The Distributional Impacts of Climate Change on Indian Agriculture: A Quantile Regression Approach

Chandra Kiran B Krishnamurthy

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India
May 2012
The Distributional Impacts of Climate Change on Indian Agriculture: A Quantile Regression Approach

Chandra Kiran B. Krishnamurthy
Post-Doctoral Research Scholar, Centre for Environmental and Resource Economics & Department of Economics
Umea School of Business and Economics
Umea University, Sweden
chandra.kiran@econ.umu.se
The Distributional Impacts of Climate Change on Indian Agriculture: A Quantile Regression Approach

Chandra Kiran B Krishnamurthy

**ABSTRACT**

Using a 30-year dataset on district-level yields, with more than 200 districts, and pairing it with a newly available gridded weather dataset, this paper estimates the impact of climate change on major food crops of India. Unlike previous work, which focuses on impacts on the conditional mean, the current approach, using newly developed methods for fixed effects in the quantile regression context, evaluates the impact on yields at different locations along full conditional distribution of yield. This approach allows for a better understanding of the differential impacts of climate change on yield, even within a given region. Further, this paper allows for differential impacts of temperature and irrigation on crops grown in different seasons. This paper reports significantly reduced yields of wheat, of up to 12%, in all regions and at most quantiles, under scenarios with reasonable temperature increase. Further, the reductions are larger at the upper quantiles, indicating significant impacts on production. For rice however, under both scenarios considered there is a very modest (up to 2%) increase in yield at the intermediate quantiles while both at the upper and lower quantiles, there are modest reductions in yield (of up to 3%). There are also significant regional differences in impacts at different quantiles. These estimates suggest significant likely loss in production of food grains under the scenarios considered and have significant implications for malnutrition in India.

**Keywords:** Climate Change, India, Agriculture, Quantile Regression, Panel Data

**JEL Classification codes:** Q54, C21, C23, O13
ACKNOWLEDGEMENTS

I thank Prof. Wolfram Schlenker for helping both, initiate the project as well as with acquiring the agricultural data. I wish to thank the IMD (Indian Meteorological Department), in particular Dr Rajeevan Nair, for providing the daily rainfall data, and Dr Andrew Robertson, at the IRI, Columbia University, for both providing the temperature data as well as for discussions regarding the use of climate model projections for India. I also wish to thank, without implicating them in any way, Prof. Bernard Salanié for providing substantive comments on an earlier draft, and my Dissertation advisors Upmanu Lall and Geoff Heal for their support. This paper also benefitted from comments by Seminar participants at Columbia University. Finally, I wish to thank the Madras School of Economics (MSE) for hosting me for a few weeks while I worked on a modified draft. In particular, I wish to thank Prof. Kavi Kumar, at the MSE, for making the visit possible and for being a good host. I alone am fully responsible for any remaining errors.
INTRODUCTION

Given the incontrovertible evidence on human-induced changes in climate, research is increasingly focussed on better and finer-scale understanding of the pathways of the impacts of such changes on economic activities. Quantifying the nature of such impacts, under a variety of possible scenarios of climate change, is a critical stage in the process of understanding the nature and magnitude of adaptations to the changes. The impacts of climate change are likely evident on a variety of time scales and climatic variables, and research to date has focused mostly on the manifestations of such change, primarily through changes in precipitation and temperature. Many regions of the globe are anticipated to undergo rapid warming at the surface, with a possibility of increasing precipitation.

Of the many sectors of an economy impacted by climate change, agriculture and forestry have been identified to be most susceptible, given that weather is a direct component of the production of these sectors. An equally important reason is the disproportionate share of agriculture in employment, especially in the developing nations. Most studies of climate change impacts on agriculture focus on the developed world, primarily the US (Schlenker et al. (2006); Schlenker and Roberts (2006, 2009); Deschenes and Greenstone (2007)), but the importance of the agricultural sector in these regions are smaller than those in the developing nations. Estimates of the impacts of climate change on these economies will presumably be a first and important step in adapting to these changes. The importance of the task is magnified by the predictions of both larger changes in climate (and therefore weather) and larger uncertainties in those changes in the developing nations (predominantly in the tropical and sub-tropical regions) than in the more developed (and temperate) ones. This paper provides direct evidence on the impact of climate change on Indian agriculture, and finds that climate change is likely to reduce yields of many food grains substantially, although there is also intriguing evidence of only small negative and even slightly positive changes for rice.

Previous studies quantifying the impact of climate change on agriculture fall into one of three categories, based on the approach used to quan-
tify such impacts: production function approach, the Ricardian approach and a panel data approach. The production function approach is based on controlled agricultural experiments in a laboratory-like setting, under agronomically optimal conditions. As a result, they reflect optimal outcomes, with only minimal adaptation to changes possible, and likely differ from realized outcomes at the farm level. The Ricardian approach, pioneered by Mendelsohn et al. (1994), attempts to control for the full range of farmer adaptations possible, by estimating a functional relationship between land values and other (appropriate spatial) climate variables (along with other controls). The main premise of the approach is that, with well functioning land markets, land prices will reflect the present discounted value of profits from all uses of land, controlling for many potentially unobservable characteristics. Naturally, the reliability of such methods depend on the tenability of such assumptions (for evidence on this approach to agriculture in the US, see Deschenes and Greenstone (2007); Fisher et al. (2009)).

A variant of this approach, which is less demanding in terms of well functioning markets, is the semi-Ricardian approach, developed in Sanghi et al. (1998) and applied to the cases of Brazil and India. This approach uses average profits (over a long period of time) at the district level as a proxy for land prices, with the assumption that land prices, if available, would be the present discounted value of such profits. Using this approach, Sanghi et al. (1998) and Kumar and Parikh (2001) find significant negative impacts on profits under a variety of climate change scenarios for India.

The panel data approach uses presumably random year-to-year variation in weather across US counties to estimate the impact of weather on agricultural yield (Schlenker et al. (2006); Schlenker and Roberts (2006, 2009)) and profits (Deschenes and Greenstone (2007)). This approach overcomes the omitted variable-type issues encountered in the Ricardian approach, and permits controlling for unobserved individual heterogeneity (via the use of county fixed effects), such as farmer/soil quality, while at the same time allowing for some farmer adaptation, providing presumably better estimates than the production function approach. A major drawback with this approach is an inability to reflect changes in technology and choices (such
as switching of crops, exit from agriculture, change in cropping seasons etc). Using this approach for India, along with gridded reanalysis weather data, Guiteras (2008) explores the impact of climate change on annual profits at the district level, and estimates moderate losses (up to 9%) for the medium term, and significant (up to 25%) losses for the long term, under a variety of climate scenarios.

It is also important to notice that all of the preceding approaches model the mean change in the underlying variable (yield, land value, average profit) by using a mean regression framework (parametric or nonparametric), in which a major assumption implicit is that covariates (weather, climate, controls, etc) impact only the mean of the underlying variable, and do not alter the shape (or scale etc) of the conditional distribution of yield.

In other words, implicit in such an approach is the idea of climate change leading to mean shift in agricultural outcomes, with no changes of underlying relationship between agricultural outcomes and climate. This is problematic for several reasons; for instance, in much of the scientific literature on climate change, focus is on changes in variability (especially in the hydrologic cycle, which determines both long and short-run availability of water supply, a critical ingredient in agriculture) as a result of an altered climate; while such changes are encapsulated in a "mean effect" framework, they are not restricted to it. Research focus in the climate literature therefore is on changes in extremes and in variability changes. Further, the assumption of a mean shift is arguably a testable hypothesis in a statistical framework, rather than an implicit axiom.

This paper develops a novel panel data approach for linking observed weather outcomes and agricultural outcomes at the district level\(^1\) at any number of chosen quantiles. In this framework, yield, at each quantile, is modeled as being a (potentially nonlinear in covariates) function of weather variables, primarily growing degree days and seasonal (and monthly) rainfall and district-specific fixed effects. The use of district-specific fixed effects allows unobserved determinants of district-specific, time-invariant effects (such as soil type, topography etc) to be controlled for. The estimation

\(^{1}\)The district is the smallest administrative unit in India, and is analogous to an American county, although generally larger.
strategy is to regress yield on observed weather outcomes and district-fixed effects, for each quantile of interest. Extensions of the framework to flexible modeling of covariate impacts, as well as to fully non-parametric settings using the penalized splines approach in Koenker et al. (1991) are also indicated. The framework developed here may therefore be seen as a generalization of the non-parametric estimation outlined in Schlenker et al. (2006) to an estimation of various features of the conditional distribution, apart from the mean.

Given that the quantile function completely characterizes the distribution function of a random variable, the idea underlying this approach (and quantile regression in general) is to parametrically obtain estimates of local features of the full conditional (upon covariates) distribution of yield, without sacrificing the interpretability and computational simplicity of the regression framework\textsuperscript{2}. Unlike in the linear panel data case, this approach allows for two distinct types of (unobserved) heterogeneity: district-specific unobserved heterogeneity (due to the inclusion of the fixed effects), and heterogeneity of covariate effects (due to the quantile nature of the regression). Region-specific cubic time trends are also included in order to disentangle the impacts of a potentially warming climate over the latter period of the twentieth century and improvements in agricultural productivity over the same period. The predicted impact of climate change at each quantile is computed as the differences in predicted yield at the historical weather and weather realizations from climate models for various climate scenarios (or uniform changes in temperature and rainfall applied to current weather).

This paper uses a unique and recently available dataset of gridded temperature and rainfall data for India, from the Indian Meteorological Department, spanning 1959-1999, and an agricultural dataset, spanning 1971-2005, for agricultural yield at the district level, for estimation. Rice and Wheat are grown in different seasons, with very different characteristics, and projections of climate changes are very different for these two seasons.

