Climate Adaptive Response of Rice Yield in Vietnam: New Insight through Panel Data Modeling with Heterogeneous Slopes

NGUYEN Viet Thanh,

Faculty of Natural Resource Sciences, School of Business and Science, University of Akureyri, Iceland

Norðurslóð 2, 600 Akureyri, Iceland

email: thanh@unak.is

SIMIONI Michel,

MoISA, INRAE, University of Montpellier, and TSE, University of Toulouse, France

Campus INRAE - Institut Agro, MoISA, Bât. 26 ,Place Viala 34060 Montpellier Cedex 1, France

email: michel.simioni@inrae.fr

and

VO Trung Hung

Faculty of Economics, Thu Dau Mot University, Vietnam

06 Tran Van On, Thu Dau Mot, Binh Duong, Vietnam

email: hungvo2910@gmail.com

September 2022

Climate Adaptive Response of Rice Yield in Vietnam: New Insight through Panel Data Modeling with Heterogeneous Slopes

Abstract

Rice production is central to the Vietnamese economy, not only in terms of contribution to Vietnam's GDP, but also to the food security of its population. However, Vietnam is one of the countries most threatened by climate change in the coming decades, and its rice production in particular. This paper focuses on rice yields and investigates their evolution over time and between provinces over the period 1987-2015, depending on climatic conditions. Special attention is devoted to the impact of heat stress. This impact is measured taking into account the potential adaptation of farmers to these extreme events. To this end, a dynamic production function allowing for spatial and temporal heterogeneity in rice yield responses of to climatic conditions is estimated. Data descriptive analysis shows that the provinces with favorable conditions for rice growth are also those that face the most risk of heat stress. Estimation results show that these provinces adapt to heat stress conditions and that their adaptation effort begins to decrease when heat stress risk becomes too high. Taking adaptation into account then makes it possible to qualify the forecasts made regarding rice yields in the presence of climate change.

Keywords: Rice yield, climate change, adaptation, Vietnam, large panel data, mean observation OLS

JEL Classifications: C23, D24, Q15, Q51, Q54, Q55

1 Introduction

Vietnam is located in Southeast Asia, with a high level of exposure to climate-related hazards and extreme weather and climate events. Vietnam is thus often presented as one of the countries that are most vulnerable to climate change.For instance, the Global Climate Risk Index 2020 ranked Vietnam as the sixth country in the world most affected by climate variability and extreme weather events over the period 1999-2018 (Eckstein et al., 2020). Moreover, at the end of the century, temperature is projected to increase from approximately 1.3°C under a scenario of low greenhouse gases global emissions (RCP2.6) and to approximately 4.2°C under a high emissions scenario (RCP8.5), with faster increases on the North of the country than in the South (Chapter 1 in Espagne et al., 2021). These temperature rises are expected to lead to what might be considered chronic heat stress in some areas, even under lower emissions pathways. Rainfall projections give mixed results. Therefore, in the 1.5°C global warming level, annual rainfall is projected to increase by 0-20% relative to the baseline 1986-2005. These rainfall projections vary significantly between the different Vietnamese regions. For larger global warming levels, significant increases (up to 30-40%) are possible for some locations. These projections suggest a significant increase in the risk of flooding in some parts of Vietnam such as the North East, the Red River Delta and Southern Vietnam. Sea-level rise projections exhibit, on average, an increase from 44cm under RCP2.6 to 73cm under RCP8.5. These projections are more severe for Mekong River Delta which is located in Southern Vietnam, whose estimated average delta plain elevation is around 80cm and which represents 54.47% of Vietnamese rice-planted area. Sae-level rise could reach 84cm under a high emissions scenario, causing large parts of the delta to fall below sea-level by the end of the century.

Despite the rapid rate of industrialization since Doi Moi reforms launched in mid 1980s, agriculture still remains a major economic sector in Vietnam with a GDP contribution of 14.85% in 2020 and employing 18.8 million Vietnamese people over a total population of 97.41 million in 2019. Rice production plays a central role in Vietnamese agriculture. Rice growing activity occupies 63% of total agricultural land, with two major producing areas: Mekong River Delta and Red River Delta which account for 54.47% and 24.80% of rice-planted area, respectively. Rice production reached 43.4 millions of tons in 2019, making Vietnam the fifth rice producing country in the world. Vietnam also ranks as the world's second largest rice exporter.

Rice production is also essential to the livelihoods of 63% of Vietnamese farming households. Moreover, rice production plays an important role in food security in Vietnam. Indeed, rice appears to the main staple food in Vietnam accounting for 29.9% (resp. 25.1%) of total food expenditures of rural (resp. urban) households in 2016 and 51.7% of their total per capita and per day calorie intake (Bairagi et al., 2020).

Agriculture is arguably the sector most affected by climate change as it is directly exposed to climate elements.This is all the more so for rice production which takes place in the open air. Due to the significant contribution of rice production in Vietnamese economy but also its role in guarantying food security of households in this country, it is important to assess its potential evolution faced to climate change. This paper aims to contribute to the literature on the impact of climate change on Vietnamese rice production by investigating the evolution of rice yields over the period 1987-2015 as well as their variability between Vietnamese provinces, as a function of the climatic conditions which these provinces faced during this period. The focus is on the impact of periods of heat stress and potential adaptation of farmers to them at the provincial level. Then, following work on the impact of climate change on crop yields (see, among others, Roberts et al., 2013), the impact of temperatures is summarized in two indicators describing how much temperatures were favorable (growing degree days) or unfavorable (killing degree days) during the growth period of rice, for a given year in each province. Adaptation is captured by estimating a dynamic production function with heterogeneous slopes, i.e. allowing spatial and temporal heterogeneity of rice yield responses to climate variations. The estimation of responses by province and varying in time is made possible by the application of the "mean-observation OLS" (MO-OLS) estimator recently proposed by Keane and Neal (2020a).

The paper is organized as follows. Section 2 provides an overview of the econometric approaches adopted to analyze the impact of climatic conditions on crop yields. Special attention is paid to approaches incorporating farmers' adaptation to these climatic conditions. We then introduce the panel model with heterogeneous slopes recently investigated by Keane and Neal (2020a). Section 3 introduces the data used. This section provides a detailed description of the temporal changes in rice yield distributions and the differences between provinces in these yields. A similar analysis is carried out for the climate indicators, i.e. growing degree days and killing degree days. Section 4 provides the results from estimation of a dynamic panel data model with heterogenous slopes and compare them with those resulting from the estimation of panel models classically used in the literature. This section then focuses on the spatial and temporal patterns of estimated heterogenous parameters associated with killing degree days, drawing new insights from adaptation to heat stress in rice production in Vietnam. Section 5 concludes. Appendices are devoted to the presentation of technical details involved in MO-OLS estimation, growing and killing degree days computation, as well as the formulation of a simple crop model with weather and adaptation.

2 Methodology

Measuring the impact of climate change on agricultural outcome (production, revenue per ha, yield...) is classically based on the estimation of a response function linking this outcome to various weather indicators (see Dell et al., 2014; Kolstad and Moore, 2020, for surveys). For instance, Chung et al. (2015) quantifies the impact of seasonal climate variability on rice yield in the Central Highland of Vietnam from 1986 to 2012. Using annual time-series of yield, temperature and precipitation for two rice growing seasons, they show the favorable impacts of an increase in average minimum temperature and average precipitation on rice yield, while an increase in average maximum temperature reduces rice yield by about 6% and 8% depending on the growing season. However, this approach relies only on short-run fluctuations in weather and, consequently, does not allow for estimation of a long-run climate response including adaptation, but rather a short-run weather response. The challenge is then to capture adaptive behavior not only at the extensive margin (e.g. increasing water use in the short run to fight against heat waves), but also at the intensive margin (e.g. adopting irrigated agriculture practices).

Two main approaches have been proposed in the literature to deal with this challenge. The first one, which was initiated by Mendelsohn et al. (1994), is called the Ricardian approach. This approach is based on the estimation of the impact of climate change on farmland value using cross-sectional data. Mendelsohn et al. (1994) measure variation in climate using 30-year average of weather conditions in different locations in the United States (US). Farmland values are known to reflect farmland profitability within a perfectly competitive market. They reflect farmers' optimization of their production technology or choice according to the climate they have faced. Thus, the main advantage of the Ricardian approach is to capture the long-run equilibrium effects of climate change, incorporating the net benefits of possible adaptation strategies. However, such approach would fail to control for many unobserved omitted variables that are correlated with the climate and also affect farmland values. For instance, lacking observations for factors such as soil quality or access to irrigation infrastructure would bias estimates of climate impacts on farm values.

Extensions of the Ricardian approach suggest using panel data (Schlenker et al., 2006; Massetti and Mendelsohn, 2011). Indeed, the inclusion of individual fixed effects plus region-by-year fixed effects in the regression model allows to control unobserved individual heterogeneity as well as unobserved region specificities and time shocks. To date, Trinh (2018) is the unique study that applies the Ricardian approach to study the impact of climate change on land value in Vietnam, using panel data. This paper uses four waves of household level data from Vietnam Household Living Standard Survey (VHLSS) over the period 2004-2014. Farmland values are approximated by net revenues, with net revenue defined as gross crop revenue (or total sales for each crop) less all cost, divided by agricultural land. Cross-sectional variation in climate is measured using time-average of monthly temperature and precipitation over 65 years (1950-2014). Results show that in the dry season, increases in temperatures are beneficial to all farms in the warmer Southern regions, while increases in precipitation damage only irrigated farms in the Central and Southern regions. The impact of higher temperature in the wet season is similar, except that it will negatively affect net revenue of irrigated farms in the long run. More rainfall in the wet season will increase net revenue only in the North region.

A limitation of the Ricaridan approach is that the classical panel data estimation method does not permit to estimate the effect of the long-run climate averages because these averages have no temporal variation. Indeed, transforming the panel data model using either within or first-difference operators makes it possible to get rid of individual fixed-effects, but at the cost of the disappearance of any time-unvarying variables such as long-run climate averages. Massetti and Mendelsohn (2011) propose the use of the two-step estimation method proposed by Hsiao (2014) to estimate the impact of time-unvarying variables in panel data models with fixed effects (see Trinh, 2018, for an application). This estimation method provides consistent estimates of the impacts of individually time-varying variables in its first step. Nevertheless, estimates of the impacts of time-unvarying variables (long-run climate averages) got in the second step are inconsistent even when the number of individuals tends to infinity, if individual fixed-effects and time unvarying variables are correlated. It is thus easy to imagine omitted variables, such as the location in a mountainous area, that can explain observed variability in farmland prices and are correlated with climate.

A second category of approaches has been recently introduced in the literature. Kolstad and Moore (2020) classify them as emerging hybrid approaches. One of these emerging approaches proposes to model the effect of weather and climate in two steps. First, the linear effect of weather variation is estimated for each location in the panel dataset, which allows the linear response to weather fluctuations to vary across space. Second, the coefficient on weather (from the first step) is modeled as a function of climate and other control variables. This approach exploits two sources of variation in the panel data: (1) the time-series variation from the natural variation of weather within each location, and (2) the cross-sectional variation in mean weather (climate) between locations. Hence, first step estimates measure short-run responses to weather variation in each location, while second step estimates capture the variation in these responses due to adaptation to climate. Indeed, in the long-term, we can think that each location has chosen a production technology adapted to its climate. For instance, hot locations have chosen a technology that performs well in hot temperatures but poorly in cold temperatures and cold locations have chosen the opposite.

