Transportation	0.0184372	0.1345266	0	1
Finance	0.0048702	0.069617	0	1
Public service	0.0665285	0.2492047	0	1
Other or no work	0.3997743	0.4898537	0	1
Own toilet	0.5857156	0.4925999	0	1
Tap water	0.1798961	0.3841024	0	1
Own house	0.8697568	0.3365721	0	1
2005	0.1933085	0.3948943	0	1
2006	0.3502451	0.4770484	0	1
2007	0.0831104	0.2760501	0	1
N	129358			

C2 Estimation results

Table C.6. Prediction models for total area and production in selected provinces in Kalimantan.

	Area with historical values from Podes	Production with historical values from Podes	Area with district forest area in 2000	Production with district forest area in 2000
Oil palm area in 2002	0.0027***			
	(0.0007)			
Palm oil production in 2002		0.0028***		
		(0.0003)		
Forest area in 2000			0.0037	
			(0.0156)	
Forest area in 2000				-0.0673***
				(0.0156)
Observations	145	93	140	88
F test for instrument	16.46***	121.94***	0.05	18.62***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province year interactions	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Instruments are interacted with predicted province area of palm oil in the area specifications and predicted province production in the production specifications.
Specifications using forest area as an instrument (columns 3 and 4) also include forest area*year interaction terms
Regression coefficients are multiplied by 1000.
* p<0.1, ** p<0.05, *** p<0.01

Table C.7. The impact of smallholder area on LOG PCE, Indonesia as a whole. OLS and IV estimates.

Dependent variable:	LOG PCE			
	OLS, All	IV; All	OLS; Only rural	IV; Only rural
Smallholder area	-0.0000012	-0.0000016	-0.0000011	-0.0000015
	(0.0000009)	(0.0000011)	(0.0000009)	(0.0000010)
Rural	-0.0056643	-0.0056262		
	(0.0139878)	(0.0139646)		
HHsize	-0.1248956***	-0.1248954***	-0.1244789***	-0.1244789***
	(0.0009753)	(0.0009737)	(0.0011793)	(0.0011770)
Female head	-0.0294674***	-0.0294662***	-0.0573325***	-0.0573308***
	(0.0040112)	(0.0040048)	(0.0033054)	(0.0032995)
Age	0.0173659***	0.0173679***	0.0172131***	0.0172159***
	(0.0004898)	(0.0004893)	(0.0004797)	(0.0004791)
Age squared	-0.0001347***	-0.0001347***	-0.0001404***	-0.0001404***
	(0.0000047)	(0.0000047)	(0.0000047)	(0.0000047)
Primary	0.1298889***	0.1298863***	0.1014460***	0.1014436***
	(0.0041794)	(0.0041718)	(0.0039561)	(0.0039482)
Junior high	0.2314918***	0.2314932***	0.1834006***	0.1834155***
	(0.0058269)	(0.0058171)	(0.0055363)	(0.0055247)
Senior high	0.3746682***	0.3746706***	0.2912961***	0.2913199***
	(0.0074358)	(0.0074236)	(0.0064856)	(0.0064767)
University	0.7146036***	0.7145970***	0.5221458***	0.5221436***
	(0.0176378)	(0.0176084)	(0.0101154)	(0.0100966)
Mining	0.1228484***	0.1228288***	0.1066864***	0.1066447***
	(0.0156327)	(0.0156082)	(0.0113402)	(0.0113220)
Manufacturing	0.0720965***	0.0721189***	0.0778137***	0.0777840***
	(0.0062807)	(0.0062664)	(0.0052530)	(0.0052467)
Electricity	0.1937256***	0.1938076***	0.1928970***	0.1929463***
	(0.0134804)	(0.0134597)	(0.0254501)	(0.0254024)
Construction	0.0150510***	0.0150994***	0.0487260***	0.0487468***
	(0.0051563)	(0.0051351)	(0.0051808)	(0.0051668)
Wholesale	0.1562005***	0.1562253***	0.1854159***	0.1854146***
	(0.0050918)	(0.0050799)	(0.0048737)	(0.0048642)
Transportation	0.0706944***	0.0707214***	0.1358958***	0.1358916***
	(0.0056915)	(0.0056730)	(0.0056565)	(0.0056462)
Finance	0.2196850***	0.2197018***	0.2072131***	0.2072214***
	(0.0088533)	(0.0088368)	(0.0172589)	(0.0172295)
Public service	0.1011716***	0.1011904***	0.1558597***	0.1558490***
	(0.0062423)	(0.0062297)	(0.0059940)	(0.0059819)
No work or other	0.0640405***	0.0640496***	0.0320173**	0.0319947**
	(0.0122866)	(0.0122644)	(0.0150085)	(0.0149783)
Employer	0.2392121***	0.2392273***	0.1748344***	0.1748818***
	(0.0093194)	(0.0093044)	(0.0058558)	(0.0058484)
Employee	0.0290314***	0.0290330***	0.0400147***	0.0400220***
	(0.0040260)	(0.0040195)	(0.0052162)	(0.0052051)
Casual worker	-0.0533570***	-0.0533772***	-0.0474412***	-0.0474667***
	(0.0037816)	(0.0037790)	(0.0042947)	(0.0042885)
Family worker no work or other	0.0213651*	0.0213744*	0.0160107	0.0160289