\textsuperscript{2}For instance, alternative and non-parametric estimation of conditional quantile functions are detailed in Koenker (2005, Chapter 7). These frameworks however are (a) not easily extended to higher dimensions (of covariates) (b) lack interpretability, especially in multi-variate settings (c) not readily generalized to fixed effects settings and (d) computationally burdensome, especially since inference is bootstrap-based.
It is therefore important to account for considerations of separate growing seasons and therefore very different impacts on crops in the different growing seasons\(^3\). This paper finds significant negative impacts, of up to 6\% of yield, for wheat, for very reasonable scenarios (up to 0.5\(^\circ\)C increase in temperature), and up to 10\% losses for larger increases (1\(^\circ\)C increase in winter temperature).

For rice, however, results indicate modest losses at the highest and lowest quantiles considered, and very mildly positive impacts at the intermediate quantiles. In addition, mildly positive impact are indicated in certain regions, primarily the northern and eastern regions, a result which is consistent with those in Guiteras (2008). These results are robust to flexible functional forms (quadratic, orthogonal polynomials and fully non-parametric) for growing degree days and to different plausible thresholds for growing degree days.

Prior approaches to estimating impacts of climate change for India suffer from two drawbacks: first, given that there are two agricultural seasons, it is not clear that ignoring completely weather (climate) in this season while including agricultural outcomes for this season as part of the response (as in Sanghi et al. (1998), Kumar and Parikh (2001) and Guiteras (2008)) provides sensible estimates. This issue aside, it is not clear how to interpret the monetary equivalent of crop yield when summed over crops grown in different growing seasons (i.e. \(\sum_k y_{itk}p_{ik}\) where \(k\) denotes the crop, \(y_{itk}\) the yield of crop \(k\) in year \(t\) and district \(i\), prices \(p_{ik}\) are fixed in time) at the district-level (Guiteras (2008)). For any interpretation of this variable as potential productivity, one must also control for all inputs, which is not possible in this case (due to the absence of such data). A final issue is the possible confounding of price impact from productivity i.e. results are

\(^3\)There are two major growing seasons in India: the summer monsoon growing season, called kharif, and the winter growing season, called rabi. There are actually two monsoon seasons over India, the most important being the summer or south western monsoon season, from June to September, during which about 70\% of annual rainfall occurs. The second monsoon season (northeastern monsoon) affects primarily south-eastern and north-eastern regions of India, of which agriculture in south-eastern states of India (Andhra Pradesh and Tamil Nadu) is significantly impacted. In this paper, we will use terms “summer growing season” (“winter growing season”) and “kharif season” (“rabi season”) interchangeably.
likely sensitive to impacts of the specific price used\textsuperscript{4}. The approach outlined in this paper avoids the issues raised above altogether by working directly with changes in physical crop yield, addressing crops grown in different seasons separately, and accounting for substantial changes in seasonal weather directly.

The plan of the paper is as follows: we first detail the agronomic and economic basis of the model, turn then to outlining the advantages of quantile regression in the current setting, indicating the benefits and interpretation of fixed-effects in quantile regressions, proceed to outline the econometric strategy adopted and finally, report results of the estimation, including impacts of climate change on yields.

**MODEL**

The objectives of the paper are to be able to obtain estimates of the impact of weather on crop yields, using a newly available high quality gridded data weather set, and using a newly developed framework for quantile regression for panel data with fixed effects. The approach outlined here allows estimation of the potentially non-linear relationship between temperature and rainfall on crop yields, at different quantiles of yield.

**Temperature and Yields**

Given the reduced form nature of the statistical approach proposed here, it is imperative to obtain a seasonally aggregate metric of temperature relevant for crop growth (given that aggregate metrics such as "average temperature" yield almost no relevant information). Following the approach outlined in Schlenker and Roberts (2009); Guiteras (2008), we postulate a temperature limit within which heat is beneficial to plants, in a linear and time-separable manner. The upper and lower limits of temperature are a matter of some debate even in the context of US agriculture (for which they were first derived) and are generally unknown for the Indian context. However, following Guiteras (2008), we fix these limits to be 8 and 34\textdegree C for the

\textsuperscript{4}It is also not evident that fixing prices at the 1960 level provides any remedy to the problem outlined for the reason that the year chosen (1966) pre-dates all significant agricultural policy changes in India, at both the national and the state level (such as support prices for rice and wheat and sugar co-operatives for Sugarcane, changes in electricity pricing for irrigation and water pricing for surface irrigation). These prices therefore are unlikely to be relevant for later periods.
summer growing season. Following Schlenker et al. (2006), who illustrate the superiority of the agronomic measure of Growing Degree Days (GDD), we compute this measure using two different methods. In the first case, which we term "Linear Degree Days", the computation is straightforward\(^5\). Given daily average temperature, \( T \), compute

\[
D(T) = \begin{cases} 
0 & T \leq T_{low} \\
T - T_{low} & T_{low} \leq T \leq T_{upper} \\
(T_{upper} - T_{low}) & T \geq T_{upper} 
\end{cases}
\]  

(1)

where the upper and lower thresholds are set to 8\(^0\)C and 34\(^0\)C for the summer growing season.

The Linear GDD metric was computed for each day, summed over the growing season (kharif and rabi seasons) to obtain a seasonal sum of GDD\(^6\). The fixing of the growing season is less to do with the endogeneity of farmers' sowing and harvesting decisions (which are, for the rainfed areas, dependent on expected onset of the summer monsoon) and more to do with bio-physical conditions, including availability of water. In order to account for potentially harmful effects of temperatures beyond 34\(^0\)C upon plant growth, we also compute "harmful degree days" as the difference between 34\(^0\)C and the daily average temperature.

The choice of temperature limits for the rabi season was made more difficult by having little India-specific literature on the topic (much of the literature use only a lower bound and do not use a specific upper bound). Given however the significant temperature constraints on wheat, it was decided to choose 6\(^0\)C as the lower threshold (temperatures only very rarely go much below this over much of the wheat growing regions) and an upper threshold of 28\(^0\)C. Choosing slightly different upper thresholds (27 and 29\(^0\)C) did not appreciably alter the results. Harmful degree days were computed as in the case of the kharif season.

\(^5\)Degree days were also computed using an alternative method of Baskerville and Emin, with very similar results (not reported here).

\(^6\)In other words, \( GDD_k = \sum_{t=1}^{T_k} D(t) \) i.e. the growing season degree days are the sum of the daily degree days over the growing season, for a given year; season length is 122 days for the summer growing season and 183 days for the winter season.
Two points regrading this formulation are worth emphasizing: (a) linearity of plant growth as a function of temperature during the growing season (b) the complete independence between temperature and precipitation, when used in the regression framework with yield (not its transformations) as the dependent variable. For instance, in the regressions in Guiteras (2008) (see for instance, Equation (5), pp 22), this implies the complete separation of the effects of temperature and precipitation upon yield, which is somewhat unrealistic in the case of the summer growing season\(^7\).

**Precipitation and Yields**

The focus in most of the econometric research on estimation of climate change impacts has been on estimation of temperature impacts upon plant growth. Agronomic and climate literature for India however, has tended to focus, for rainfed agriculture, on the significance of summer monsoon characteristics. However, in the case of India, even without climate change, climatic variability, especially the variability in monsoon rainfall, has been of major interest for agriculture. This is due to (a) the season: the earlier part of the summer growing season from May-Sept are also the hottest months of the year and temperatures generally reduce only with rainfall (b) lack of water availability: seasonal characteristics (see footnote 7) and the largely rainfed nature of agriculture in India (current estimates indicate that only about 40-50% of agriculture in India is fully irrigated) means that sowing operations are critically dependent on the arrival of rainfall.

The significance of monsoon rainfall for Indian agriculture is well documented in both, Agricultural and Development economics (for instance, Rosenzweig and Binswanger (1993)) and in the Agriculture and Climate literature (Gadgil, 2000 etc). In these regions, it is well known that the major constraint is availability of water during the critical sowing period, May to June, and therefore, onset of the monsoon is of some importance. Further, contrary to popular perception of the monsoon as marching southwest to

\(^{7}\)An important point to bear in mind is the summer growing season in India is the hottest period of the year (May, June and July are the hottest months of the year in large parts of India), where maximum temperatures rise up to 40 - 46\(^\circ\)C in many agricultural regions, with average daily temperatures of 32 - 36\(^\circ\)C. These temperatures are mitigated only by precipitation i.e. the differences in maximum temperatures on days with and without rainfall tend to be substantial, especially early in the season.
northwest, the monsoon has what are known as “break periods”, periods during which there are long dry spells (of up to a fortnight or more in duration) over large parts of India. There is some speculation on the harmful effects of these periods, but little actual evidence about its importance for agriculture.

If (as in Guiteras (2008)) identification is based on weather variations from district mean weather, we argue that it is important to (a) account for rainfall variations which are not captured by monthly totals (b) condition monthly and seasonal total rainfall on onset and dry spells. The latter accounts for the fact that years with late onset and years with normal rainfall are not exclusive and therefore, two years with very similar total rainfall can have very different onset days; further, given the concentrated nature of monsoonal rainfall in India, years with a large number of “dry spells” can easily be years with average monthly rainfall (especially true in the semi-arid parts of India with highly variable rainfall patterns). In other words, monthly rainfall is not a sufficiently fine proxy for rainfall distribution within the month, given the relatively concentrated nature of rainfall in India. However, total seasonal rainfall may be a somewhat inadequate measure of such variability due to the concentrated nature of rainfall in many regions in India (for instance, more than 80% of total rainfall in India is documented as occurring in less than 100 hrs). We therefore use more direct measures on the two aspects of monsoon most relevant for agriculture: onset of the monsoon and the number of dry spells during the season. Monsoon onset is defined simply as the first day after April 1 that rainfall on 5 successive days is above 5mm without being followed by 5 successive days with rainfall below 5 mm. Defining as a “dry spell” any 7-day period with rainfall below 5mm, we obtain a count of such spells for every district for every year (see Appendix A for more details on these variables).