A first application of this approach has been proposed by Butler and Huybers (2013). Their application deals with maize yields in US and makes use of county-level data observed from 1981 to 2008. Butler and Huybers (2013) uses the classical framework where the influence of temperature on yield is parametrized by growing degree days (GDD) and killing degree days (KDD) (see, for instance, Roberts et al., 2013). GDD are a commonly used measure for the cumulative warmth a crop has experienced and benefitted over the growing season. By contrast, KDD capture the detrimental effect of high temperatures by accumulating the total number of hours with harmful temperatures over the growing season. As a consequence, in a multiple regression of yields on GDD and KDD, the first measure generally enters positively while the second measure enters negatively. Thus, in a first step, Butler and Huybers (2013) estimate for each county:

$$y_{it} = c_{0i} + c_{1i}t + \beta_{1i}GDD_{it} + \beta_{2i}KDD_{it} + \varepsilon_{it},\ t = 1, \ldots, T \tag{1}$$

where y_{it} denotes maize yield (in logarithm), and GDD_{it} and KDD_{it} denote growing degree days and killing degree days in county i in year t, respectively. The linear time term in t accounts for technological and other steady changes over the time period considered and ε_{it} is the residual error.

β_{1i} and β_{2i} measure yield sensitivity of county i to GDD and KDD, respectively.

In a second step, Butler and Huybers (2013) put the focus on adaptation to warming climate and thus regress estimated county values of KDD sensitivities, or $\widehat{\beta}_{2i}$, on average county values of KDD over the period considered, or $\overline{KDD}_i$, i.e.

$$\widehat{\beta}_{2i} = \alpha_0 + \alpha_1 \log \overline{KDD}_i + \eta_i, \; i = 1, \ldots, N \tag{2}$$

This specification with the regressor expressed in logarithm appears to be the best in terms of fit quality when compared with other linear specifications with transformed covariates. County time-average value of KDD is used as an indicator of the long-run spatial heterogeneity among counties in terms of climate.

The estimated value of parameter α_1 shows a significant negative impact of county time-average value of KDD (in logarithm) on county KDD sensitivity. The relationship between these two variables appears to be concave, i.e. county KDD sensitivity increases up to a given threshold for time-average KDD, from which it starts to remain stable or even to decrease. Or, in other words, hotter counties exhibit significant adaptation to climate over the considered time period, these counties are becoming less and less sensitive to yield losses from heat.

The two-step approach originally proposed by Butler and Huybers (2013) has been recently applied to the measure of the impact of climate change on residential electricity and natural gas consumption in California (Auffhammer, 2022), and on mortality in the U.S. (Heutel et al., 2021). Although the two-step approach uses the temporal and individual dimensions to identify short and long term climate impacts, it does not fully exploit the panel dimension of the data. Moreover, this approach does not allow to investigate adaptation over time within counties. As recently emphasized by Keane and Neal (2020a), due to adaptation to climate, we could expect the function mapping weather conditions into crop yields to exhibit regional and time fixed effects in both intercepts and slopes. Keane and Neal (2020a) then propose an estimation strategy in panel data modeling that addresses adaptation across regions and time in a flexible way. More precisely, they consider estimation of the following panel data model:

$$
\begin{aligned}
y_{it} = \;& c_i + \tau_t + \beta_{1it} GDD_{it} + \beta_{2it} KDD_{it} + \beta_{3it} Prec_{it} + \beta_{4it} Prec_{it}^2 + \varepsilon_{it}, \\
& i = 1, \ldots, N, \; t = 1, \ldots, T
\end{aligned} \tag{3}
$$

where y_{it} denotes crop yield (in logarithm), and GDD_{it} and KDD_{it} denote growing degree days and killing degree days in county i in year t, respectively. They add precipitations in county i in year t, denoted by $Prec_{it}$, and its squared value, or $Prec_{it}^2$, in order to investigate a potential nonlinear impact of precipitations on crop yield (see, among others, Burke and Emerick, 2016). This model generalizes the classical two-way fixed effects specification, with c_i and τ_t denoting the individual and time fixed effects, respectively, by considering both spatial and temporal heterogeneity in the slope coefficients, or β_{kit}, $k = 1, \ldots, 4$. This approach is more general and flexible than those proposed by Butler and Huybers (2013). This approach does not assume a specific form of non-linearity for the KDD coefficient such as Eq. (2). Instead, slope heterogeneity is allowed to be correlated with the regressors. The nature of the relationship between β_{2it} and KDD_{it} can be assessed in a second step by regressing estimates of β_{2it} provided by estimation of Eq. (3), on KDD_{it}.

Restrictions must be imposed in order to estimate coefficients in Eq. (3). Indeed, this equation involves more coefficients to be estimated than data points. Keane and Neal (2020a) then recommend to restrict attention to additive heterogeneity across the region and time dimensions, i.e.

$$
\beta_{kit} = \beta_k + \lambda_{ki} + \theta_{kt}, \; k = 1, \ldots, 4 \tag{4}
$$

As a consequence, each region's relative sensitivity to weather is assumed to be fixed over time. Moreover, time effects shift all region's sensitivities up or down to the same degree.

The panel data model defined by Eq. (3) and (4) can be estimated using the "mean observation OLS" (MO-OLS) procedure developed by Keane and Neal (2020a). The MO-OLS estimator is constructed by first running pooled OLS to obtain $\widehat{\beta}$, then running regressions by region to collect $\widehat{\beta}_i$, $i = 1, \ldots, N$, and lastly a set of regressions by year to collect $\widehat{\beta}_t$, $t = 1, \ldots, T$. A biased preliminary estimator for each β_{it} is given by $\widehat{\beta}_{it} = \widehat{\beta}_i + \widehat{\beta}_t - \widehat{\beta}$. Keane and Neal (2020a) show how the bias can be calculated to arbitrary accuracy and removed using an iterative procedure. Specifically, they show that the expression for the bias is a Cauchy sequence such that the new bias introduced in each iteration gets smaller and smaller. Details on MO-OLS estimation are provided in Appendix A.

Keane and Neal (2020b) compare the predictive performances of MO-OLS estimator with two other approaches: classical two-way fixed effect panel data estimator and the nonparametric technique of deep neural networks (DNNs). They show that both DNNs and MO-OLS estimators outperform classical two-way fixed effect panel data estimator for predicting yields, both in an exercise using the US county-level corn yield data from 1950 to 2015 and a Monte-Carlo cross-validation exercise. Moreover, MO-OLS estimation substantially outperforms the two other approaches in forecasting yield in a 2006-2015 sample.

3 Data

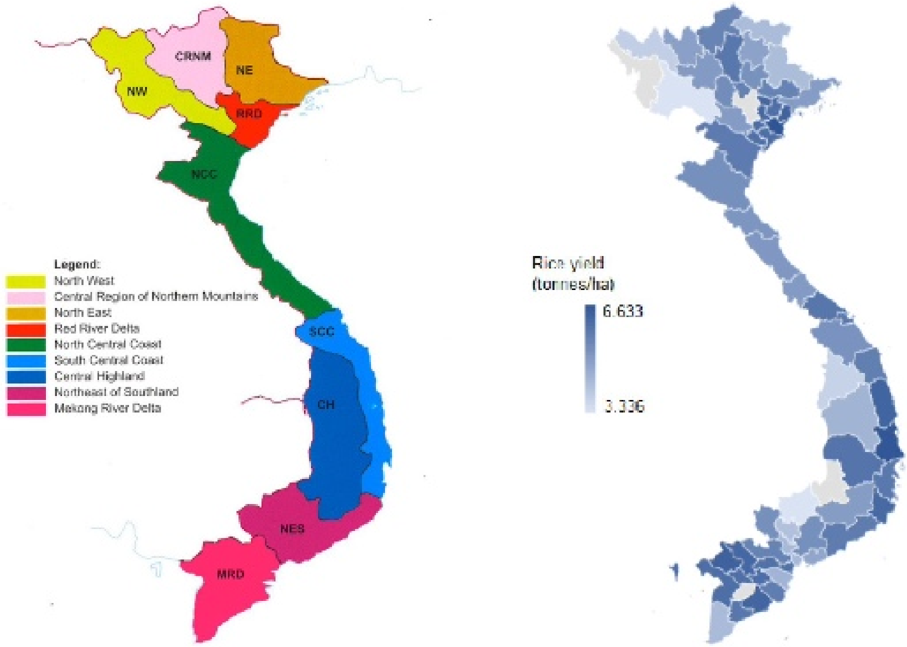
3.1 Rice data

The rice yield database is obtained from the International Rice Research Institute[1], containing information on annual agricultural statistics such as rice production, area harvested, rice yields across 64 sub-national units nationwide from 1987 to 2015. In this period, Vietnam embarked on major structural reforms, or Doi Moi reforms, which gradually converted the country from a highly centralized planned economy to a socialist-oriented market economy with more openness to international trade.

The sub-national units refer to the first tier administrative units that are made up of municipalities and provinces in Vietnam. Our time period of study involved splitting of divisions with high population density and merging of geographically adjacent units. From 1987 to 2003, Vietnam was made up of 61 administrative units. In 2003-2004, Dien Bien was set apart from the remainder of Lai Chau province; Dak Lak was subdivided into Dak Lak and Dak Nong; Hau Giang was seperated from Can Tho city. Hence, from 2004-2007, there were 64 administrative units in Vietnam. Since 2008, the government incorporated Ha Tay province into Hanoi city, resulting in 63 administrative units left in

[1]IRRI is a research organisation that promotes agro-research and development, faring in reviving rice seedlings and improving crop yield in the world. Information about the institute can be found at https://www.irri.org/

Figure 1
Map of rice yield distribution (tonnes/ha) in 2015

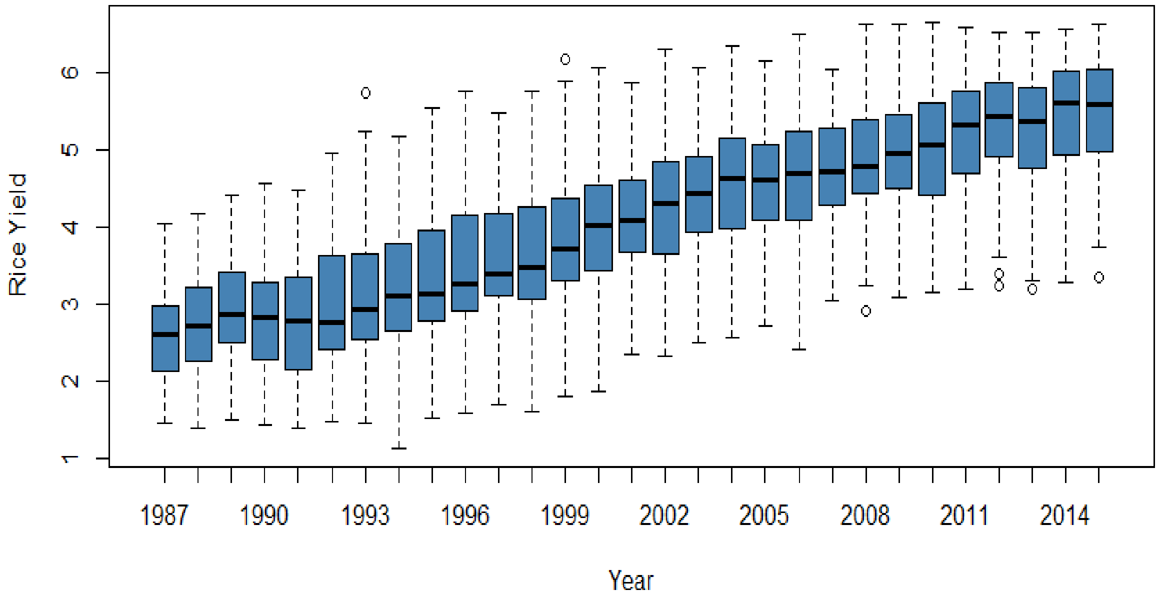


Note: The map on the left shows the nine agro-ecological zones of Vietnam marked by their particular types of soil, topography and climate context. The map on the right displays rice yields across 60 provinces in VN. The grey area is where rice yields were not available. (Source: Authors' own elaboration from IRRI's data).