Dependent variable:	LOG PCE			
	OLS, All	IV; All	OLS; Only rural	IV; Only rural
	(0.0117266)	(0.0117068)	(0.0150408)	(0.0150095)
Observations	839918	839918	518724	518724
F-test 1st stage		46.41***		51.85***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

In the IV-regressions historical values from PODES2003 as an instrument. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.

* p<0.1, ** p<0.05, *** p<0.01

Table C.8. The impact of smallholder production on LOG PCE, Indonesia as a whole. OLS and IV estimates.

Dependent variable	LOG PCE			
	OLS, All	IV; All	OLS; Only rural	IV; Only rural
Smallholder production	-0.0000005	-0.0000009*	-0.0000003	-0.0000006
	(0.0000004)	(0.0000005)	(0.0000003)	(0.0000004)
Rural	-0.0056234	-0.0054932		
	(0.0139874)	(0.0139694)		
HHsize	-0.1248942***	-0.1248928***	-0.1244773***	-0.1244757***
	(0.0009753)	(0.0009737)	(0.0011794)	(0.0011774)
Female head	-0.0294821***	-0.0294915***	-0.0573323***	-0.0573275***
	(0.0040102)	(0.0040031)	(0.0033047)	(0.0032990)
Age	0.0173648***	0.0173688***	0.0172097***	0.0172136***
	(0.0004896)	(0.0004894)	(0.0004796)	(0.0004795)
Age squared	-0.0001347***	-0.0001347***	-0.0001403***	-0.0001404***
	(0.0000047)	(0.0000047)	(0.0000047)	(0.0000047)
Primary	0.1298847***	0.1298748***	0.1014491***	0.1014457***
	(0.0041802)	(0.0041726)	(0.0039567)	(0.0039479)
Junior high	0.2314791***	0.2314719***	0.1833819***	0.1834027***
	(0.0058274)	(0.0058167)	(0.0055371)	(0.0055256)
Senior high	0.3746523***	0.3746449***	0.2912499***	0.2912660***
	(0.0074352)	(0.0074206)	(0.0064872)	(0.0064753)
University	0.7145830***	0.7145490***	0.5221821***	0.5222146***
	(0.0176366)	(0.0176056)	(0.0101132)	(0.0100966)
Mining	0.1229404***	0.1229682***	0.1067866***	0.1067790***
	(0.0156283)	(0.0155937)	(0.0113399)	(0.0113123)
Manufacturing	0.0721297***	0.0722138***	0.0778320***	0.0777704***
	(0.0062822)	(0.0062652)	(0.0052543)	(0.0052488)
Electricity	0.1937820***	0.1940357***	0.1928441***	0.1929225***
	(0.0134872)	(0.0134550)	(0.0254591)	(0.0254150)
Construction	0.0150650***	0.0151985***	0.0486913***	0.0487113***