To summarize, we use the following variables for summer monsoon season measures of rainfall: seasonal (monthly) total rainfall and onset day, dry spells. For the winter season we use, in addition to seasonal rainfall, a bi-monthly rainfall s1m.
ECONOMETRIC STRATEGY

We turn now to characterizing our econometric strategy. We begin with a brief interpretation of linear quantile regression as an extension of linear regression to the heterogeneous effects of covariates case, and then turn to indicating the major ideas that underlie an extension of quantile regression to the fixed effects setting.

Quantile Regression and Heterogeneous Covariate Effects

Consider the so-called location-scale model:

\[ Y_{it} = X_{it} \beta + (X_{it} \gamma) \sigma_{it} \]

where where \( c_{it} \) is an error term (not necessarily i.i.d) and \( \gamma \neq 0 \) represents "heterogeneity" (observed). It is evident that the condition \( \gamma = 0 \) is sufficient for the conditional quantiles to be parallel to one another, and this is generally a maintained assumption in mean regressions, parametric or non-parametric. In other words, most mean regression frameworks assume that the conditional quantiles are all parallel to one another and that the covariates exert an effect only on the mean of the distribution. Evidently, this assumption is restrictive. In this context, quantile regression maybe viewed as an extension of regression to the "location-scale" models i.e. covariates also affect the scale parameter (variance, for the normal distribution) of the conditional distribution of the response, due to which the quantiles are no longer parallel to one another⁸.

For instance, in the current application, it is quite plausible that weather covariates have differential impacts on distinct quantiles of the yield⁹; further, if the effect of climate change for India are likely manifested in increases in variance along with changes in mean (as is currently speculated, especially for precipitation), then impacts computed based on mean regression are likely to understate the true magnitude. Similarly, if there is a change in seasonal distribution of rainfall, computing the mean

---

⁸In the case of the simple location-scale model above, note that the conditional variance of \( Y \) is now a function of \( X \) and of \( \gamma \).

⁹A very common economic application of quantile regression is in labor economics, in the analysis of differential impacts of covariates on different parts of the wage distribution. See Fitzenberger et al. (2002) for a list of applications to mostly this area of economics.
impact will tend to smooth out variations in yield at different quantiles.

**Quantile Regression for Fixed Effects Panel Data Models**

There is an extensive literature in economics, particularly in fields such as labor economics, on the usage of quantile regressions (see for instance Fitzenberger et al. (2002)). Effectively, linear quantile regression is the simplest parametric framework for a scenario in which covariates are expected to exert differential impacts on the distribution of the response.

Panel data with fixed effects allow applied researchers to base identification purely on "within" variation in the covariates. Further, this approach allows arbitrary correlation between the covariates and the fixed effects. There has been however little research on the intersection of the two methods. Two approaches have recently been developed to deal with the fixed effects quantile regression (FE-QR) case, differing primarily (from an applied perspective) in how the fixed effects are viewed.

Consider the following two regression frameworks:

\[ \mathbb{E}(Y_{it}|\alpha_i, X_{it}) = \alpha_i + X_{it}\beta \]  \hspace{1cm} (3)

\[ Q_{Y_{it}}(\tau|\alpha_i, X_{it}) = \alpha_i(\tau) + X_{it}\beta(\tau) \]  \hspace{1cm} (4)

where \( Q_{Y_{it}}(\tau|X_{it}, \alpha) \) is the \( \tau^{th} \) conditional quantile of \( Y_{it} \). Model (3) maybe seen to be obtained from

\[ Y_{it} = \alpha_i + X_{it}\beta + \epsilon_{it} \]  \hspace{1cm} (5)

by imposing the following orthogonality condition: \( \mathbb{E}(\epsilon_{it}|X_{it}, \alpha_i) = 0 \), while model (4) might appear, at a first glance, to be obtained by imposing the following restriction on the model in (5): \( Q_{\epsilon_{it}}(\tau|X_{it}, \alpha_i) = 0 \) (along with the separability assumption on \( \alpha_i \) and \( \epsilon_{it} \)). However, the two models are **not** equivalent, except under special conditions, as indicated in detail in Rosen (2009) and we do not further pursue this path here.

In both cases, the \( \alpha_i \)\(^{10} \) are (a) unobservable (b) potentially cor-

---

\(^{10}\)In general, we will not index the parameters \( \alpha_i \) and \( \beta \) by \( \tau \), except when such an indexing is required, in order to avoid a proliferation of notation, with the understanding
related with the $X_{it}$ and with the residuals of the model (accounting for omitted variables effects) and we maintain the assumption that the $X$ are (conditionally) exogenous i.e. $X|\alpha \perp \epsilon$. The $\{\alpha_i\}$ in the equation (3) represent “unobserved” heterogeneity while such an interpretation in the case of equation (4) is somewhat misleading, since even in the absence of $\{\alpha_i\}$, there is considerable heterogeneity allowed by the model. In fact, the $\{\alpha_i\}$ are not “fixed” in the true sense, since they vary across quantiles, and provide additional flexibility in modeling unobserved heterogeneity (since they vary over the distribution of $Y$), due to which Harding and Lamarche (2009) term them “Quantile Fixed Effects”.

A fundamental issue in both equations is how the $\{\alpha_i\}$ are viewed; as “nuisance” parameters to be estimated or as ancillary parameters aiding identification. In the case of model (2) above, both perspectives yield, from the point of view of identification and estimation of $\beta$, simple solutions; identification is straightaway guaranteed under simple orthogonality condition above while estimation is facilitated by noting that the linearity of the expectation operator ensures that the mean of the difference is the difference of the means. In the case of the quantiles, however, conditions on identification of $\beta$ are not always clear and the “de-meaning” estimation procedure is no longer valid since quantiles of differences are only rarely equal to differences of quantiles.\(^{11}\)

Koenker (2004); Harding and Lamarche (2009); Lamarche (2010) pursue the statistics-based approach of treating the $\{\alpha_i\}$ as “nuisance” parameters and focus on consistent estimation while minimizing variance. In more detail, Koenker (2004) views the least squares problem in model (1) in the context of a penalized estimator with an $l_2$-penalty, the rather special form of the penalty inducing a separation between the estimation of $\alpha_i$, the

\(^{11}\)To see this in a more precise manner, note that the quantile operator does not commute with the “subtraction” operator; formally, this means that while $E(Y_{it} - Y_{i,t-1}) = E(Y_{i,t}) - E(Y_{i,t-1}), Q_{Y_{i,t} - Y_{i,t-1}}(\tau) \neq Q_{Y_{i,t}}(\tau) - Q_{Y_{i,t-1}}(\tau)$, where for notational simplicity we do not explicitly indicate that all of these operations are conditional on $X_{i,t}, \alpha_i$. This implies, in particular, that “first differencing”, which in the case of the conditional mean regression yields a separation in the estimation of the $\beta$, is no longer valid, since the quantile of the first difference is not guaranteed to be identical to the difference of the quantiles.
fixed effects, and $\beta$. Koenker (2004) then proposed both a penalized, with an $l_1$-penalty$^{12}$, and an unpenalized version of FE-QR. The two proposals therein were

$$\min_{\alpha, \beta} \sum_{k=1}^{q} \sum_{j=1}^{n} T_i(w_k \rho_{\tau_k}(y_{it} - \alpha_i - X_{it}\beta(\tau_k))) + \lambda \sum_{i=1}^{n} |\alpha_i| \tag{6}$$

with the unpenalized estimator obtained by setting $\lambda = 0$, where $w_k$ are the weights on the quantiles. The weighting QR pursued here may be seen as yet another variance reduction device. However holding the $\{\alpha_i\}$ fixed across the quantiles imposes strong restrictions on the type of dependence between $\alpha_i$ and $X_i$ as well as on the omitted variable bias such effects are typically used to control for in applied studies$^{13}$.

This approach follows a rich statistical literature and views the problem in (4) above as one of regularization, since the benefits of an $l_1$-penalty are well documented in the statistics literature. In effect, the (scalar) $l_1$-penalty term “shrinks”$^{14}$ the estimates of $\alpha$ in order to reduce the variability induced in the estimation of $\beta$. While appealing from a theoretical perspective, the major drawbacks in this framework are (a) selection of the shrinkage parameter (b) Inference. Both the estimates and the variance matrix are functions of the shrinkage parameters, and while there are results in the case of univariate covariates (Lamarche (2010)), there is no generalization to a multi-variate framework of a method for selection of the shrinkage parameter. An important point to note in the formulation above is that the $\{\alpha_i\}$ are viewed as purely location shift (since they are $\tau$-invariant).

$^{12}$The $l$-spaces mentioned herein refer to the “little $l$” norms in functional analysis, the discrete analogue of the $L_p$-norms. In other words, the $l_1$-penalty corresponds to the absolute value penalty, the $l_2$ to the squared penalty.

$^{13}$For instance, viewing the fixed effects as capturing the effect of unobserved, individual specific variables, i.e. if $\alpha_i = Z_i \gamma(\tau)$, restricting them to be invariant across the quantiles implies that the coefficients of such variables are $\tau$-invariant i.e. $\gamma(\tau) = \gamma$.

$^{14}$There is an extensive statistical literature on the benefits of $l_1$-penalty and shrinkage. In effect, the main arguments are the statistical and computational benefits offered by this approach. For more details, see Tibshirani (1996), as well as [http://www-stat.stanford.edu/~tibs/lasso.html](http://www-stat.stanford.edu/~tibs/lasso.html)
Kato et al. (2010) provide clearer proofs of the asymptotic properties of the unpenalized estimator and investigate the properties of different types of bootstraps for inference in this framework. Further, they also allow the "fixed effects" to vary across quantiles. We therefore follow this approach in our analysis\textsuperscript{15}.