Vietnam. To have a consistent sample size, we construct spatially consistent geographic units over the study period. In split or merger cases, new geographical entities are created with the rice production and area harvested to be combined from the corresponding units split or merged. Rice yields for these new entities are obtained by dividing the modified rice production by the area harvested. Therefore, there are 60 administrative units presented in our panel data.

Figure 1 maps the rice productivity across the 60 provinces under study in 2015. Rice cultivation is mainly concentrated in Mekong River and Red River Delta. The most productive rice growing area is Thai Binh at a record of around 6.64 tons per hectare. Rice farming is also heavily concentrated along the coast with South Central Coast as another rice production heartland of Vietnam. The area of light blue signifies poor productivity for rice in Vietnam and can be found in landlocked regions. The provinces with the lowest rice yield of Vietnam is Son La in the Northwest of Vietnam. Rice production in rural areas is apparently higher than that of the urban zones because the rural population

Figure 2
Rice yield distributions from 1987 to 2015

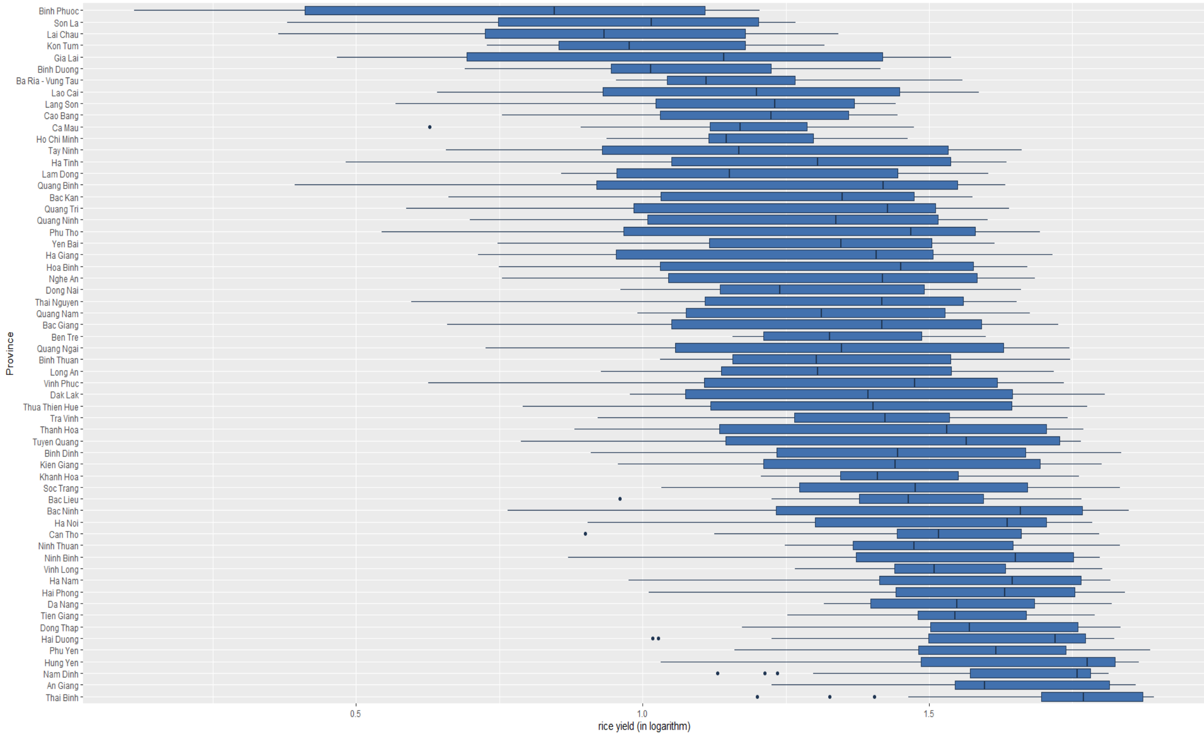


still depends on the agricultural sector for livelihood.

Figure 2 shows the evolution of rice yields over the period from 1987 to 2015. After stagnation between 1987 and 1992, rice yields have grown steadily since 1992, and this growth has affected all Vietnamese provinces. It is then interesting to see if they have all progressed in the same way or if some of them have seen their yields increase while others have not. Figure 3 reports the distributions of rice yields for the 60 provinces. Provinces are ordered according to rice yields. Figure 3 shows a stability over time in order between provinces, the most productive remaining the most productive and so on, and this even if the respective positions of certain neighboring provinces in terms of yield may have changed certain years. This descriptive analysis therefore highlights a persistence in provinces' rice yields, i.e. provinces tend to "stick" with their previous position in the ranking, which econometric modeling must take into account.

Figure 3

Rice yield distributions by province



3.2 Climate data

Weather data used in this study are daily temperatures and precipitations. Temperature data comes from the Climate Prediction Center (CPC) database developed by the National Oceanic and Atmospheric Administration (NOAA). It provides historical data on daily maximum and minimum temperature for a grid of 0.50×0.50 degree of latitude and longitude.[2] Precipitation data comes from Asian Precipitation-Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) project.[3] The rainfall dataset has mean daily precipitation (millilitre per day) gridded at a 0.25×0.25 degree resolution used as a proxy for moisture conditions. The time units are days from January 1st 1987 to December 31st 2015.

Geo-spatial interpolation method was applied to have temperature estimates on a 0.25x0.25 degree

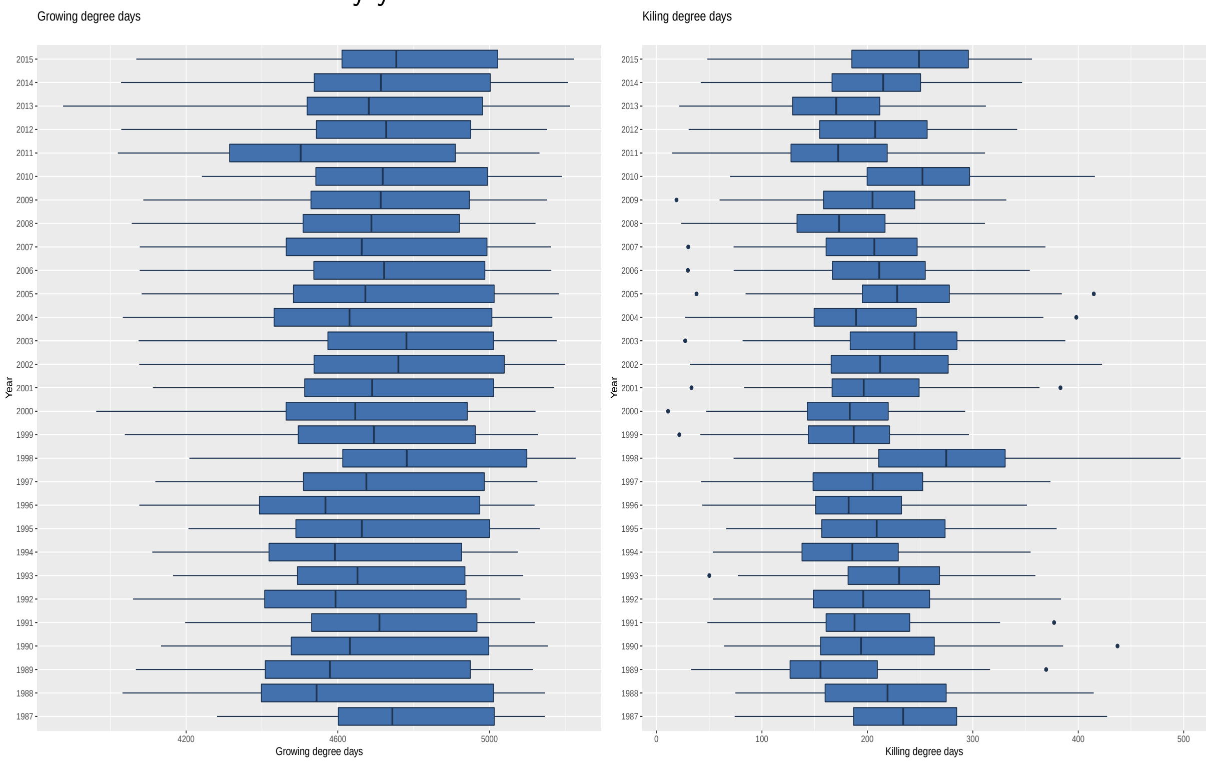
[2] The Global Positioning System (GPS) coordinates utilized in our study range from 104.875 °E to 109.375 °E for longitude and from 8.625° N to 23.125 °N.