Dependent variable	LOG PCE			
	OLS, All	IV; All	OLS; Only rural	IV; Only rural
	(0.0051553)	(0.0051269)	(0.0051830)	(0.0051718)
Wholesale	0.1562608***	0.1563739***	0.1854302***	0.1854419***
	(0.0050944)	(0.0050799)	(0.0048743)	(0.0048621)
Transportation	0.0707209***	0.0708107***	0.1358755***	0.1358426***
	(0.0056909)	(0.0056627)	(0.0056562)	(0.0056470)
Finance	0.2197977***	0.2199346***	0.2072149***	0.2072394***
	(0.0088498)	(0.0088255)	(0.0172493)	(0.0172153)
Public service	0.1012496***	0.1013620***	0.1558607***	0.1558328***
	(0.0062413)	(0.0062228)	(0.0059958)	(0.0059842)
No work or other	0.0640915***	0.0641570***	0.0320346**	0.0319911**
	(0.0122787)	(0.0122473)	(0.0150037)	(0.0149653)
Employer	0.2391808***	0.2391929***	0.1747631***	0.1748168***
	(0.0093235)	(0.0093070)	(0.0058666)	(0.0058550)
Employee	0.0290267***	0.0290266***	0.0400275***	0.0400614***
	(0.0040249)	(0.0040186)	(0.0052167)	(0.0052040)
Casual worker	-0.0533724***	-0.0534362***	-0.0474479***	-0.0475245***
	(0.0037786)	(0.0037746)	(0.0042902)	(0.0042837)
Family worker no work or other	0.0213652*	0.0213887*	0.0160051	0.0160488
	(0.0117218)	(0.0116960)	(0.0150368)	(0.0149963)
Observations	839918	839918	518724	518724
F-test 1st stage		44.79***		49.97***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

In the IV-regressions historical values from PODES2003 as an instrument. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.

* p<0.1, ** p<0.05, *** p<0.01

Table C.9. The impact of smallholder area on log PCE in Sumatra and Kalimantan. OLS and IV estimates.

Dependent variable	LOG PCE		
	OLS;	IV; PODES	IV; FOREST
Smallholder area	-0.0000010	-0.0000016	-0.0000045*
	(0.0000009)	(0.0000011)	(0.0000025)
Observations	323114	323114	317720
F-test 1^{st} t stage		46.34***	9.59***
Household controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes

Robust standard errors in parenthesis
Household controls include urban/rural, household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.10. The impact of smallholder area on LOG PCE in Sumatra and Kalimantan. OLS and IV estimates, only rural households.

Dependent variable	LOG PCE		
	OLS; Only Rural	IV; PODES; Only rural	IV; FOREST ;Only rural
Smallholder area	-0.0000009	-0.0000015	-0.0000119*
	(0.0000009)	(0.0000010)	(0.0000065)
Observations	214312	214312	212216
F-test 1^{st} stage		51.72***	4.18**
Household controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes

Robust standard errors in parenthesis
Household controls include household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.11. The impact of smallholder production on LOG PCE in Sumatra and Kalimantan, OLS and IV estimates, all households.

Dependent variable	LOG PCE		
	OLS	IV; PODES	IV; FOREST
Smallholder production	-0.0000004	-0.0000009*	-0.0000024*
	(0.0000004)	(0.0000005)	(0.0000014)
Observations	323114	323114	317720
F -test 1st stage		44.93***	7.0***
Household controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes

Robust standard errors in parenthesis
Household controls include urban/rural, household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.12. The impact of smallholder production on LOG PCE in Sumatra and Kalimantan, OLS and IV estimates, only rural households.

Dependent variable	LOG PCE		
	OLS; Only Rural	IV; PODES, Only Rural	IV; Forest, Only Rural
Smallholder production	-0.0000002	-0.0000006	-0.0000032
	(0.0000003)	(0.0000004)	(0.0000024)
Observations	214312	214312	212216
F-test 1st stage		50.18***	3.54*
Household controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
Year fixed effects			
Province*Year Interaction	Yes	Yes	Yes

Robust standard errors in parenthesis
Household controls include household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.13. The impact of smallholder area on asthma. OLS and IV estimates, Indonesia as a whole.