**Identification and Interpretation**

There are three important issues to note with the framework above. First, following Powell (2009), it is important to note that the quantiles are defined, in the approaches outlined above, based on each individual unit’s residual i.e. based on $\epsilon_{i,t} = Y_{i,t} - \alpha_i - X_{it}\beta$ (the alternative definition being $(\alpha_i + \epsilon_i)$) as a result $\alpha$: which one "loses" the position of an observation in the distribution, compared to the cross-section distribution. However, given that the effects identified here (weather on yield) require a panel, and do not make sense in a cross-section, it is evident that the definition used here is the most sensible one. This, however, leads to the interpretation of the coefficients being somewhat different from those in the purely cross-sectional case\textsuperscript{16}.

Second, if year fixed effects are not included in the above framework, then, as pointed out in Powell (2009), the resulting $\tau^{th}$ quantile corresponds to different parts of the distribution for different years, a possibly undesirable artifact in applications. Third, unlike in the case of the linear panel data models (where identification and consistency follows from a simple orthogonality condition between $X$ and $\epsilon$), conditions for identification and consistency of $\beta$ involve high level conditions on the marginal density of the residuals evaluated at zero and on moment conditions on $X_i$ (Assumption A3 in Kato et al. (2010)), due to which they have no interpretable or intuitive content.

\textsuperscript{15}All of the frameworks considered above are subject to the incidental parameter problem (see Lancaster (2000) for an detailed introduction to the topic in econometrics), which is "solved" in all cases above by allowing both $n$ and $T$ to diverge at specific rates.

\textsuperscript{16}It is, in this context, important to note that if the use of a panel framework is for the purposes of identifying the effects of a variable which is believed to be endogenous in the cross-section case (such as for instance in the hedonic case), then the framework outlined above is likely inappropriate, and a better suited one is the GMM approach outlined in Powell (2009).
Covariance Matrix Estimation

For the unpenalized case, Koekker (2004) derives the asymptotic variance of the estimator and points out that bootstrap inference maybe more accurate (another cause for concern is that the asymptotic variance requires $T$ to grow "fast enough", compared to $N$; in particular, $\frac{N^a}{T} \to 0$ for some $a > 0$). Kato et al (2010), for the same unpenalized model, also point out that asymptotics require, in their case, $T$ to grown faster than $N^2$, a clearly restrictive condition, and they recommend the cross-sectional bootstrap as being the better approach, after comparing both cross-sectional and Time bootstraps for their coverage accuracy. We note however that their comparisons are for coverage for *individual coefficients* only.

There is a large literature on the use of the bootstrap for confidence interval estimation in a cross-section QR setting (see for instance Buchinsky (1995); Fitzenberger (1998); Fitzenberger et al. (2002) among others). For the Panel data case, there are three possible basic version of the bootstrap, which, following Kapetanios (2008), are termed "Time" bootstrap, with resampling across the time dimension, "Cross section", with resampling across the "individual unit" dimension and a combination. We note that all three versions correspond to the "paired" or the "x-y" bootstrap, involving bootstrapping the data matrices rather than the residuals. We use the simple time bootstrap for inference, which accounts for the unknown spatial dependence but not for time dependence. Details of estimation of the standard errors (and therefore, the t-statistic) are provided in the Appendix B.

Extension to Semi-parametric Quantile Regression

We outline an extension of the above linear conditional QR-FE estimation strategy to semi-parametric settings. Consider the linear FF-QR estimation detailed above:

---

17In more detail, denoting $t_{i_1}, \ldots, t_{i_T}$ the resampled time indices (assuming for now a balanced panel data), time resampling involves forming the full bootstrapped data, $Z^* = (Y^*, X^*)$ where $Y^* = (Y_{t_{i_1}}, Y_{t_{i_2}}, \ldots, Y_{t_{i_T}})$ and $X^* = (X_{t_{i_1}}, \ldots, X_{t_{i_T}})$ where $Y_t, X_t$ denote respectively the $N \times 1$ and $N \times K$ matrix of observations for all $N$ individual units. Cross-sectional resampling is entirely analogous, with $I_{j_1}, I_{j_2}, \ldots, I_{j_N}$ denoting the resampled individual unit indices, $Y^* = (Y_{I_{j_1}}, Y_{I_{j_2}}, \ldots, Y_{I_{j_N}})$, $X^* = (X_{I_{j_1}}, \ldots, X_{I_{j_N}})$ with $Y_{I_{j_n}}$ being the resampled $T \times 1$ vector of observations on individual $I_{j_n}$ for all $T$ time periods and $X_{I_{j_n}}$ the $T \times K$ matrix of covariates for individual unit $I_{j_n}$. 
\[ Q_{Y_{it}}(\tau|\alpha_{it}, X_{it}) - \alpha_{i}(\tau) + X_{it}\beta(\tau) \]  

(7)

and observe that, once a particular estimation framework is settled upon, the problem above is in essence a straightforward linear QR problem (with more involved conditions for consistency). Consider instead the more general semi-parametric conditional quantile estimation problem:

\[ Q_{Y_{it}}(\tau|\alpha_{i}, X_{it}) - \alpha_{i}(\tau) + X_{i,t}^I\gamma + g(Z_{it}) \]  

(8)

where \( Z_{it} \) is (for now) a univariate variable to be modeled non-parametrically. There exist two possible approaches to non-parametric modeling of quantiles (see Koenker (2005, Chapter 7) for a survey), the local-polynomial based one, and a penalized spline version. We note however, that even without these options, the

function \( g(Z_{it}) \) may always be approximated by an orthogonal polynomial of a suitable order, and the problem is one of estimation of the order of the polynomial. Inference is even more problematic in this setting; however, since consistency is likely guaranteed (under possibly additional conditions), a bootstrap approach maybe easily used in this case (and this is the approach pursued here).

The alternative approach, of penalized splines, is a different variant of the same approach, in that the scalar penalty term must be chosen, generally through means of AIC-like measures (see Koenker (2010) for details and an application); we briefly pursue this approach here, more as a robustness check, but note that in most non-parametric QR cases, inference is generally problematic and computationally expensive, and this is especially true in the FE case.

**Model Specification**

The basic model we consider is of the following form, for a given crop:
\[ y_{it} = \alpha_i + \beta_1 GDD_{it} + \beta_2 GDD^2_{it} + \beta_3 Precip_{it} + \beta_4 Precip^2_{it} \\
+ \sum_{j=1}^{J} \sum_{k=1}^{K} \beta_{5+k,j} Trend^k \ast R_j \]

where \( y_{it} \) is the log of yield, \( R_j \) is a region indicator (we use 5 regions, North, West, South, East and Central with "North" being the omitted category), \( K = 3 \) is the order of the time trend. A second possible specification where we use, in addition, monsoon onset at the district level and the number of dry spells for the summer growing season, is

\[ y_{it} = \alpha_i + \beta_1 GDD_{it} + \beta_2 GDD^2_{it} + \beta_3 Precip_{it} + \beta_4 Precip^2_{it} \\
+ \sum_{k}^{K} \sum_{j}^{J} \beta_{5+j,k} Trend^k \ast R_j + \beta_{5+J+1} Onset_{it} + \beta_{5+J+2} dryspells_{it} \]

The third specification is where we use monthly precipitation data instead of using seasonal totals as above.

\[ y_{it} = \sum_{k=1}^{K_1} \text{precip}_{sit} \beta_k + GDD_{it} \beta_{K+1} + GDD^2_{it} \beta_{K+2} + \sum_{j=1}^{K} \text{precip}^2_{kit} \delta_j \\
+ \sum_{j}^{J} \sum_{k}^{K} \gamma_{k,j} Trend \ast R_j + \nu Onset_{it} + \zeta dryspells_{it} \]

with \( K_1 = 6 \) being the number of months in the winter (rabi) growing season, and \( K_1 = 4 \) the number of months in the summer growing season (kharif), with onset and dry spells included for the summer growing season only. Another covariate added to many regressions is (where available) the ratio of irrigated area to un-irrigated area at the district level. Finally, for the winter season which spans 6 months, instead of including rainfall for all 6 months separately, in one specification, we aggregate them into 3 separate
periods, early, middle and late season rainfall\textsuperscript{18}.

\textbf{Data and Summary Statistics}

\textbf{Agricultural Data}

We use district-level data on agricultural outcomes from the commercially available \textit{Indian Harvest} database from the CMIE. The database includes crop-specific data on district-level yield, area sown, farm harvest price and a very few input data, in addition to irrigated area by crop. Data on inputs are very unreliable, and we do not make use of them. The dataset nominally spans the time period 1950-2006 but data for a large proportion of the districts are available only from 1970 onwards and we therefore use data from 1971. The dataset covers most of the major crops (including what are termed “cash crops” but not plantation crops such as “Tea”, “Coffee” and “Rubber”) and all the states and districts. The present analysis is restricted to the major cereals rice and wheat. The data are essentially a compilation of the official statistics reported by the state governments and the Ministry of Agriculture of the Govt. of India.

The CMIE’s \textit{Indian Harvest} database has only been available since 2003. Previous research for India on this topic used a dataset compiled by the World Bank (WB) (Sanghi et al. (1998)) which pre-dates the CMIE’s database and is essentially similar. The main differences stem in geographical coverage and homogeneity of district definition. The WB dataset covers a large proportion of the agriculturally important states while the CMIE database covers all states. Further, data on the large and agriculturally important states of Bihar and West Bengal are mostly missing in the extension by Duflo and Pande (2007).