[3] Netcdf format data can be retrieved from `https://www.cpc.ncep.noaa.gov` and / `http://aphrodite.st.hirosaki-u.ac.jp/`

resolution grid. Indeed, the original data detailed on a 0.5x0.5 degree resolution grid does not sample all Vietnamese provinces. Generalized additive models (Wood, 2017) were estimated in order to capture the complex nonlinear relationship between temperature and location using original data. Fitted models were used to predict temperature data over the finest grid.[4]

Figure 4

GDD and KDD distributions by year



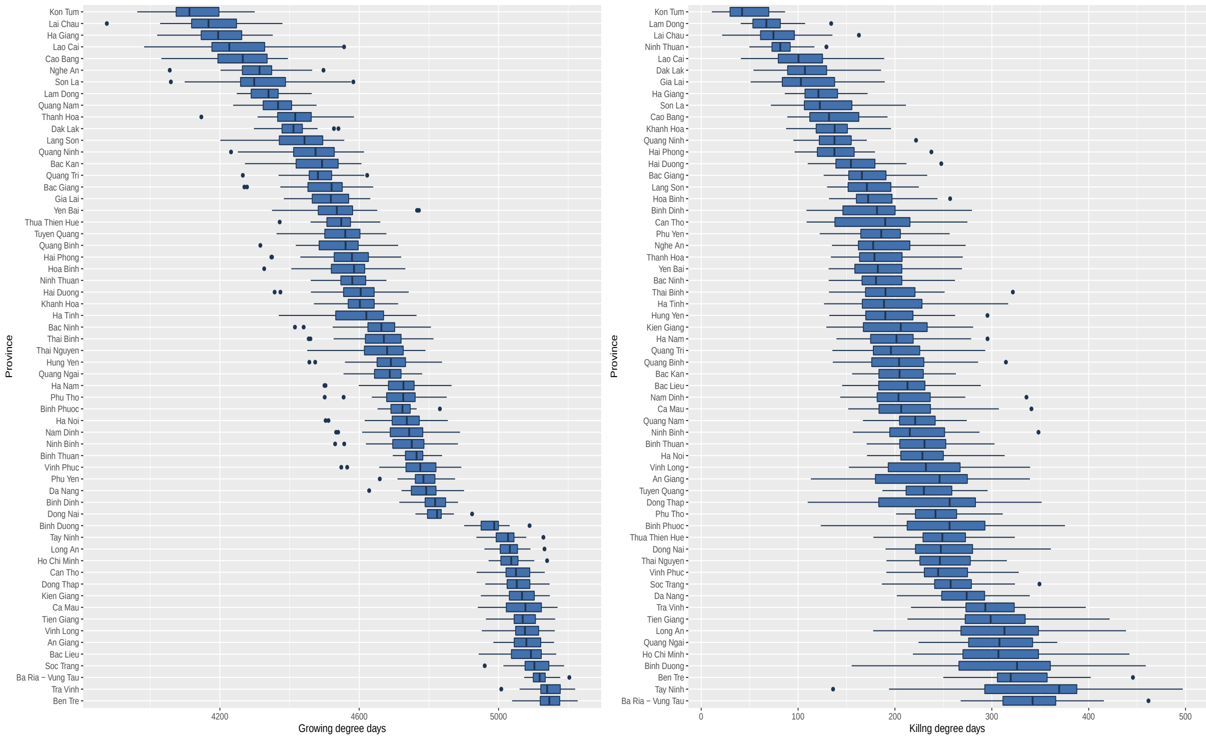
The interpolated weather data were then up-scaled to match the yield measurement spatial unit. Administrative boundaries were first overlaid with the gridded weather dataset. Then, for aggregated maximum (or minimum) temperature of a particular province, we found the maximum (or minimum) record of all observed maximum (or minimum) daily temperature values gridded within that province. We proceeded similarly for precipitation, i.e we took a simple average of observed precipitations within the provinces.

In line with the current literature on climate change impact on crop yields (Roberts et al., 2013), we finally computed growing degree days (GDD) and killing degrees days (KDD) for each province and

[4] More information on the chosen interpolation strategy can be found at `https://swilke-geoscience.net/post/spatial_interpolation/`

Figure 5

GDD and KDD distributions by province



year to assess the effect of weather conditions during the growing season on yields. Computational details on GDD and KDD computation are provided in Appendix B.

The changes in the distributions of GDD and KDD over the studied period do not appear to be characterized by any identifiable pattern as shown in Figure 4 . The year 1998 stands out from the others because several provinces show very high values for Killing Degree Days. Indeed, this year was characterized by a large decrease in rainfall during the dry season (only about 30-70% of the annual average), and prolonged heatwave, all this causing severe droughts in main rice producing provinces.

On the other hand, the analysis of the distributions of GDD by province (see Figure 5) shows a clear ranking of the provinces in terms of rice growing conditions. A similar fact is also observable for the distributions of KDD by province. Although the two rankings does not fit perfectly, the Spearman correlation between them is equal to 0.731 and is significantly different from zero with a p-value much smaller than usual significance levels. The Vietnamese provinces with the most favorable climate for the growth of rice are also those that experience the greatest risks in terms of unfavorable temperatures for this growth, and vice versa. This observation is corroborated when looking at the spatial distribution of average values of GDD and KDD (see Figure 6).

Figure 6
Spatial distributions of mean GDD and KDD in Vietnam

To sum up, Table 1 gives summary statistics for all the variables used below.

Table 1
Descriptive statistics for variables used for econometric analysis

Variables	Mean	Standard Deviation	Minimum	25% Percentile	75% Percentile	Maximum
Rice Yield	4.048	1.205	1.120	3.077	4.969	6.637
GDD	4,687.027	289.332	3,875.414	4,479.303	4,964.526	5,227.362
KDD	206.485	78.791	10.976	153.596	256.828	497.137
Precipitation	1,331.774	272.169	246.384	1,148.470	1,509.492	2,261.893
Crop acreage	118,296.000	121,029.900	5,400	43,775	150,025	770,400
Observations	1740					

Note: Precipitation is the simple average of precipitation values within province.

4 Results

4.1 Classical panel data models

Table 2 reports results from estimation of the three following models:

$$
\begin{aligned}
y_{it} &= c + \rho y_{i,t-1} + \beta_1 GDD_{it} + \beta_2 KDD_{it} + \beta_3 Prec_{it} + \beta_4 Prec_{it}^2 + \varepsilon_{it} && (5)\\
y_{it} &= c_i + \tau_t + \rho y_{i,t-1} + \beta_1 GDD_{it} + \beta_2 KDD_{it} + \beta_3 Prec_{it} + \beta_4 Prec_{it}^2 + \varepsilon_{it}, && (6)\\
y_{it} &= c_i + \tau_t + \rho y_{i,t-1} + \beta_1 GDD_{it} + \beta_{20} KDD_{it} + \beta_{21} \left(\log(KDD_{it}) \times KDD_{it} - KDD_{it}\right) \\
&\quad + \beta_3 Prec_{it} + \beta_4 Prec_{it}^2 + \varepsilon_{it}, && (7)\\
&\quad i = 1, \ldots, N,\ t = 1, \ldots, T
\end{aligned}
$$

These three models integrate the lagged rice yield value (in logarithm) in addition to classical weather variables, i.e. KDD_{it} and $Prec_{it}$, in the set of regressors. The objective is to capture the strong temporal dependence observed in rice yields highlighted in the descriptive presentation of the data. In other words, the lagged rice yield value captures state dependence as the assumption that rice yield is redrawn randomly in each period is inconsistent with evident shown in the data.

Model (5) is the classical Pooled-OLS model that does not take into account any unobserved individual or time heterogeneity while model (6), or Fixed-Effect model, captures it by adding individual and time fixed effects in its specification. As shown in Keane and Neal (2020a) (see also Appendix C), this last specification can be deduced from a simple formalization of the rice production function that does not take into account any adaptation to climate change from farmers. This simple modeling can be generalized in order to incorporate adaptation behavior of farmers to extreme temperatures. The resulting prediction of a log linear relationship in farmers' high temperature response to KDD leads to the model (7) which includes as an additional regressor a nonlinear function of KDD to capture adaptation to high temperatures. β_{20} is thus expected to be negative and β_{21} positive.

A first striking result appears in Table 2: the coefficient associated with lagged rice yield value is found to be positive and significantly different from zero (at 1% significance level), regardless of the estimated model. The expectation that Pooled-OLS estimator overestimates the true coefficient on

Table 2
Classical panel data estimates of the impacts of temperature and precipitation on rice yields in Vietnam

	Pooled-OLS	**FE-OLS without adaptation**	**FE-OLS with adaptation**
Lagged rice yield	0.93707***	0.65740***	0.65998***
(in logarithm)	(0.00575)	(0.02516)	(0.02514)
GDD	0.00003***	-0.00006	-0.00002
	(0.00001)	(0.00004)	(0.00004)
KDD	-0.00011***	-0.00005	-0.00211**
	(0.00003)	(0.00009)	(0.00094)
log(KDD) $\times$ KDD - KDD	—	—	0.00037**
			(0.00017)
Precipitation	0.19831**	0.17185*	0.17533*
	(0.07591)	(0.08930)	(0.09014)
Precipitation$^2 (\div 1000)$	-0.08174***	-0.08213***	-0.08396***
	(0.02803)	(0.03073)	(0.03102)
Constant	-0.10187	—	—
	(0.06997)		
Observations	1680	1680	1680
R squared	0.928	0.912	0.912
Fixed Effects	No	Province, Year	Province, Year

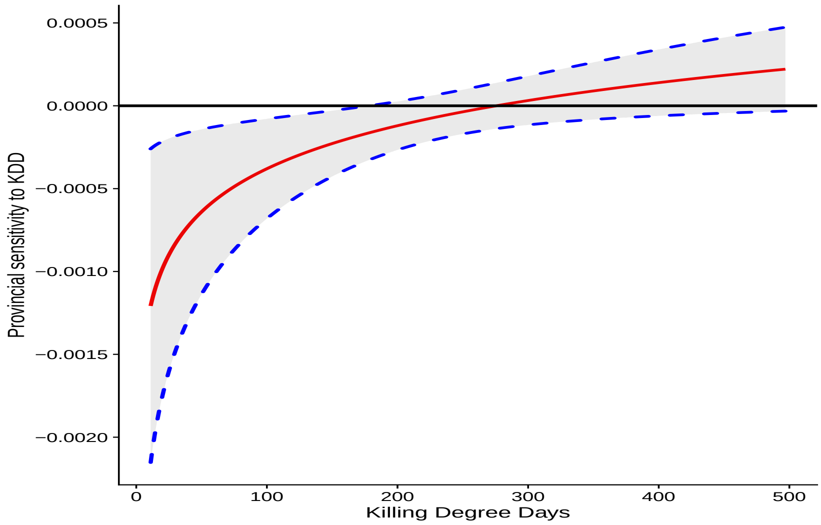
Note: Standard errors are reported in parentheses, and are clustered at the province level.

the lagged rice yield value, whereas the Fixed effects estimator will underestimate it (Bond, 2002), is satisfied. Hence, the true value of ρ lies between 0.66 and 0.94. As a consequence, rice yields exhibit large state dependence, even when unobserved individual heterogeneity is taken into account.

The results of the estimation of the three models show a positive and significant effect of precipitation on rice yield. However, this effect becomes less and less pronounced as the amount of precipitation increases. The results regarding the effect of precipitation are similar whether or not we take into account the presence of individual and temporal fixed effects or the possibility of farmers' adaptation.

The two Fixed-Effect models (6) and (7) do not exhibit a significant and positive relationship between rice yield and GDD, while Pooled-OLS model (5) does. Results regarding the impact of high temperatures are more mixed. A significant and negative impact is shown when estimating the

Figure 7
Estimated marginal effect of KDD on rice yields using FE-OLS with adaptation estimates (with 95% confidence interval)



Pooled-OLS model, while the impact cannot be disentangled from zero when estimating the Fixed-Effect model without adaptation. The introduction of the nonlinear regressor in model (7) makes the direct impact of KDD negative and significant, as expected. In turn, the added regressor is positive and highly significant. Accordingly, as KDD increases, the total marginal effect of KDD on rice yield gets smaller as follows:

$$\frac{\partial y_{it}}{\partial KDD_{it}} = -0.00211 + 0.00037 \ln\left(KDD_{it}\right). \tag{8}$$

Figure 7 reports this estimated marginal effect drawn as a function of KDD. In accordance with the economic modeling of adaptation behaviors proposed by Keane and Neal (2020a) (see Appendix C), this function is increasing and concave. As the climatic conditions for rice production worsen in terms of high temperatures, farmers are making an increasing effort to adapt to these conditions. This effort seems sufficient to annihilate any impact of high temperatures from a threshold of KDD around 200 degree days.