Dependent variable	Asthma			
	OLS, All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Smallholder area	-0.000000077	-0.000000098	-0.000000121	-0.000000217
	(0.000000057)	(0.000000152)	(0.000000080)	(0.000000258)
Rural	0.001756534***	0.001758080***		
	(0.000528876)	(0.000528163)		
Female	-0.004886622***	-0.004886811***	-0.006230871***	-0.006232467***
	(0.000409749)	(0.000409108)	(0.000563123)	(0.000562278)
Age	-0.001027870***	-0.001027884***	-0.001055035***	-0.001054923***
	(0.000076052)	(0.000075927)	(0.000098749)	(0.000098563)
Age squared	0.000027847***	0.000027847***	0.000030525***	0.000030524***
	(0.000001173)	(0.000001171)	(0.000001467)	(0.000001465)
Primary	-0.008753253***	-0.008753871***	-0.006650451***	-0.006655219***
	(0.001353090)	(0.001350815)	(0.001469867)	(0.001467378)
Junior high	-0.010575081***	-0.010575923***	-0.007730931***	-0.007736118***
	(0.001371681)	(0.001369345)	(0.001512481)	(0.001509884)
Senior high	-0.011492281***	-0.011492888***	-0.008667908***	-0.008670951***
	(0.001394463)	(0.001392107)	(0.001546539)	(0.001543837)
University	-0.011539996***	-0.011541187***	-0.009684137***	-0.009692914***
	(0.001480986)	(0.001478346)	(0.001864071)	(0.001861057)
Mining	-0.002847722	-0.002847228	-0.005091228*	-0.005094420*
	(0.002402432)	(0.002398542)	(0.002860485)	(0.002854893)
Manufacturing	0.004351963***	0.004353298***	0.004199438***	0.004203149***
	(0.000809709)	(0.000809120)	(0.001118752)	(0.001117170)
Electricity	-0.003046438	-0.003045983	-0.003698976	-0.003683301
	(0.003022995)	(0.003017566)	(0.005612277)	(0.005601227)
Construction	0.000784609	0.000784983	-0.001133953	-0.001133954
	(0.000960388)	(0.000959134)	(0.001381070)	(0.001378491)
Wholesale	0.002868798***	0.002868842***	0.002658418***	0.002658247***
	(0.000670605)	(0.000669570)	(0.000911374)	(0.000909760)
Transportation	0.001736797*	0.001737330*	0.000704264	0.000703941
	(0.000903877)	(0.000902696)	(0.001255757)	(0.001253127)
Finance	0.003192720**	0.003191868**	0.002144800	0.002149169
	(0.001364593)	(0.001362320)	(0.003305702)	(0.003300029)
Public service	0.001440625**	0.001440003**	0.000433264	0.000430157
	(0.000702872)	(0.000701589)	(0.001051375)	(0.001049259)
No work or other	0.009421569***	0.009421652***	0.011037300***	0.011037199***
	(0.000658996)	(0.000657995)	(0.000803264)	(0.000801673)
Own toilet	-0.005040103***	-0.005038785***	-0.005472076***	-0.005463953***
	(0.000409088)	(0.000407477)	(0.000558241)	(0.000555656)
Tap water	-0.000560063	-0.000559010	-0.000596983	-0.000591168
	(0.000533309)	(0.000532272)	(0.000876993)	(0.000874634)
Own house	-0.002582503***	-0.002582799***	-0.003026005***	-0.003028220***
	(0.000439880)	(0.000439262)	(0.000748691)	(0.000747810)
Observations	825670	825670	503484	503484
F-test 1st stage		49.01***		53.80***
District fixed effects	Yes	Yes	Yes	Yes

Dependent variable	Asthma			
	OLS, All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Dependent variable: dummy variable indicating whether individual suffered from breathing problems. In the IV regressions historical values from PODES as an instrument. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.

* p<0.1, ** p<0.05, *** p<0.01

Table C.14. The impact of smallholder production on asthma, OLS and IV estimates, Indonesia as a whole.

Dependent variable	Asthma			
	OLS, All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Smallholder production	-0.000000009 (0.000000023)	-0.000000027 (0.000000067)	-0.000000017 (0.000000029)	-0.000000062 (0.000000106)
Household controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Observations	825670	825670	503484	503484
F -test 1st stage		45.65***		52.17***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Dependent variable: dummy variable indicating whether individual suffered from breathing problems. In the IV regressions historical values from PODES as an instrument. Household controls include rural/urban and gender of the head and dummy variables indicating whether household has own toilet, tap water and own house. Individual controls include age, education and industry. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.15. The impact of smallholder area on asthma, OLS and IV estimates. Only Sumatra and Kalimantan.