Indian district boundaries change periodically, with larger districts generally split into smaller ones (sometime combinations of districts are split into larger numbers but this is rare), creating difficulties for defining a district as the unit of observation. Further, no year-by-year spatial definition of the districts are available, and it is not always clear what year a district was changed; we therefore take the year in which data for a district

\textsuperscript{18}For instance, there is generally little rainfall during the months of Feb-March, for instance, and therefore including them separately provides little additional knowledge.
appears in the agricultural data as the “year” in which the said district was created. Yet another difficulty lies in that the “mother” district, from which the new district is created, exists and has the same name as before. However, in many cases, the agricultural database indicates that a change has occurred\textsuperscript{19}, and this allows a largely\textsuperscript{20} comprehensive accounting of (a) which new districts are created (b) which existing districts have altered and at what point of time.

There are two ways to proceed to obtain uniform district data sets. The first approach, pursued in Duflo and Pande (2007) (an extension of the earlier dataset compiled by Sanghi et al. (1998)), is one in which districts were defined based on the most primitive definition and are therefore consistent in time. The second approach, and the one we pursue, is to (a) track when new districts are created (b) treat the new districts as being distinct from the original ones\textsuperscript{21}. We argue that this is a better indicator of the districts for the following reason: in general, districts are altered for ease of administration and/or for historical reasons. It is then evident that we anticipate that two cistricts to have different characteristics (in addition to the obvious one of different area), and therefore, separate fixed effects for each is more appropriate when the fixed effects are viewed as accounting for “unobserved heterogeneity”. In order to obtain geographical co-ordinates for each district (primarily the district center), we make use of GIS maps of India for two time periods, 1991 and 2001. This allow us to keep track of how district centroids change with time, following the approach outlined above in treating districts which undergo any changes as separate districts (although most districts are already small enough that even after a change, their centroids are either almost unchanged or change at most by half a

\textsuperscript{19}The timing of the change is generally restricted to the census years, due to the nature of reporting of data.

\textsuperscript{20}This is “largely” but not completely comprehensive since it is not clear if all districts which have altered even slightly have indicators in the database but this issue is likely to be minor enough to be ignored.

\textsuperscript{21}Thus, if district “X” was split in year “Y” into districts “X(1)” and “Z” (the “(1)” is simply an indicator—generally a phrase such as “up to 1981” etc—allowing one to distinguish between the pre- and post districts), then we treat districts “X”, “X1” and “Z” as all distinct, with district “X” being in existence prior to year “Y”, and districts “X1” and “Z” existing post year “Y”. This approach of course leads to the number of districts being different at different time points (generally, the census years).
degree)

Weather Data

Temperature Data

Schlenker and Roberts (2006, 2009) have illustrated the superiority of using fine-scale temperature measures to account for the cumulative effects of variations in temperature on crop yield. Much of this research has been confined to the developed countries (an exception is Guiteras (2008)) due to lack of availability of such data for the developing nation setting. Given the paucity of station-level temperature and rainfall data for India, prior work (Guiteras (2008)) makes use of a corrected reanalysis dataset, consists of gridded (at the 1° × 1° resolution) 6-hourly measurements of temperature from 1949-2000. This study makes use of gridded (1° × 1° resolution) daily rainfall and temperature products recently created and made available by the IMD (Srivastava et al. (2009)), and to our knowledge, this is the first paper in economics to make use of this dataset, spanning 1969-2005.

To create daily district-level data from gridded, we use a weighted (using standard weights in the Climate literature, the inverse square root of distance) average of data from all grids which are within a 100 KM radius from the district centroid. Due to the nature of agricultural policy, yields across space are expected to be related (especially within a state and region); further, the mode of creation of the weather variables (gridding) induces additional dependence across space, due to which dependence of the regression residuals are likely to be dependent across space. Following recent literature on bootstrapping the variance estimators for panel data with spatial dependence (Gonçalves (2008)), we illustrate a simple bootstrap approach which accounts for this, even in the relatively complicated estimation framework employed here.

---

22 The approach in the CMIE database to new states is to treat these states and their districts to have existed for the entire period and to simply relabel the districts originally in another state to have been in the new state. For instance, districts which were in Madhya Pradesh and were later a part of Chhattisgarh are treated as being, for all time periods, in Chhattisgarh. We follow this convention since it is also more sensible when reporting effects at the state level.

23 Alternative radii of up to 200 km yielded very similar seasonal summary statistics.
Precipitation

For precipitation, we again make use of a gridded daily rainfall data set product, spanning 1951-2005, developed by the IMD (Rajeevan et al. (2005)). To our knowledge, this is this first paper in Economics to make use of this dataset. Creation of district-level data from gridded proceeds in the same way as for temperature. The significance of using actual data, in comparison to reanalysis data, for India is illustrated elsewhere (and is available upon request) but the main differences are that the fidelity to actual rainfall varies with reanalysis datasets, and the NCC dataset in particular appears poorly constrained due to the lack of actual data for India.

Summary Statistics

Summary statistics of the key variables used, by region, are presented in Table (1) for wheat and Table (2) for rice. Three points are noteworthy about Table (1): the predominance of the Northern region in wheat production, its high yields and the very high irrigation ratios. Turning now to rice, it is evident that the Eastern region dominates in rice production, followed by the Southern and Northern regions; in addition, there is a significant difference in rainfall between the Eastern regions and the rest of India. Finally, the relatively late onset of the monsoon is evident in the Northern and Western regions.

Regression Results

As indicated earlier, for each of rice and wheat, two sets of regressions are carried out for each specification of GDD: using seasonal rainfall and using monthly rainfall. For the sake of brevity, and since in our climate change specifications we do not use monthly rainfall changes, we report and discuss results using only seasonal rainfall. The regression strategy is as follows: interest centers on the estimation of the $\tau^{th}$ quantile of $\log(yield)$, as in equation (9) for the parametric case and equation (8) for the non-

---

24 The CMIE database, and in general Agricultural statistics in India, reports irrigated area by crop and year, instead of by season. However there are, for certain districts in south eastern India, two rice growing seasons due to which it has not yet been able to establish the extent of irrigation during the kharif season. We therefore do not have potentially important information regarding irrigation in the kharif season.

25 Nearly identical results were obtained when monthly (in case of kharif) and bi-monthly (in case of rabi) rainfall were used instead.
parametric case. The actual equations estimated, however, are (9), for the rabi season, and (10) and (11) for kharif, substituted into equation (6) (with equal weights and no penalty, as indicated earlier).

We follow much of the literature (see Fitzenberger et al. (2002)) and chose the first three quartiles, along with the 90th percentile: \( \tau = (0.25, 0.5, 0.75, 0.9) \) for estimation. All regressions include region-specific cubic time trends whose coefficients, along with those of individual fixed effects, are not reported since they are not of intrinsic interest. We avoid the use of year fixed effects for two reasons, to better capture region-specific heterogeneity and to avoid numerical issues with the linear programming problem. An undesirable consequence of this choice is that the interpretation of the coefficients at each quantile are somewhat different from those in a pure cross-section, as indicated earlier. As indicated above, both asymptotic and (time) bootstrap t-statistics are reported.

**Wheat**

Results of estimation are provided in Tables (3) and (4). Given the log-linear nature of the regressions, we may interpret the coefficients as approximately corresponding percentage changes. Turning first to effects of temperature, two conclusions are evident. First, there appears to be an inverted U shape relationship between yield and temperature, indicating that temperature up to a certain point is beneficial, and harmful thereafter. This is in keeping with the results obtained in Schlenker and Roberts (2009) and in Guiteras (2008), using very different methods. A second, and more surprising, one is that the differences across quantiles of yield of the inflection point is minimal, indicating a significant degree of uniformity in impacts across differing crop and ecology types (controlling for irrigation and rainfall). Rainfall also appears to have a similar relationship to yield as does temperature and interestingly, the coefficients across different quantiles is very different (for instance, in absolute terms, the coefficients in many instances differ by upto 50%), reflecting the high spatio-temporal variability in winter season

---

26It proved difficult to estimate the year fixed effects, given the unbalanced nature of the regression sample and the fact that both individual fixed effects and year fixed effects must be estimated. For the same reason, it is even more infeasible in this approach to estimate region-specific or district-specific year fixed effects.
rainfall (evident in Table (1)).

Yet another interesting point is that while harmful degree days appear to have (for the most part) the correct sign, they are never significant, indicating that, when accounting for a richer interaction between temperature and rainfall (as the log specification does\textsuperscript{27}), the additively separable nature of harmful degree days yields little additional information. Irrigation, unsurprisingly, is highly significant and positive at all quantiles, with a possibly smaller coefficient at the 90\textsuperscript{th} percentile. Results using bi-monthly rainfall totals are almost identical (and are available upon request), although given the highly variable and spatially concentrated nature of rainfall, it is difficult to interpret the coefficients on rainfall.

Finally, in order to assess if the results are influenced by the quadratic specification for GDD, we use two, very different, non-parametric approaches; first, we use an orthogonal polynomial expansion for growing degree days, using polynomials of degree 4, and second, we use the penalized spline approach outlined above. The coefficients for other variables, using either approach, appeared to be identical (we do not therefore report coefficients estimated with the spline specification), as illustrated for the case of orthogonal polynomials in Table (4). The estimated effect of gdd on yields using non-parametric splines, in figure (6), indicates very clearly an initial increase, followed by a subsequent decrease (the inflection point also appears very close to the ones obtained using simple quadratic specification above). Thus, the results are robust to functional form assumptions.