Table 3
MO-OLS estimates of the impacts of temperature and precipitation on rice yields in Vietnam

	Mean	**Weighted Mean**	**Median**	**Standard Deviation**	**25% Percentile**	**75% Percentile**
Lagged rice yield (in logarithm)	.369017*** (.0304878)	.3732414	.3797735	.1989875	.2411912	.5178423
GDD	.0005283*** (.0000946)	.0004151	.0003268	.0007192	.0000527	.0008113
KDD	-.0006413*** (.0001617)	-.0003812	-.0003281	.0011987	-.0009688	.0000899
Precipitation	.3980065** (.1991514)	.3239101	.236618	1.477125	-.5835178	1.408696
Precipitation$^2(\div 1000)$	-.1718048*** (.0755211)	-.134721	-.0927474	.563858	-.5249135	.2224721
Constant	-1.750997*** (.4387644)	-1.193757	-1.186775	3.300954	-3.578941	.601102
R squared	.9737651		Observations	1680		

Note: Standard errors are reported in parentheses.

4.2 Heterogeneous slope model (MO-OLS)

It should be noted that this result is closely related to the parametric specification chosen in Eq. (7). It is therefore interesting to investigate it in more detail using a more general specification that allows for both province and time heterogeneity in parameters on temperature and precipitation, or

$$\begin{aligned} y_{it} &= c_{it} + \rho_{it} y_{i,t-1} + \beta_{1,it} GDD_{it} + \beta_{2,it} KDD_{it} + \beta_{3,it} Prec_{it} + \beta_{4,it} Prec_{it}^2 + \varepsilon_{it} \\ & \quad i = 1, \ldots, N, \; t = 1, \ldots, T \end{aligned} \tag{9}$$

Table 3 reports results from the estimation of Eq. (9) using MO-OLS presented above. This table gives unweighted and weighted means of the estimated coefficients, using crop acreage of each province as weight, as well as their standard deviations, medians and 25% and 75% percentiles. Due to the asymptotic normality of the distribution of the unweighted mean, it is possible to associate a standard error with it and therefore to test its significance, which cannot be done for each province and time varying parameters.

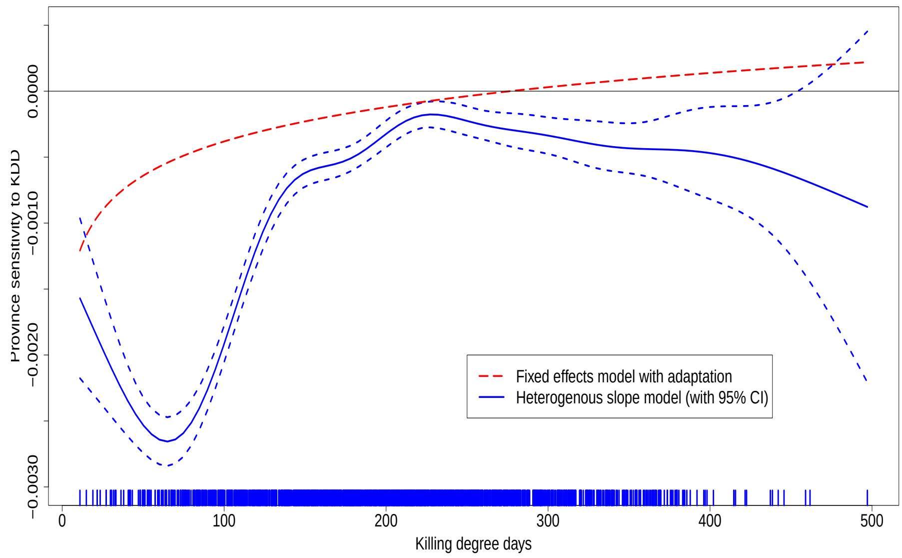
The unweighted mean parameter on lagged rice yield is equal to 0.369 and is significantly different from zero, showing existence of true persistence in rice yield. High rice yields in current years are driven by high rice yields in past years, and vice versa, after controlling for observed and unobserved heterogeneity among provinces. Nevertheless, this average effect is smaller than the marginal effect of 0.660 obtained in the last column of Table 2 where we adaptation is modelled parametrically using the nonlinear KDD parameter. Unobserved factors seems to play a greater role in explaining persistence in rice yield when spatial and time heterogeneity are considered in production function modeling. Finally, the variation coefficient, or 0.539, is small and shows a concentration of individual estimates around their average values.

Results further highlight an inverse U-shaped relationship between rice yield and precipitation. The rice yield increases as the precipitation increases but this positive effect gradually diminishes until it is canceled or eventually begins to decrease. Furthermore, this effect appears more pronounced when the spatial and temporal heterogeneity are taken into account in the modeling of the production function. Indeed, the mean values of the parameters associated with precipitation shown in Table 3 are twice as large as the estimated values shown in Table 2. Nevertheless, the statistics summarizing the distribution of estimated individual and temporal values show a high variability of the responses of rice yield to precipitation with respect to geography and time.

The unweighted average parameter on GDD is equal to 0.00053 and is significantly different from zero, implying that one extra degree days of temperature between $7°$C and $29°$C causes a 0.053% increase in rice yield. Note that no significant and positive effect of GDD appears when estimating production function models with only individual and temporal fixed effects. Here too, individual responses to GDD do not appear to be widely dispersed around their mean.

The unweighted mean parameter on KDD is negative, as expected, and is significantly different from zero, showing that one extra degree days of temperature above $29°$C causes a -0.064% decrease in rice yield. The standard deviation of the KDD parameter, or 0.00120 is twice as large as the estimated mean value, with a $75/25$ percentile range of 0.00009 to -0.00969. The MO-OLS estimates display thus substantial heterogeneity in estimated KDD sensitivity. Moreover, estimated KDD sensitivities appear to be positive for some provinces and some years. Such positive values

Figure 8
Estimated marginal effect of KDD on rice yield as a function of KDD



contradict the expected results with respect to the sign of KDD sensitivities. Nevertheless, in the absence of any result pertaining to the asymptotic distribution of the estimated heterogeneous slopes, it is impossible to say whether these values are statistically significantly positive.

It is therefore interesting to investigate the shape of the relationship linking KDD sensitivity to KDD. Figure 8 reports the nonparametric regression of estimated heterogeneous KDD sensitivities, or $\widehat{\beta}_{2,it}^{\text{MO-OLS}}$, on corresponding values of KDD, or KDD_{it} and the corresponding 95% confidence interval.[5] First, the estimated curve shows that, on average, KDD sensitivity is always negative and significantly different from zero.[6] Second, the estimated curve suggests a more complex relationship between yield response and KDD that the parametric models fail to capture (Figure 8 also reports the estimated marginal effect in FE-OLS model with adaptation). Therefore, as KDD moves from 0 to 60, rice yields respond more negatively to heat stress. Adaptation seems to be overlooked in cooler regions. These regions are not the ones with the best conditions for rice growth as shown in the description of the data.

[5]Nonparametric estimation is based on recent advances in estimating generalized additive models from splines (Wood, 2017).

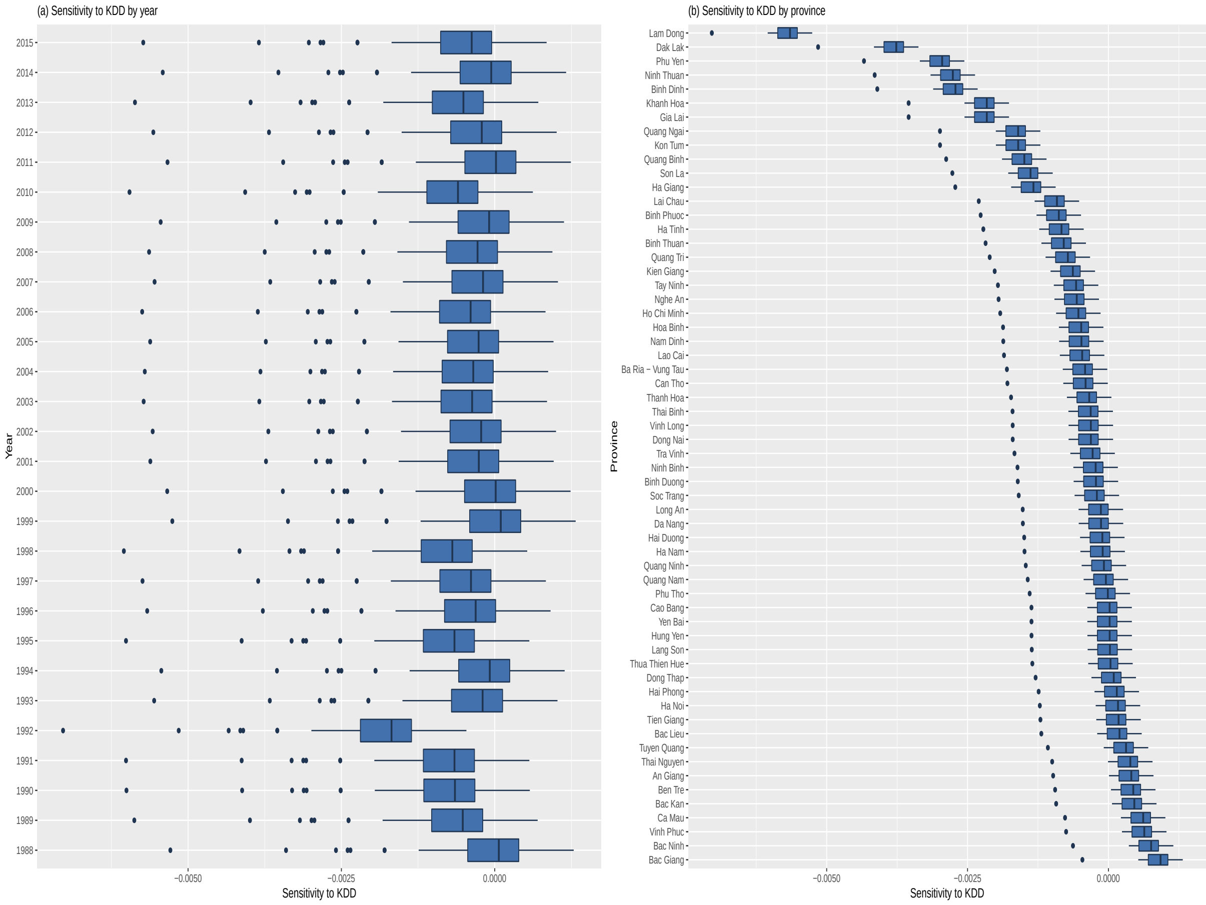
[6]Estimated degrees of freedom associated with the estimated curve are equal to 8.549. They are larger than 1, indicating the nonlinearity of the curve. Moreover, the null hypothesis of joint nullity of all the parameters involved in the spline basis approximation of the smooth function is clearly rejected. Indeed, a low p-value, beyond classical significance levels, is associated with the estimated value of the F-statistics involved in the test.