Dependent variable	Asthma			
	OLS, All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Smallholder area	-0.000000043 (0.000000044)	-0.000000112 (0.000000121)	-0.000000090 (0.000000064)	-0.000000264 (0.000000208)
Observations	1073420	1073420	700250	700250
F-test 1st stage		48.85***		53.93***
Household controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

Dependent variable: dummy variable indicating whether individual suffered from breathing problems. In the IV regressions historical values from PODES as an instrument. Household controls include rural/urban and gender of the head and dummy variables indicating whether household has own toilet, tap water and own house. Individual controls include age, education and industry. In education the reference category is no education, in industry the reference category is agriculture.

* p<0.1, ** p<0.05, *** p<0.01

Table C.16. The impact of smallholder production on asthma, OLS and IV estimates. Only Sumatra and Kalimantan.

Dependent variable	Asthma			
	OLS, All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Smallholder production	-0.000000010	-0.000000050	-0.000000025	-0.000000101
	(0.000000019)	(0.000000060)	(0.000000028)	(0.000000096)
Observations	1073420	1073420	700250	700250
F-test 1st stage		45.89***		52.23***
Household controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Dependent variable: dummy variable indicating whether individual suffered from breathing problems. In the IV regressions historical values from PODES as an instrument. Household controls include rural/urban and gender of the head and dummy variables indicating whether household has own toilet, tap water and own house. Individual controls include age, education and industry. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table C.17. Robustness check with rainfall. The impact of smallholder production on log household per capita expenditure, OLS and IV estimate, Indonesia as a whole.

Dependent variable	LOG PCE	
	OLS	IV, Podes
Smallholder production	-0.0000005	-0.0000009*
	(0.0000004)	(0.0000005)
Rainfall	-0.0085450**	-0.0085066**
	(0.0038878)	(0.0038898)
Observations	837170	837170
F-test 1st stage		44.67***
Household controls	Yes	Yes
District fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Province*Year Interaction	Yes	Yes

Robust standard errors in parenthesis
Lagged values of deviation of annual rainfall from its historical mean included as an additional control.
Household controls include household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.18. Robustness check with rainfall. The impact of smallholder area and smallholder production on log PCE, IV estimates, only Sumatra and Kalimantan.

Dependent variable	LOG PCE			
	AREA, IV; PODES	AREA, IV; Forest	PRODUCTION, IV; PODES	PRODUCTION, IV; Forest
Smallholder area	-0.0000016	-0.0000044*		
	(0.0000012)	(0.0000023)		
Smallholder production			-0.0000008	-0.0000023*
			(0.0000005)	(0.0000013)
Rainfall	-0.0124548	-0.0143570*	-0.0113173	-0.0113943
	(0.0080980)	(0.0086547)	(0.0082314)	(0.0089083)
Observations	323114	317720	323114	317720
F-test 1st stage	50.58***	9.65***	44.93***	7.04***
Household controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Province *Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

Lagged values of deviation of annual rainfall from its historical mean included as an additional control. Household controls include household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.

* p<0.1, ** p<0.05, *** p<0.01

Table C.19. The impact of total palm oil area on log PCE in selected provinces in Kalimantan. OLS and IV estimates.