Rice

In the case of rice, GDD appear to have a U-shaped relationship, quite contrary to intuition and agronomic belief, as indicated in Table (5). Further, this behavior is consistent across quantiles, indicating that the relationship is global (in the distribution function). The relationship with rainfall, however, appears to be a more conventional \(\cap\)-shape, and rainfall is seen to be significant at all quantiles. Onset day appears to be a poor predictor, with a positive significant coefficient at the lower quantiles, with no effects at the

\textsuperscript{27}We note that the use of a log specification implies some amount of substitutability between heat and rainfall, which is realistic. It also limits the interactions to be multiplicative.
upper quantiles. Dry spells, on the other hand, have the expected negative sign, and is significant at all quantiles. In order to assess if the results are impacted by the simple, quadratic functional form assumptions, both orthogonal polynomials and penalized splines were used to model GDD.

Since the quadratic specification for GDD performs adequately (apart from a well known tendency to inflate variance of the remaining parameters), we do not report results from an investigation of the orthogonal polynomial coefficients, which are qualitatively similar. Further, the fully non-parametric approach to estimating the effect of GDD, as indicated in figure (5), reinforces the robustness of an inverted U-shape relationship. Thus, it is clear that the relationship outlined, and the lack of significance, are robust to various functional forms.

These results are unaltered when monthly rainfall is used instead of seasonal rainfall. Yet another concern could be the omission of irrigation, which is of some importance for many regions, especially the northern region. Given the lack of reliable data on irrigation at the district level for rice by season, it is not possible to directly address this concern. However, we argue that it is very plausible that the possible bias due to omission of irrigation is unlikely to drive our results, by noting that onset days are already not significant, at the upper part(s) of the distribution, which correspond (imperfectly but largely) to irrigated districts. Thus, the concern that these coefficients are larger (or significant) due to the omission of irrigation is unlikely to be substantially responsible for these results.

Intriguingly, however, recent results using farm level data on irrigated agriculture (Welch et al. (2010)) indicate that maximum temperatures are beneficial for rice crops, despite agronomic evidence to the contrary. While the study had only one site in India, in Tamil Nadu, with maximum temperature significantly lower than those in say Northwestern India, the impact of the interaction of high temperatures with irrigation may well explain a part of the results obtained here.

**Comparison with the Standard, Conditional Mean Approach**

In order to illustrate the benefits of the current approach, consider the results of using an identical specification in a standard, linear panel data
framework, results for which are in the final columns of Table (3) to (5). Since the comparison is identical when using orthogonal polynomials, we focus on the simple, quadratic-in-gdd specification.

Consider first, the case of wheat, in Table (3), where magnitudes of differences in coefficients across quantiles are smaller than those for rice. Two points are evident from a perusal of the results: (a) results of the mean regression are very similar to those of the quantiles when there is little difference in coefficients across the quantiles, as for the case of GDD (b) when there are substantial differences across quantiles (as in the case of irrigation and rainfall), mean regression provides a very incomplete characterization of the conditional distribution. For instance, in the case of irrigation ratio, the coefficient in the mean regression is 7.95, which is very close the impact of regression at the 90th percentile, and is substantially (about 25%) smaller than at the first quartile.

These differences are even larger in the case of rice (Table (5)). Consider for instance the coefficient on either onset day or dry spells; these coefficients are, in many instances, substantially (a factor of 4, in the case of onset) different. More importantly, the variation in coefficient(s) across quantiles (which are statistically significant, see below), are smoothed over by a focus on the mean.

It is therefore clear that substantial heterogeneity in relationships exist at different quantiles and that the mean regression framework, in this case, provides a very incomplete characterization of the conditional distribution of yield. The substantial differences in impacts using these two frameworks are fully illustrated below.

**Climate Change Scenarios**

Climate change scenarios for econometric studies of impacts on agriculture has focused on GCM projections, for the most part the Hadley Center models (Schlenker and Roberts (2009); Guiteras (2008); Deschenes and Greenstone (2007)). However, for the US, Schlenker and Roberts (2009) used downscaled GCM projections while Guiteras (2008) uses the coarser GCM daily output. GCM daily projections, however, are known to have significant biases which make their use for agricultural purposes quite tricky (Ines
et al. (2010)). Further the coarseness of the GCM output tends to lead to a large degree of spatial smoothing, a very undesirable feature. While we hope to report on downscaling GCM output for India and their applications in a future iteration, we will use simpler scenarios for the future at present.

Consistent with IPCC projections and with prior literature on crop modeling for India, we consider the following scenarios for kharif (summer growing) season: uniform increases in temperature over the growing season of 1 and 2°C, coupled with uniform increases in rainfall of 10%. It is clear that, using this approach, the increases in rainfall are significantly overstated (since model projections expect reduction in rainfall over certain regions of India, especially the Northwest). Further, many GCM’s also project increase in monsoon rainfall variability, which we interpret as increases in probability of late monsoon (by a uniform 10%) and increases in dry spells (again by a uniform 10%), given model projections of more intense rainfall coupled with reduction in rainfall frequency. However, for the results below, we do not apply changes in sub-seasonal characteristics, in order to minimize dutter (and since modest changes in these variables do not substantially affect yield).

For the rabi (winter growing) season, there have been very few studies on possible impacts; however, most indications are for increased minimum (therefore average) temperatures, and little changes in rainfall. We translate this, in line with prior crop modeling efforts, into uniform 0.5 and 1°C increase in temperature, and an unchanged rainfall distribution. We emphasize that these scenarios are consistent with those used in the crop modeling literature (Mall et al. (2006b,a)) as well as prior econometric approaches (Sanghi et al. (1998); Guiteras (2008)).

**Climate Change Impacts**

The impacts of climate change are computed, at each quantile, as the differences in predicted values under current and future weather. In more detail, denoting with superscript *future* realizations of weather, we compute

\[
\Delta Y_{it}(\tau) = \bar{Y}_{it}^f(\tau) - \bar{Y}_{it}(\tau)
\]

(12)

as the predicted change. In other words, predicted values are ob-
tained, using coefficients estimated under current weather, for both current and future weather and the difference in predicted values are interpreted as the impact of change in climate. We present the full distribution of predicted changes, instead of focusing on summary changes (such as changes at the mean of the covariates) in order to assess the full range of changes predicted. The distribution of the changes are summarized using box plots.

**Wheat**

There are two possible scenarios for winter, as indicated above. Given the signs of the coefficients, one anticipates significant losses in yield under increased warming. The results in Table (6) indicate that increasing temperatures indeed are harmful. Further, the magnitude of losses are decreasing in quantiles for every scenario considered, indicating a significant degree of heterogeneity in the direction (not magnitude) of impact across a range of growing conditions. Finally, median losses range from 2 to 5% in the relatively benign 0.5°C temperature increase scenario to a substantial 5–10% in the 1°C warming scenario, with many districts losing as much as 22% of yield. Figure (1) provides a visual description of the magnitude of losses: for scenario 2, losses rise to as high as 20% for the lower percentiles, indicating significant losses for the already lower productive districts.\(^{28}\)

Turning now to assessing regional impacts, we note that the southern region, already quite hot, has the highest losses (although of less importance in wheat production nationally) in both scenarios, as indicated in figure (2) (we do not include the figure for regional losses under scenario 2, since the regional distribution is identical). However, at all percentiles, western and central regions, regions of already large poverty and substantial wheat production, also lose significantly, up to 6% of yield with a uniform

\(^{28}\)The interpretation of the coefficients are somewhat different from the cross-section case, since the quantiles are defined for \(i\) and \(t\); thus, one can only speak of the “low” or “high” yielding district-year combinations, instead of low (high) yielding “districts” or “years” separately. However, for substantial parts of India, yields tend to have relatively low variability and only regions with low yield on average tend to have high variability. Therefore, we loosely interpret district-year combinations with low and high yields as “districts” with low and high yield. We note that such terminology is clearly inappropriate for the quantiles 0.5 and 0.75, since these can correspond to “low” (“high”) years in high (low) yielding districts, and therefore, we can only refer to them by the less useful “district-year” terminology.
increase of $0.5^0C$ and 9%, with an increase of $1^0C$. Losses in the Northern region are smaller, at about 2 – 5%, whose impacts however can translate into large welfare losses due to the very high production losses embodied.

**Rice**

The impact of climate change on rice (from Table (6)) appears somewhat dissimilar. There are minor increases in yield, of up to 0.5% (median) at the 50th percentile, with moderate reductions in yield, of up to 1.67%, at the higher quantiles, in Scenario 1. Under Scenario 1, reductions are predominant everywhere except at the median. The results so far can be summarized as indicating moderate reductions in yield at most low and high quantiles, which increased reductions under enhanced warming. However, as clearly seen in figure (3) (unlike in the case of wheat wherein there are only a very few districts whose impacts differ in sign) a substantial number of district-years, at most quantiles, experience significant (>5%) yield reductions, indicating significant heterogeneity in impacts.

Turning now to a closer analysis of the regional distribution of losses and gains, it is striking, from figure (4) that most of the gains are driven by the Eastern and Northern regions, which has a very large share in area and production of rice, while most of the losses are driven by losses in parts of the Southern region, and this is consistent across quantiles and scenarios (figures for scenario 2 not shown but are available upon request). The results of the regional analysis help inform the pathways to possible gains. The impact of climate change on rice cultivation in eastern India has been investigated, with some studies indicating possible increases in yield. Further, the study by Welch et al. (2010) already cited reports a similar increase in south-eastern India. This study therefore adds to the weight of evidence on this important issue.

The result of increased rice yields in Northern India is more puzzling; this is the region with among the highest average temperatures (see Table 1) and most crop models predict reduced yields under any warming scenario. However, this is also the region with among the highest rice yields in India, with yields rising over a period (1970-2000) presumably along with increased surface temperatures. The fact that the result persists with an
econometric specification which includes year-fixed effects (but not at the district-specific level) is quite intriguing. It is possible that what is picked up by the specification is the district- or region-specific productivity growth, which we intend, in a subsequent iteration, to capture using region-specific year fixed effects.