The costs of adapting to high temperatures may therefore not be offset by yield gains in these regions. As KDD moves from 60 to 220, rice yields are less and less sensitive to high temperatures. Adapting to high temperatures is becoming increasingly profitable. The increase in KDD is accompanied by an increase in GDD for the provinces concerned. They are therefore experiencing increasing incentives to adapt to high temperatures and to limit the resulting losses in terms of yield. For KDD values exceeding 220, the estimated curve shows a slow decrease and even stabilization in the yield response. Although the provinces concerned are those which experience the best conditions in terms of rice cultivation, their efforts to adapt to heat stress seem to be less and less effective, this stress becomes a great concern. For KDD values exceeding 400, the curve estimate becomes increasingly imprecise due to the scarcity of available observations (see the rug of observations on the x-axis of Figure 8).

No clear trend appears regarding the evolution of the distribution of provincial responses to KDD over time, as shown in left panel of Figure 9. The beginning of the period, 1988-1992, is characterized by a deterioration in these responses. Subsequently, from 1993 to 1998, the evolution is characterized by wide fluctuations. These fluctuations diminish considerably from 1999 to 2009 and pick up again in the last years studied. We do not therefore see an increasingly pronounced adaptation process to heat stress over the period, as observed for wheat in the US by Keane and Neal (2020a).

However, a very clear pattern appears when looking at the differences between the distributions of responses to KDD of Vietnamese provinces. The right panel of Figure 9 clearly shows a ranking of Vietnamese provinces from the most KDD sensitive to the least. The descriptive analysis of the climatic condition distributions revealed a clear ranking of the Vietnamese provinces, showing that the more they experience favorable conditions for rice growth, the more they face the risk of heat stress. So, is there a link between this classification and that regarding sensitivity to heat stress? The Spearman correlation coefficient between the ranking of provinces relative to GDD and that relative to KDD sensitivity is 0.246 and is not significantly different from zero. The correlation coefficient between now the ranking of provinces relative to GDD and that relative to KDD sensitivity, or 0.380, is higher and significantly different from zero at classical significance levels. This clearly reflects an adaptation to heat stress which is getting stronger as this stress is high. Despite being significant, this correlation remains weak because there is no perfect fit of the two rankings. A perfect fit would

Figure 9
Distributions of $\beta_{2,it}$



have resulted in a linear relationship between KDD sensitivity and KDD value. However, as shown in Figure 8, this relation appears to be log-linear for a part of the potential values of KDD, and even decreasing for the low or very high values of this climatic indicator.

5 Conclusion

Our paper contributes to the growing literature of climate change impact in Vietnam's agricultural sector in several ways. There have been many discussions around agricultural communities about the potential effects of climate change related to changes in both temperature and rainfall at the national scale. Nevertheless, the literature has treated the yield response without accounting for adaptations to get a full measure of climate damage. The need for large-scale adaptive measures to vulnerability to increased warming might be overlooked. Vietnamese farmers have engaged in all kinds of adaptations to protect their crop against heat stress over the half past century. Models that do not allow for adaptation might result in biased estimates of the effects of high temperatures on plant growth.

In addition, the conventional empirical models treated the country as one region which implies that climate change adaptation remains the same everywhere. The panel data model with heterogeneous slopes recently proposed by Keane and Neal (2020a) provides us a tool to investigate how the responsiveness of rice yield to climate conditions varies over time and across Vietnam's regions. Indeed, heterogeneity in the parameters of the production function linking the rice yield to climatic conditions makes it possible to infer different impacts on heat stress tolerance under different environmental conditions and adaptation measures. This heterogeneity takes a fixed effect form not only in intercepts, as in classical linear panel data models, but also in slopes. In addition, as a byproduct, this approach makes it possible to infer the link between heterogeneous responses and heat stress, without fixing its form as in the fixed effects model with adaptation also proposed by Keane and Neal (2020a) following Butler and Huybers (2013).

A dynamic version of the production function capturing not only heterogeneity in slopes but also state persistence in rice yields is estimated using panel data for the 1987-2015 period across 60 subnational units. We use the MO-OLS estimator proposed by Keane and Neal (2020a). Results clearly show strong temperature effects on rice yield. We find that rice yield is magnified when beneficial temperatures (warming to the 7–29 °C range in a growing season) increase. Extreme heat (average temperatures exceed 29 °C) intensifies yield loss. Rice yield also relies heavily on rainfall for growth, but excessive rainfall can bring significant damage.

Results also show that the heat stress response function is nonlinear in extreme heat. This function seems to have a natural logarithm form when using parametric specifications, as shown by Butler and Huybers (2013) and Keane and Neal (2020a) when dealing with wheat yield in the U.S. There is good evidence from the parametric models that adaptation increases plant heat stress tolerance. However, whether crop is still resilient beyond an extreme climate threshold is questionable. We thus estimate nonparametrically the function, in order to overcome the caveats of classical parametric specifications.

Results then suggest that rice yield sensitivity to heat stress is maximized in the middle range of killing degree days, but declines significantly if they rise above an upper threshold. Put differently, farmers adapt all the more to heat stress as it becomes important. And this holds true up to a threshold beyond which the gain resulting from their adaptation effort is exceeded by the losses due to heat stress.

From a policymakers' perspective, location is very important to consider when thinking about developing food security and sustainability solutions. There are big disparities, yet well-defined spatial patterns in sensitivity across regions. In general, moving towards colder environments, the impact of higher temperatures on yield is more severe. There is a significant difference between crops in rice production hubs such as Mekong River Delta, Red River Delta and other parts of the country in their response to heat stress. The finding also implies that increases in temperature in the coming decades could have even more disruptive effects on agriculture in the coastal part.

This paper focuses on the impact of climatic conditions, and, more specifically, heat stress, on rice yields in Vietnam, in presence of farmers' adaptive behavior. Further studies should be undertaken to go into more detail about rice production in Vietnam. Therefore, it would be interesting to consider separately the two, and, sometimes, three rice growing seasons of rice in Vietnam, each season having its own climates requirements. Such a distinction was not possible with the available data. Similarly, other meteorological indicators such as humidity, wind spreed, sunshine duration, evaporation, could be considered (Zhang et al., 2017)

References

Arnold, C. Y. (1960). Maximum–minimum temperatures as a basis for computing heat units. *Proc. Am. Soc. Hort. Sci. 76*, 682–692.

Auffhammer, M. (2022). Climate adaptive response estimation: Short and long run impacts of climate change on residential electricity and natural gas consumption. *Journal of Environmental Economics and Management 114*, 102669.

Bairagi, S., S. Mohanty, S. Baruah, and H. T. Thi (2020). Changing food consumption patterns in rural and urban Vietnam: Implications for a future food supply system. *Australian Journal of Agricultural and Resource Economics 64*(3), 750–775.

Baskerville, G. L. and P. Emin (1969). Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology 50*(3), 514–517.

Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Journal of Economics 1*, 141–162.

Burke, M. and K. Emerick (2016). Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy 8*(3), 106–40.

Butler, E. E. and P. Huybers (2013). Adaptation of US maize to temperature variations. *Nature Climate Change 3*, 68–72.

Chung, N. T., A. Jintrawet, and P. Promburom (2015). Impacts of seasonal climate variability on rice production in the Central Highlands of Vietnam. *Agriculture and Agricultural Science Procedia 5*, 83–88.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature 52*(3), 740–798.

Eckstein, D., V. Künzel, L. Schäfer, and W. M. (2020). Global climate risk index 2020. Who suffers most from extreme weather events? Weather-related loss events in 2018 and 1999 to 2018. Report, Germanwatch, Bonn, Germany, https://www.germanwatch.org/en/17307.

Espagne, E., T. Ngo-Duc, M.-H. Nguyen, E. Pannier, M.-N. Woillez, T. P. L. Drogoul, A.and Huynh, T. T. Le, T. T. H. Nguyen, T. T. Nguyen, T. A. Nguyen, F. Thomas, C. Q. Truong, Q. T. Vo, and C. T. Vu (2021). Climate change in Vietnam: Impacts and adapatation. a COP26 assessment report of the GEMMES Vietnam project. Technical report, Agence Française de Développement, Paris, France, https://www.ird.fr/gemmes-vietnam-report-climate-change-vietnam-impacts-and-adaptation.

Gilmore, E. C. and J. S. Rogers (1958). Heat units as a method of measuring maturity in corn. *Agronomy Journal 50*(10), 611–615.

Heutel, G., N. H. Miller, and D. Molitor (2021). Adaptation and the mortality effects of temperature across U.S. climate regions. *Review of Economics and Statistics 103*(4), 740–753.

Keane, M. and T. Neal (2020a). Climate change and U.S. agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics 11*(4), 1391–1429.

Keane, M. and T. Neal (2020b). Comparing deep neural network and econometric approaches to predicting the impact of climate change on agricultural yield. *The Econometrics Journal 23*(3), S59–S80.

Kolstad, C. D. and F. C. Moore (2020). Estimating the economic impacts of climate change using weather observations. *Review of Environmental Economics and Policy 14*(1), 1–24.

Krishnan, P., B. Ramakrishnan, K. R. Reddy, and V. R. Reddy (2011). High-temperature effects on rice growth, yield, and grain quality. In D. L. Sparks (Ed.), *Advances in Agronomy*, Volume 111, pp. 87–206. Burlington: Academic Press.

Lobell, D., C. Bänziger, M.and Magorokosho, and B. Vivek (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change 1*, 42–45.

Massetti, E. and R. Mendelsohn (2011). Estimating Ricardian models with panel data. *Climate Change Economics 2*(4), 301–319.

Mendelsohn, R., W. D. Nordhaus, and D. Shaw (1994). The impact of global warming on agriculture: A Ricardian analysis. *The American Economic Review 84*(4), 753–771.

Roberts, M. J., W. Schlenker, and J. Eyer (2013). Agronomic weather measures in econometric models of crop yield with implications for climate change. *American Journal of Agricultural Economics 95*(2), 236–243.

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2006). The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *The Review of Economics and Statistics 88*(1), 113–125.

Snyder, R. L. (1985). Hand calculating degree days. *Agricultural and Forest Meteorology 35*, 353–358.

Trinh, T. A. (2018). The impact of climate change on agriculture: Findings from households in Vietnam. *Environmental and Resource Economics 71*(4), 897–921.

Wassmann, R., S. V. K. Jagadish, K. Sumfleth, H. Pathak, G. Howell, A. Ismail, R. Serraj, E. Redona, R. K. Singh, and S. Heuer (2009). Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. In D. L. Sparks (Ed.), *Advances in Agronomy*, Volume 102, pp. 91–133. Burlington: Academic Press.

Wood, S. N. (2017). *Generalized additive models: an introduction with R.* CRC press.

Yoshida, S. (1978). Tropical climate and its influence on rice. Technical Report 20, International Rice Research Institute, Los Banos, Philippines.

Zhang, P., J. Zhang, and M. Chen (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of Environmental Economics and Management 83*, 8–31.