Dependent variable	LOG PCE			
	OLS; All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Total oil palm area	-0.0000081	-0.0000015	-0.0000048	0.0000053
	(0.0000060)	(0.0000215)	(0.0000064)	(0.0000204)
Rural	-0.1532083***	-0.1533002***		
	(0.0383034)	(0.0376476)		
HHsize	-0.1292383***	-0.1292000***	-0.1306497***	-0.1305553***
	(0.0036003)	(0.0035581)	(0.0045508)	(0.0045608)
Female head	-0.0492159***	-0.0490491***	-0.0805764***	-0.0804585***
	(0.0119970)	(0.0116397)	(0.0116686)	(0.0115545)
Age	0.0172670***	0.0172536***	0.0168387***	0.0168308***
	(0.0022441)	(0.0022083)	(0.0024168)	(0.0023800)
Age squared	-0.0001359***	-0.0001358***	-0.0001448***	-0.0001448***
	(0.0000213)	(0.0000209)	(0.0000211)	(0.0000208)
Primary	0.1064695***	0.1063708***	0.0865836***	0.0863817***
	(0.0173626)	(0.0171565)	(0.0169618)	(0.0167169)
Junior high	0.1990454***	0.1988897***	0.1528461***	0.1522497***
	(0.0230067)	(0.0227716)	(0.0205168)	(0.0201350)
Senior high	0.3504356***	0.3505598***	0.2707878***	0.2708126***
	(0.0324450)	(0.0317237)	(0.0282032)	(0.0275539)
University	0.6369796***	0.6366828***	0.4514402***	0.4498097***
	(0.0459421)	(0.0456967)	(0.0410451)	(0.0395059)
Mining	0.1325434***	0.1322394***	0.0726715	0.0737862
	(0.0479453)	(0.0473095)	(0.0469530)	(0.0461782)
Manufacturing	0.0757893**	0.0755635**	0.0359641	0.0365038*
	(0.0349942)	(0.0343292)	(0.0218740)	(0.0214547)
Electricity	0.1704404***	0.1694042***	0.1472547	0.1454658
	(0.0510389)	(0.0500760)	(0.1146475)	(0.1109908)
Construction	-0.0497427**	-0.0503331**	0.0027179	0.0035183
	(0.0221623)	(0.0216566)	(0.0226954)	(0.0215900)
Wholesale	0.1895762***	0.1887812***	0.2277927***	0.2273067***
	(0.0211072)	(0.0203628)	(0.0204843)	(0.0202635)
Transportation	0.0224078	0.0218556	0.0816987**	0.0829037**
	(0.0305636)	(0.0293906)	(0.0376675)	(0.0377358)
Finance	0.2197376***	0.2187426***	0.1014548	0.1047451
	(0.0524008)	(0.0528579)	(0.0734243)	(0.0757542)
Public service	0.0797096***	0.0791599***	0.0917124***	0.0930355***
	(0.0248497)	(0.0242071)	(0.0297234)	(0.0286864)
No work or other	-0.0392218	-0.0396752	-0.0782925	-0.0775087
	(0.0671887)	(0.0670383)	(0.0856610)	(0.0856091)
Employer	0.2702395***	0.2706344***	0.2058356***	0.2047722***
	(0.0216875)	(0.0212777)	(0.0274434)	(0.0270414)
Employee	0.0628784***	0.0630666***	0.1197775***	0.1192262***
	(0.0199656)	(0.0199551)	(0.0244730)	(0.0240669)

Dependent variable	LOG PCE			
	OLS; All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Casual worker	-0.0385402	-0.0385867	-0.0096343	-0.0100053
	(0.0240679)	(0.0235939)	(0.0244982)	(0.0239983)
Family worker no work or other	0.1047354	0.1045631	0.1444057	0.1438388*
	(0.0732878)	(0.0724841)	(0.0865976)	(0.0860851)
Observations	32726	32726	21995	21995
F-test 1st stage		1.57		0.8
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
In the IV regressions historical values from PODES2003 as an instrument. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.20. The impact of total palm oil production on log PCE in selected provinces in Kalimantan. OLS and IV estimates.

Dependent variable	LOG PCE			
	OLS	IV; PODES	OLS; Only Rural	IV; PODES; Only Rural
Total oil palm production	-0.0000007	-0.0000002	-0.0000006	0.0000000
	(0.0000009)	(0.0000008)	(0.0000007)	(0.0000006)
Household controls	Yes	Yes	Yes	Yes
Observations	24214	24214	17705	17705
F-test 1st stage		216.59***		278.44***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis
Household controls include urban/rural, household size, the gender, age, education, industry category and work type of the household head. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed.
* p<0.1, ** p<0.05, *** p<0.01

Table C.21. The impact of total area of palm oil on prevalence of asthma, selected provinces in Kalimantan. OLS and IV estimates.