There is substantial prior literature for India employing crop simulation models for an investigation of impacts of climate change on yields of various crops (see for instance Mall et al. (2006b); Mall and Aggarwal (2002); Mall et al. (2006a); Attri and Rathore (2003); Kumar and Parikh (2001); Chatterjee (1998)). The results of such experiments have been varied, but for the crops considered here, quite clear; most studies report minimal direct impact (defined as simply an increase in daily temperatures) of climate change on yields of major crops, including rice, grown in the kharif season, with modest or substantial positive impacts on rice yields, in particular, for various regions (specifically, minimal impact on Northwestern and Central India). This is consistent with the results obtained here.

Second, most studies report significant negative impacts on crops grown in the rabi (winter) season, primarily due to increase in night time temperature. Our results also indicate, for the moderate scenarios considered here, substantially negative impacts on wheat yields, mostly in the southern and Central regions, which is consistent with the results obtained in many crop modeling efforts, and modest increases in rice yield, driven for the most part by increases in the eastern and northern regions.

**Comparison with the Mean (fixed effects panel) Regression**

In order to compare the predictions obtained from a FE-QR framework with those of a linear FE framework, we consider box plots of predicted changes (as for the FE-QR case) in figure (7)\(^\text{29}\). For wheat, the predicted median impact under the two scenarios, of \(-3\) and \(-7\)% are comparable to those of the median, in figure (1). However, it is evident that under the reasonable scenarios considered here, mean impacts cannot fully characterize the substantial differences in impacts of changes in weather on yield. A similar

\(^{29}\text{In the interests of brevity, a table similar to the FE-QR case is not presented. It is however available upon request.}\)
result is seen with rice, from figure (3), i.e. that the mean regression framework is not suited to characterizing the heterogeneity of impacts of (changes in) weather on yield. The inadequacy of FE panel regression in capturing heterogeneity in many interesting applications has been repeatedly emphasized in Galvao (2009), Harding and Lamarche (2009) and Powell (2009) and this study adds to the weight of evidence.

CONCLUSIONS
This paper employed a newly developed panel data quantile regression methodology to first estimate the relationship between current weather and agricultural yields of the two most important food grains for India, rice and wheat. The framework used allows a fuller exploration of the conditional distribution of yield beyond just the conditional mean, which is the focus of most studies. Results of the estimation, when projected onto moderate, and almost universally accepted uniform climate change scenarios for India, indicate a significant negative impact on wheat yields, of up to 11%, primarily in Southern and Central India, with more moderate losses in the more important Northern region. Further, these impacts were seen to be most negative on the most productive districts, indicating losses in production which are likely significant.

For rice, however, impacts of uniform warming of up to $2^\circ C$ were shown to be moderately negative, with reductions in yield (upto 3% at the highest quantile) concentrated at the upper quantiles, or in the most productive areas, while impacts at the lower and intermediate quantiles are seen to be very mildly positive. This is likely to translate into a moderate reductions in production of rice. The results here are consistent with many estimates of changes in rice yields obtained using various crop models, under a variety of climate change scenarios, as well as a recent study of rice cultivation in Asia under irrigated conditions (Welch et al. (2010)).

These results indicate significant reduction in wheat production, and more moderate reductions in rice production. However, given the substantial number of outliers in rice with yield losses, these estimates likely translate into potentially substantial decreases in rice production, at least at the local scale. In summary, this study indicates that in the absence of
significant changes in agricultural practices and technology, climate change is likely to lead to increases in food insecurity for India’s poor, as a result of decreased yield at the national (or regional) level.
REFERENCES


APPENDIX A

PRECIPITATION MEASURES

The monsoon is, at its most basic, a large scale atmospheric phenomenon which determines moisture availability and whose interaction with local features determines actual local rainfall. Therefore, a definition of "onset" of the monsoon is only climatologically meaningful at a suitable large scale, generally much larger than a district. This is the major reason that the IMD defines onset as occurring at a particular location on the Southwestern coast of India (Moron and Robertson (2009) and references therein). On the other hand, these definitions of the monsoon are completely divorced from local rainfall and are uninformative to the farmer.

Recently, there has been renewed interest among the Climatologists in attempting to hydrologically define onset of the monsoon for India. Fasullo and Webster (2003) provide a definition of onset in terms of vertically integrated moisture transport, which is valid only for larger regions than considered here. Moron and Robertson (2009) provide a more local, at the grid-level, definition of monsoon onset and provide evidence that it does correspond to actual onset at selected locations. Their measure of onset, at the grid-level, maybe defined simply as the first day after April 1 that rainfall for X days exceeds a certain threshold without being followed by Y days in succession without rainfall. The rationale for this is straightforward: onset is a phenomenon involving rainfall on a few successive days, while the second condition ensures that there are no "false start". As before, we compute district-level onset from grid-level onset. Finally, in order to proxy for substantial periods with no rainfall, we compute the number of "dry spells" in the season, wherein a dryspell is defined as at least 7 days with rainfall below 5mm, a very low threshold. Again, this is computed at the district level.

APPENDIX B

DETAILS OF THE BOOTSTRAP PROCEDURE USED

For unbalanced panels, which in practice are the commonest forms of panel data, the modifications to the time bootstrap are straightforward but algebraically tedious and will not be illustrated here. We will however point out that the same method outlined above maybe applied with resampling now being specific to each individual unit. In the absence of dependence in the
error structure, we may simply pursue a time bootstrap. However, given (at least in our application), the suspicion of strong dependence in both space and time, such time-bootstrap-based inference will clearly be biased (in uncertain directions). A general but heuristic solution, when spatial dependence is suspected, is to use the time bootstrap and when spatial dependence is suspected, to use the cross-section bootstrap (as indicated in for instance Kapetanios (2008)). In the fixed effects quantile regression setting it is not clear that there is any asymptotic justification for this approach to the covariance matrix estimation (due to the lack of theoretical results on the second order accuracy of the bootstrap).

Denoting the individual coefficients (on the covariates of interest) \( \beta \), which is of dimension \( K \times 1 \), the estimated variance of the \( k^{th} \) component of \( \beta \) as \( \hat{\sigma}_{\beta_k}^2 \), \( n = \sum_i^N T_i \), the sample size, we have that the \( t - statistic \) for the test \( H_0 : \beta_k = 0 \) of the \( k^{th} \) coefficient is:

\[
t_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}}
\]

\[
\hat{\sigma}_{\beta_k} = \sqrt{\frac{C_{n,k}^*}{n}}
\]

where, in the notation of Gonçalves and White, 2005, \( C_{n,k}^* \) is the \((k,k)^{th}\) element of the bootstrapped variance matrix \( C_n^* \) with

\[
C_n^* = \frac{n}{B} \sum_{j=1}^B \left( \beta(j) - \beta^a \right) \left( \beta(j) - \beta^a \right)^T
\]

(13)

where the pivotal values used, \( \beta^a \), is the bootstrapped mean i.e.

\[
\beta^a = \bar{\beta}^* = \frac{1}{B} \sum_{j}^B \beta(j)
\]

A final point to note is that one may obtain the critical values by bootstrapping the variance of \( \sqrt{n}\hat{\sigma}_{\beta_k}^* \), an approach corresponding to the so-called “double bootstrap”. A major drawback of this approach is the computational burden imposed, due to which we do not pursue the approach in the present instance. The simplified “Fast Double Bootstrap” of
Davidson and MacKinnon (2007) is likely to provide, in this case, the required improvements without the same computational burden and we leave this approach to future research.

Note again that the validity of the above approach, including the double bootstrap, has been analysed in detail in Gonçalves and White (2005) for the case of the cross-section regression models, and they conjecture that the results are likely to be valid for the case of cross-sectional quantile regression. If such is the case, then it is likely to also be valid for the case of quantile regression for fixed-effects panel data, under slightly different conditions.
## Tables

Table 1: Descriptive Statistics of Regression Sample for Wheat (rabi season).

<table>
<thead>
<tr>
<th></th>
<th>Central</th>
<th>North</th>
<th>South</th>
<th>West</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yield (kg/hectare)</strong></td>
<td>1054.8</td>
<td>2056.1</td>
<td>786.1</td>
<td>1313.0</td>
<td>1551.2</td>
</tr>
<tr>
<td>(506.5)</td>
<td>(829.5)</td>
<td>(438.5)</td>
<td>(592.0)</td>
<td>(849.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Production (000's of ton)</strong></td>
<td>80.24</td>
<td>335.4</td>
<td>7.089</td>
<td>84.51</td>
<td>193.5</td>
</tr>
<tr>
<td>(63.28)</td>
<td>(297.1)</td>
<td>(15.09)</td>
<td>(106.3)</td>
<td>(251.9)</td>
<td></td>
</tr>
<tr>
<td>**Seasonal rainfall (mm)</td>
<td>88.13</td>
<td>105.0</td>
<td>185.0</td>
<td>67.69</td>
<td>102.7</td>
</tr>
<tr>
<td>(68.31)</td>
<td>(73.48)</td>
<td>(101.9)</td>
<td>(65.68)</td>
<td>(81.44)</td>
<td></td>
</tr>
<tr>
<td><strong>Growing degree days (deg Celsius)</strong></td>
<td>2808.2</td>
<td>2444.1</td>
<td>3383.1</td>
<td>2845.5</td>
<td>2708.1</td>
</tr>
<tr>
<td>(217.9)</td>
<td>(311.5)</td>
<td>(164.4)</td>
<td>(351.3)</td>
<td>(417.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Harmful degree days (deg Celsius)</strong></td>
<td>8.499</td>
<td>4.146</td>
<td>32.16</td>
<td>15.94</td>
<td>10.43</td>
</tr>
<tr>
<td>(8.898)</td>
<td>(5.702)</td>
<td>(23.50)</td>
<td>(14.92)</td>
<td>(14.70)</td>
<td></td>
</tr>
<tr>
<td><strong>Average seasonal temp (deg Celsius)</strong></td>
<td>21.44</td>
<td>19.41</td>
<td>24.74</td>
<td>21.68</td>
<td>20.90</td>
</tr>
<tr>
<td>(1.233)</td>
<td>(1.727)</td>
<td>(0.993)</td>
<td>(1.985)</td>
<td>(2.353)</td>
<td></td>
</tr>
<tr>
<td><strong>irr_ratio</strong></td>
<td>0.407</td>
<td>0.849</td>
<td>0.596</td>
<td>0.673</td>
<td>0.692</td>
</tr>
<tr>
<td>(0.293)</td>
<td>(0.180)</td>
<td>(0.334)</td>
<td>(0.274)</td>
<td>(0.301)</td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>911</td>
<td>2053</td>
<td>486</td>
<td>893</td>
<td>4343</td>
</tr>
<tr>
<td><strong>Number of district</strong></td>
<td>45</td>
<td>97</td>
<td>28</td>
<td>52</td>
<td>222</td>
</tr>
</tbody>
</table>