Appendices

A The MO-OLS estimator

Recently, Keane and Neal (2020a) consider the estimation of a linear panel data model with heterogeneous coefficients like

$$y_{it} = \boldsymbol{\beta}'_{it}\boldsymbol{x}_{it} + u_{it},\ i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{A1}$$

where $\boldsymbol{x}_{it} = (1, x_{1it}, \ldots, x_{Kit})'$ is a $(K+1) \times 1$ vector of regressors, $\boldsymbol{\beta}_{it} = (\beta_{0it}, \beta_{1it}, \ldots, \beta_{Kit})'$ is a $(K+1) \times 1$ vector of coefficients that vary across individuals and overtime, and u_{it} is the idiosyncratic error term. The vector of regressors $\boldsymbol{x}_{it}$ includes a constant term, which allows for intercept heterogeneity across i and t, and it may also include lags of the dependent variable or any of the regressors as needed. Regressors are assumed to be weakly exogenous and the idiosyncratic error term to have finite conditional second order moment given $\boldsymbol{x}_{it}$.

Linear panel data model with coefficients varying both over individuals and periods are in general overparameterized, having $N \times T$ coefficients and disturbance second order moments. This by far exceeds the number of parameters estimable from one panel data set. Keane and Neal (2020a) propose then the parsimonious alternative where

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t.$$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)'$ is the vector of the constant effects across all observations, $\boldsymbol{\lambda}_i = (\lambda_{0i}, \lambda_{1i}, \ldots, \lambda_{Ki})'$ are the individual effects that vary across every unit in the panel, and $\boldsymbol{\theta}_t = (\theta_{0t}, \theta_{1t}, \ldots, \theta_{Kt})'$ are time effects that vary between each time period.

Keane and Neal (2020a) propose an iterative estimation procedure, they call "mean observation OLS" (MO-OLS), to get consistent estimates of $\boldsymbol{\beta}_{it}$. This procedure consists in running a series of feasible regressions and then removing the resulting biases. Three preliminary sets of regressions are initially run. First, pooled OLS estimation is applied to Eq. (A1) rewritten as

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it} \text{ with } v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\lambda}_i + \boldsymbol{x}_{it}'\boldsymbol{\theta}_t + u_{it},, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T$$

to obtain $\hat{\boldsymbol{\beta}}$, or

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it} y_{it} \right)$$

Second, regressions by individual using Eq. (A1) rewritten for each indivudual i as

$$y_{it} = x_{it}' \left(\boldsymbol{\beta} + \boldsymbol{\lambda}_i \right) + v_{it} \text{ with } v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\theta}_t + u_{it}, \ t = 1, \ldots, T$$

are run to get $\hat{\boldsymbol{\beta}}_i$, or

$$\widehat{\boldsymbol{\beta}}_i = \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it} y_{it} \right)$$

Third, Eq. (A1) is rewritten for each year t as

$$y_{it} = x_{it}' \left(\boldsymbol{\beta} + \boldsymbol{\theta}_t \right) + v_{it} \text{ with } v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\lambda}_i + u_{it}, \ i = 1, \ldots, N$$

and estimated in order to get

$$\widehat{\boldsymbol{\beta}}_t = \left(\frac{1}{N} \sum_{t=1}^{N} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^{N} \boldsymbol{x}_{it} y_{it} \right)$$

A preliminary biased estimator of $\boldsymbol{\beta}_{it}$ is then given by $\widehat{\boldsymbol{\beta}}_{it}^{\text{Prel}} = \widehat{\boldsymbol{\beta}}_i + \widehat{\boldsymbol{\beta}}_t - \widehat{\boldsymbol{\beta}}$. Indeed, using matrix calculus, it can be shown that

$$\widehat{\boldsymbol{\beta}}_{it}^{\text{Prel}} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t + R(\boldsymbol{\lambda}_i, \boldsymbol{\theta}_t) + Q(\boldsymbol{x}, \boldsymbol{u}) \tag{A2}$$

where the first bias term $R(\boldsymbol{\lambda}_i, \boldsymbol{\theta}_t)$ that depends on unknown values of parameters in $\boldsymbol{\lambda}_i$ and $\boldsymbol{\theta}_t$, arises from correlation between the regressors and the heterogeneity, and the second bias term $Q(\boldsymbol{x}, \boldsymbol{u})$ only involves the regressors and the errors. The latter vanishes asymptotically given the weak exogeneity assumption for regressors.

Keane and Neal (2020a) show how the bias can be calculated to arbitrary accuracy and removed. This requires replacing $\boldsymbol{\lambda}_i$ and $\boldsymbol{\theta}_t$ by their estimates $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\boldsymbol{\beta}}_t$ to form a biased estimate of $R(\boldsymbol{\lambda}_i, \boldsymbol{\theta}_t)$.By replacing $R(\boldsymbol{\lambda}_i, \boldsymbol{\theta}_t)$ by its estimate corrected for its bias, we eliminate the original bias from $\widehat{\boldsymbol{\beta}}_{it}^{\textsf{Prel}}$, while introducing a new bias. Keane and Neal (2020a) show that this new bias is smaller than the original bias.

This process can be repeated using $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\boldsymbol{\beta}}_t$ to approximate the new bias. In turn, this generates a new bias which is smaller in magnitude. In fact, this process can be repeated L times to render the heterogeneity biases arbitrarily small and form the final estimates:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{it} &= \widehat{\boldsymbol{\beta}}_i + \widehat{\boldsymbol{\beta}}_t - \widehat{\boldsymbol{\beta}} + \sum_{\ell=0}^{L}(-1)^{\ell+1}\left(\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell} + \boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell}\right.\\ &\qquad \left. -\boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell} + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell}\right)\right)\end{aligned} \tag{A3}$$

with $\boldsymbol{Q}_{xx,N} = \left(\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'/N\right)$, $\boldsymbol{Q}_{xx,T} = \left(\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'/T\right)$, and $\boldsymbol{Q}_{xx,NT} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'/NT\right)$. Moreover, $\Gamma_{1,\ell} = Q_{xx,T}^{-1}\left(\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell-1}/T\right)$ and $\Gamma_{2,\ell} = Q_{xx,N}^{-1}\left(\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell-1}/N\right)$ when $\ell >$ 0 , and $\Gamma_{1,0} = \widehat{\boldsymbol{\beta}}_i$ and $\Gamma_{2,0} = \widehat{\boldsymbol{\beta}}_t$. This is a Cauchy sequence in ℓ, so a suitable L can be chosen by terminating the sequence once it converges to a desired tolerance. In practice, small values of L are usually adequate. Eq. (A3) is simple to construct as it is a function of only the preliminary estimates $\left(\widehat{\boldsymbol{\beta}}_i, \widehat{\boldsymbol{\beta}}_t, \widehat{\boldsymbol{\beta}}\right)$ and the covariates $\boldsymbol{x}_{it}$.

Keane and Neal (2020a) show consistency of $\widehat{\boldsymbol{\beta}}_{it}$ when N and T together tend to infinity. They then define the Mean Observation OLS (MO-OLS) estimate as the simple average of of estimates of the observation-level coefficients, or

$$\widehat{\boldsymbol{\beta}}_{\textsf{MO}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{\beta}}_{it} \tag{A4}$$

Finally, they provide results for consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_{\textsf{MO}}$.

B GDD and KDD computation

Calendar days do not provide reliable information about the timing of crop development because crop grows by the accumulation of heat over the growing season (Gilmore and Rogers, 1958). When the air temperature exceeds the base temperature for a certain length of time, rice will grow. If the temperature falls below the base temperature, development slows. Degree day accumulations are used by growers to monitor the development of biological processes and thus are used in crop and pest management. Following Arnold (1960), Baskerville and Emin (1969) and Snyder (1985), degree days are usually computed from daily minimum and maximum air temperature, or $T_{\min}$ and $T_{\max}$, using a sinusoidal approximation, or

$$T = M + W\sin(t)$$

where t is time expressed in radians from $-\pi/2$ to $3\pi/2$, or $t = \pi(h-6)/12$ where h varies form 0 to 24 hours, $M = (T_{\max} + T_{\min})/2$, and $W = (T_{\max} - T_{\min})/2$. This approximation is depicted on Figure A1 when $T_{\min} = 5°$C and $T_{\max} = 30°$C.

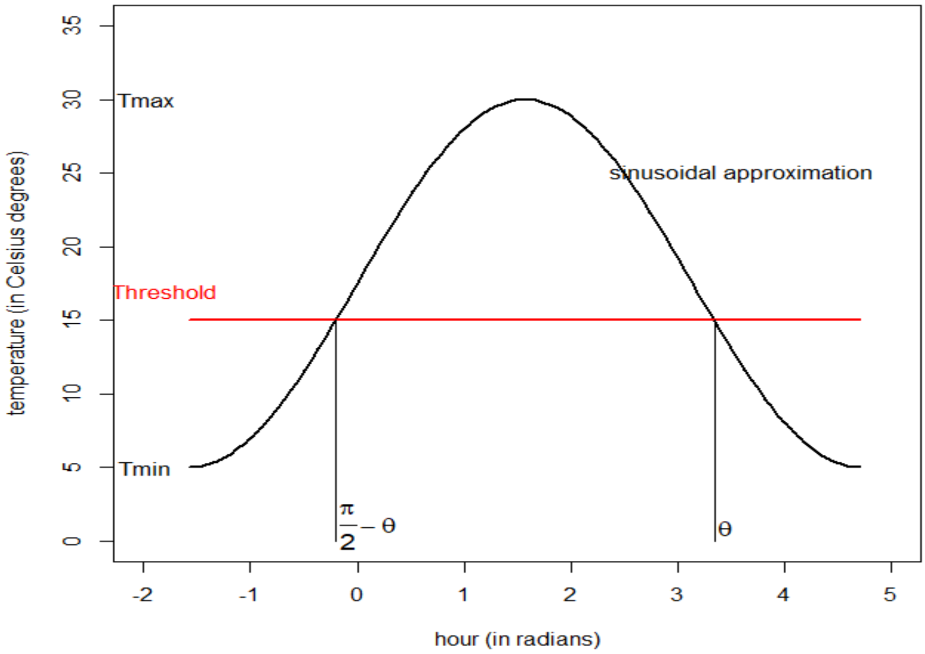
Degree days at a given threshold temperature are measured by integrating the aera under the sinusoidal curve above this threshold temperature. In Figure A1, this area corresponds to the difference between the area under the sinusoidal curve for time between $\pi/2-\theta$ and θ, and the aera of the rectangle whose base goes from $\pi/2-\theta$ to θ and height is equal to the threshold temperature value. This results in the following expression of degree days as a function of the value C of the temperature threshold:

$$\mathrm{DD}_C = \begin{cases} 0 & \text{if } C > T_{\max} \\ T_{\textsf{avg}} - C & \text{if } C < T_{\min} \\ \left((T_{\textsf{avg}} - C)\,(\pi - 2\sin^{-1}(\theta)) + (T_{\max} - T_{\min})\cos(\sin^{-1}(\theta))\right)/2\pi & \text{otherwise} \end{cases}$$

where $T_{\textsf{avg}} = (T_{\max} + T_{\min})/2$ and $\theta = (2C - T_{\max} - T_{\min})/(T_{\max} - T_{\min})$.