Dependent variable	Asthma			
	OLS; All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
Total oil palm area	0.0000002	0.0000009**	0.0000002	0.0000009*
	(0.0000001)	(0.0000004)	(0.0000001)	(0.0000006)
Rural	0.0072916***	0.0069361***		
	(0.0017608)	(0.0017639)		
Female	-0.0026238***	-0.0026307***	-0.0042630***	-0.0042751***
	(0.0009576)	(0.0009489)	(0.0012735)	(0.0012572)
Age	-0.0011467***	-0.0011497***	-0.0011578***	-0.0011649***
	(0.0001547)	(0.0001536)	(0.0001742)	(0.0001730)
Age squared	0.0000311***	0.0000312***	0.0000346***	0.0000347***
	(0.0000029)	(0.0000029)	(0.0000032)	(0.0000031)
Primary	-0.0168377***	-0.0167954***	-0.0142712***	-0.0142113***
	(0.0024001)	(0.0023258)	(0.0029466)	(0.0028642)
Junior high	-0.0181877***	-0.0181412***	-0.0147202***	-0.0146326***
	(0.0023316)	(0.0022614)	(0.0029445)	(0.0028738)
Senior high	-0.0190660***	-0.0191264***	-0.0151601***	-0.0151821***
	(0.0024195)	(0.0023569)	(0.0033315)	(0.0032670)
University	-0.0213831***	-0.0213923***	-0.0175994***	-0.0177467***
	(0.0026339)	(0.0025652)	(0.0039590)	(0.0038572)
Mining	0.0017525	0.0013216	0.0026665	0.0023131
	(0.0026232)	(0.0025880)	(0.0036128)	(0.0035934)
Manufacturing	0.0013731	0.0016246	0.0030493	0.0032321
	(0.0023732)	(0.0023227)	(0.0029634)	(0.0028879)
Electricity	0.0049986	0.0066078	0.0243424	0.0262043
	(0.0091638)	(0.0088972)	(0.0171438)	(0.0159936)
Construction	0.0061621***	0.0060357***	0.0041609	0.0040405
	(0.0021427)	(0.0021738)	(0.0034853)	(0.0035063)
Wholesale	0.0024573	0.0024569	0.0015648	0.0015194
	(0.0015254)	(0.0015386)	(0.0023630)	(0.0023620)
Transportation	0.0000881	0.0003577	-0.0023664	-0.0019981
	(0.0030036)	(0.0030611)	(0.0045174)	(0.0046342)
Finance	0.0005726	0.0004991	0.0032658	0.0032551
	(0.0034282)	(0.0032635)	(0.0085039)	(0.0078497)
Public service	0.0020097	0.0021296	0.0023218	0.0023197
	(0.0015785)	(0.0015824)	(0.0027438)	(0.0026824)
No work or other	0.0081088***	0.0080705***	0.0111497***	0.0110571***
	(0.0016311)	(0.0015929)	(0.0021266)	(0.0020960)
Own toilet	-0.0056535***	-0.0055205***	-0.0068390***	-0.0067309***
	(0.0010615)	(0.0010287)	(0.0012466)	(0.0012037)
Tap water	0.0002756	0.0000685	0.0002765	0.0002656
	(0.0014634)	(0.0014921)	(0.0022456)	(0.0021841)
Own house	-0.0019994**	-0.0020678**	-0.0012588	-0.0014482
	(0.0008805)	(0.0008755)	(0.0014263)	(0.0013808)
Observations	181070	181070	119475	119475

Dependent variable	Asthma			
	OLS; All	IV; PODES, All	OLS; Only Rural	IV; PODES, Only Rural
F test 1^{st} stage		13.13***		6.17**
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province*Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

Dependent variable: dummy variable indicating whether individual suffered from breathing problems (Asthma).In the IV regressions historical values from PODES as an instrument. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed

* p<0.1, ** p<0.05, *** p<0.01

Table C.22. The impact of total palm oil production on the prevalence of asthma, selected provinces in Kalimantan. OLS and IV estimates.

Dependent variable	Asthma			
	OLS	IV; PODES	OLS; Only Rural	IV; PODES; Only Rural
Total production	0.000000068* (0.000000035)	0.000000143*** (0.000000045)	0.000000076 (0.000000046)	0.000000125** (0.000000051)
Household controls	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes
Observations	129358	129358	93438	93438
F -test 1^{st} stage		413.32***		507.7***
District fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Province Year Interaction	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis

Dependent variable: dummy variable indicating whether individual suffered from breathing problems. Household controls include rural/urban and gender of the head and dummy variables indicating whether household has own toilet, tap water and own house. Individual controls include age, education and industry. In education the reference category is no education, in industry the reference category is agriculture and in work type the reference category is self employed

* p<0.1, ** p<0.05, *** p<0.01