Figure A2, which has been adapted fromYoshida (1978) by Krishnan et al. (2011), provides us a guideline for choosing the boundaries of the temperature for rice development.As shown in Figure

Figure A1
Graphical illustration of DD computation

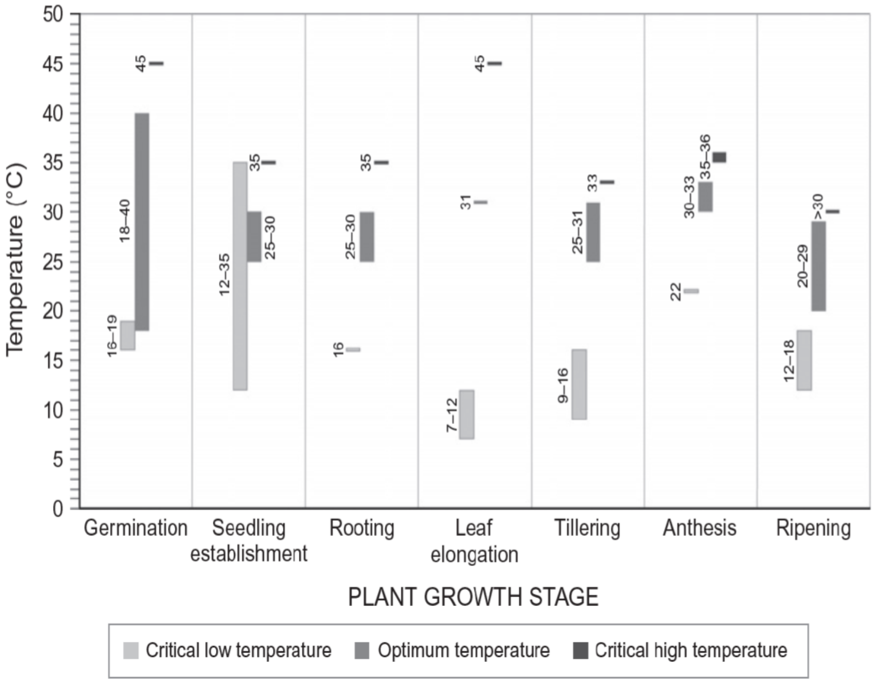


A2, rice crop only develops if there is adequate heat, i.e. if the minimum temperature for the day is above a base temperature. Rice Crop will develop faster with more heat between a base temperature and an optimum temperature. Growth will slow between this optimum and an upper temperature and eventually cease occurring outside the upper temperature range. We choose optimum growing temperature between 7°C and 29 °C to ensure our low and high temperature thresholds are within the acceptable range for the positive growth rate of rice in Vietnam [7]. Beyond this range, high temperature could reduce yield by delaying flowering and shortening the duration of grain-filling. Hence, growth and productivity would rapidly decrease.

Cumulative beneficial temperatures ("growing degree days") are calculated by adding up beneficial temperature per day over a season to predict when rice crop will reach maturity, i.e. $\mathrm{GDD}_{it} =$

[7] The optimum growing temperature chosen in our study also counts for the fact that the temperature experienced by the crop itself is normally higher than the measured air temperature above the plant canopy as noted by Butler and Huybers (2013) and Keane and Neal (2020a). We also find that the other choices of threshold reduce the predictive power of the fit in our regressions.

Figure A2
Rice plant response to varying temperature at different growth stages.



$\mathrm{DD}_{7,it} - \mathrm{DD}_{29,it}$ where $\mathrm{DD}_{d,it}$ is the accumulation of degree days at temperature threshold d for province i over the growing season in year t. Harmful temperatures ("killing degree days") indicating the amplified warming that might be detrimental to crops are calculated as $\mathrm{KDD}_{it} = \mathrm{DD}_{29,it}$.

The rice yield data we have is annual. Accordingly, the different harvests that take place in North and South Vietnam during one year are summarized in only one growing season going from March to October (Wassmann et al., 2009).

C A simple model of agricultural yield with weather and adaptation

This appendix replicates section 1 in Keane and Neal (2020a). In this section, they present a simple model of agricultural yield with adaptation. Their aim is to provide a coherent framework for the empirical work they present in their paper. This model starts with a production function that incorporates measures of temperature:

$$\frac{Y_{it}}{C_{it}} = A_t \mu_i I_{it} \left(1 + \beta_1 (GDD_{it} - GDD_{min}) + \beta_2 KDD_{it}\right) \tag{A5}$$

Y_{it} denotes the crop output (in tonne) for farm i at time t, while C_{it} indicates the land area (hectare) planted. Thus, Y_{it}/C_{it} measures crop productivity or yield (tonne per hectare). Basic factors of production, denoted by I_{it}, are capital, labor, and fertilizers that vary across farms and time. μ_i is an indicator of unobserved soil characteristics pertaining to farm i, and A_t indicates total factor productivity that enhances crop productivity at time t.

The yield fluctuation associated with temperature variation is characterized by growing degree days, or GDD_{it}, and killing degree days, or KDD_{it}, as defined in Appendix B. $GDD_{\min}$ stands for the minimum level of GDD needed for crop to experience a positive yield. Intuitively, β_1 is expected to be positive whereas β_2 is expected to be negative. $(\beta_1(GDD_{it} - GDD_{min}) + \beta_2 KDD_{it})$ measures the percent shift in crop yield due to temperature factors.

Taking the log on both sides of Eq. (A5), and making use of the approximation $ln(1 + x) \approx x$ when x is small, we obtain:

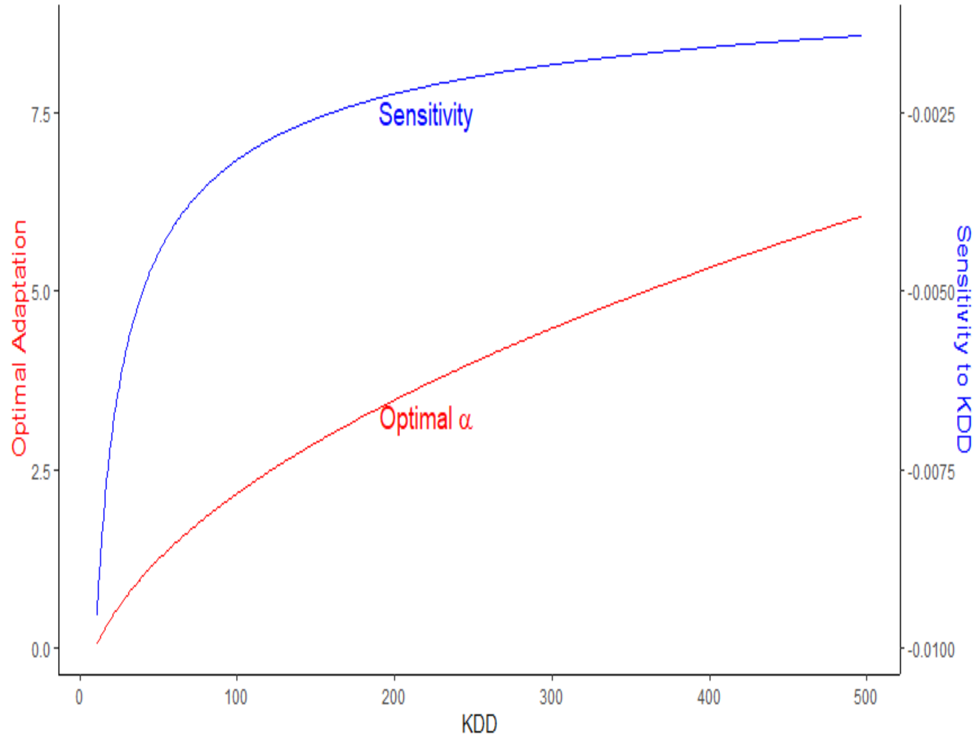
$$y_{it} = \ln(A_i) + (\ln(\mu_i) - \beta_1 GDD_{min}) + \ln(I_{it}) + +\beta_1 GDD_{it} + \beta_2 KDD_{it} \tag{A6}$$

where $y_{it} = \ln(Y_{it}/C_{it})$. Keane and Neal (2020a) then notice that Eq. (A6) is akin that estimated in several recent papers (See, among others, Lobell et al., 2011; Burke and Emerick, 2016). In these papers, fixed effects over i and i are used to capture the A_t, μ_i, and I_{it} terms.The estimation of model (6) is part of this approach which neglects any adaptation to climate change by farmers.

Subsequently, the simple model is extended to present a more sophisticated adaptive farm management approach. Climate change imply that the weather experienced by a farmer would change. A farmer who deal with a potential loss of crop revenue due to increased temperatures would adjust production choices by adapting its production technology to the new climatic situation. For instance, in drier and hotter periods, yields may be lower than normal and seeds with heat or drought tolerance traits can be selected to better protect the crop. Farmers adapt crop production by adopting best practices that decrease sensitivity of yield to extreme heat or droughts at a monetary cost.

Figure A3

Relationship between optimal adaptation, or α^*_{it}, yield sensitivity, or $\beta^*_{2.it}$, and KDD_{it}



Let now the coefficient that indicates yield response to KDD varying across farmer an time according to $\beta_{2it} = s/(1+\alpha_{it})$ where α_{it} denotes units of adaptation purchased in farm i in period t, and s, which is negative, indicates the adverse impact of heat stress on crop yield when $\alpha_{it} = 0$. Determining optimal responses to extreme temperature due to climate climate involves trading off the benefits of the adaptation choice against its cost. Profit for farm i in period t is total revenue minus total cost, i.e.

$$\pi_{it} = p_t Y_{it} - \gamma \alpha_{it} - r_t I_{it} \tag{A7}$$

where p_t is the price of the crop,γ is the price of adaptation, and r_t is the rental rate per unit of production factor in period t. To maximize profit, farmers purchase the optimal level of adaptation.

Setting $\partial\pi/\partial\alpha = 0$ leads to

$$\alpha^*_{it} = \sqrt{\frac{p_t(C_{it}A_i\mu_i I_{it})(-s)KDD_{it}}{\gamma}} - 1 \tag{A8}$$

Hence, the optimal level of adaptation increases in correspondence with KDD_{it}. In hot-enduring regions, farmers have more incentive to protect their crop through investing in various adaptation activities. Figure A3 displays the optimal level of adaptation α^*_{it} and the corresponding sensitivity to KDD $\beta^*_{2,it}$ as a function of KDD.[8] The relationship between KDD and $\beta^*_{2,it}$ looks like a log-linear function. Taken in absolute value, the sensitivity to KDD tends to become negligible when KDD increases, highlighting a better adaptation to the extreme temperatures of the hotter provinces. The prediction of this simple economic model is in line with the empirical findings of Butler and Huybers (2013). This prediction motivates the estimation of model (7) where the marginal effect of KDD_{it} on y_{it} approximates a log linear dependence, or

$$\frac{\partial y_{it}}{\partial KDD_{it}} = \beta_{20} + \beta_{21}\ln\left(KDD_{it}\right). \tag{A9}$$

The prediction assumes that β_{20} is negative while β_{21} is positive, assumptions that can be tested.

[8] As Keane and Neal (2020a), we fix $p_t(C_{it}A_i\mu_i I_{it}) = 1$ and $s = -0.01$.