Note: Mean coefficients reported, with standard deviations in parenthesis.
<table>
<thead>
<tr>
<th>Table 2: Descriptive Statistics of Regression Sample for Rice (kharif season).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>log(yield) (kg/hectare)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Production (000's of tonnes)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Seasonal rainfall (mm)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Growing degree days (deg. celsius)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Harmful gdd (deg. celsius)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Day of monsoon onset (days from April 1)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dry spells (number of days)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total number of dry days</td>
</tr>
<tr>
<td>Longest dry spell</td>
</tr>
<tr>
<td>Average seasonal temp (deg. celsius)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Number of Districts</td>
</tr>
</tbody>
</table>

*Note: Mean coefficients reported, with standard deviation in parenthesis*
Table 3: Quantile and Linear Panel Regression with Fixed-effects for Wheat, Using Seasonal Rainfall Data.

<table>
<thead>
<tr>
<th></th>
<th>tau=0.25</th>
<th>tau=0.5</th>
<th>tau=0.75</th>
<th>tau=0.9</th>
<th>Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDD</td>
<td>2.05E-03</td>
<td>1.94E-03</td>
<td>1.51E-03</td>
<td>1.41E-03</td>
<td>2.04E-03</td>
</tr>
<tr>
<td></td>
<td>{4.106}</td>
<td>{5.08}</td>
<td>{4.86}</td>
<td>{4.634}</td>
<td>{6.626}</td>
</tr>
<tr>
<td></td>
<td>3.862</td>
<td>4.802</td>
<td>4.633</td>
<td>4.621</td>
<td></td>
</tr>
<tr>
<td>GDD sqf</td>
<td>-4.79E-07</td>
<td>-4.21E-07</td>
<td>-3.20E-07</td>
<td>-3.32E-07</td>
<td>-4.35E-07</td>
</tr>
<tr>
<td></td>
<td>{-5.133}</td>
<td>{-5.941}</td>
<td>{-5.367}</td>
<td>{-4.493}</td>
<td>{-7.98}</td>
</tr>
<tr>
<td></td>
<td>-4.811</td>
<td>-5.632</td>
<td>-5.221</td>
<td>-5.303</td>
<td></td>
</tr>
<tr>
<td>Seasonal Rain</td>
<td>5.03E-04</td>
<td>2.91E-04</td>
<td>6.41E-04</td>
<td>3.69E-04</td>
<td>4.95E-04</td>
</tr>
<tr>
<td></td>
<td>{2.19}</td>
<td>{1.88}</td>
<td>{3.885}</td>
<td>{1.139}</td>
<td>{3.3}</td>
</tr>
<tr>
<td></td>
<td>2.446</td>
<td>1.69</td>
<td>3.233</td>
<td>1.576</td>
<td></td>
</tr>
<tr>
<td>Seasonal rain sqf</td>
<td>-1.26E-06</td>
<td>-9.97E-07</td>
<td>-1.74E-06</td>
<td>-4.09E-07</td>
<td>-1.28E-06</td>
</tr>
<tr>
<td></td>
<td>{-1.772}</td>
<td>{-2.483}</td>
<td>{-3.168}</td>
<td>{-0.586}</td>
<td>{-3.15}</td>
</tr>
<tr>
<td></td>
<td>-2.032</td>
<td>-2.143</td>
<td>-2.698</td>
<td>-0.852</td>
<td></td>
</tr>
<tr>
<td>hgd</td>
<td>7.58E-04</td>
<td>4.10E-04</td>
<td>5.45E-05</td>
<td>1.27E-03</td>
<td>3.79E-04</td>
</tr>
<tr>
<td></td>
<td>{0.827}</td>
<td>{0.5346}</td>
<td>{0.07862}</td>
<td>{1.149}</td>
<td>{0.654}</td>
</tr>
<tr>
<td></td>
<td>-0.7863</td>
<td>-0.5126</td>
<td>0.07665</td>
<td>1.027</td>
<td></td>
</tr>
<tr>
<td>Irr_ratio</td>
<td>1.00E+00</td>
<td>9.75E-01</td>
<td>9.27E-01</td>
<td>7.75E-01</td>
<td>7.95E-01</td>
</tr>
<tr>
<td></td>
<td>{14.83}</td>
<td>{15.34}</td>
<td>{12.88}</td>
<td>{8.697}</td>
<td>{22.92}</td>
</tr>
<tr>
<td></td>
<td>15.25</td>
<td>15.72</td>
<td>14.59</td>
<td>9.705</td>
<td></td>
</tr>
</tbody>
</table>

Note: All regressions include region-specific cubic time trends and district fixed-effects whose coefficients are not reported. T-statistics for the test H0: = 0 based on the asymptotic variance matrix (\{\}) and time bootstrap for panel data (\[\]) are reported.

Table 4: Quantile and Linear Panel Regression with Fixed-effect for Wheat, Using Orthogonal Polynomials (of degree 4) for GDD.

<table>
<thead>
<tr>
<th></th>
<th>tau=0.25</th>
<th>tau=0.5</th>
<th>tau=0.75</th>
<th>tau=0.9</th>
<th>Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal Rain</td>
<td>5.05E-04</td>
<td>2.38E-04</td>
<td>5.67E-04</td>
<td>3.90E-04</td>
<td>4.84E-04</td>
</tr>
<tr>
<td></td>
<td>{2.232}</td>
<td>{1.544}</td>
<td>{3.183}</td>
<td>{1.874}</td>
<td>{3.23}</td>
</tr>
<tr>
<td></td>
<td>{2.519}</td>
<td>{1.42}</td>
<td>{2.9}</td>
<td>{2.124}</td>
<td></td>
</tr>
<tr>
<td>Seasonal rain sqf</td>
<td>-1.29E-06</td>
<td>-8.63E-07</td>
<td>-1.52E-06</td>
<td>-4.82E-07</td>
<td>-1.26E-06</td>
</tr>
<tr>
<td></td>
<td>{-1.851}</td>
<td>{-2.19}</td>
<td>{-2.686}</td>
<td>{-0.6702}</td>
<td>{-3.09}</td>
</tr>
<tr>
<td></td>
<td>2.118</td>
<td>1.902</td>
<td>2.447</td>
<td>0.7924</td>
<td></td>
</tr>
<tr>
<td>hgd</td>
<td>-8.45E-04</td>
<td>-6.62E-04</td>
<td>-6.85E-05</td>
<td>1.28E-03</td>
<td>3.13E-04</td>
</tr>
<tr>
<td></td>
<td>{-0.9137}</td>
<td>{-0.6013}</td>
<td>{-0.09655}</td>
<td>{1.163}</td>
<td>{0.53}</td>
</tr>
<tr>
<td></td>
<td>-0.8716</td>
<td>-0.597</td>
<td>-0.0917</td>
<td>1.647</td>
<td></td>
</tr>
<tr>
<td>Irr_ratio</td>
<td>1.01E+00</td>
<td>9.71E-01</td>
<td>9.22E-01</td>
<td>7.76E-01</td>
<td>7.95E-01</td>
</tr>
<tr>
<td></td>
<td>{14.99}</td>
<td>{15.33}</td>
<td>{12.32}</td>
<td>{8.58}</td>
<td>{22.92}</td>
</tr>
<tr>
<td></td>
<td>15.32</td>
<td>15.75</td>
<td>14.26</td>
<td>9.307</td>
<td></td>
</tr>
</tbody>
</table>

Note: Same as in Table 3.
Table 5: Quantile and Linear Panel Regression with Fixed-effects for Rice, Using Seasonal Rainfall Data.

<table>
<thead>
<tr>
<th>Coefs</th>
<th>tau=0.25</th>
<th>tau=0.5</th>
<th>tau=0.75</th>
<th>tau=0.9</th>
<th>linear panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>gdd</td>
<td>-2.94e-03</td>
<td>-2.54e-03</td>
<td>-2.85e-03</td>
<td>-2.77e-03</td>
<td>-2.73e-03</td>
</tr>
<tr>
<td>gdd sqn</td>
<td>5.59e-07</td>
<td>4.93e-07</td>
<td>5.19e-07</td>
<td>4.90e-07</td>
<td>4.36e-07</td>
</tr>
<tr>
<td></td>
<td>[2.2]</td>
<td>[3.3]</td>
<td>[3.2]</td>
<td>[2.4]</td>
<td>[2.1]</td>
</tr>
<tr>
<td>seasonal rain</td>
<td>3.03e-04</td>
<td>2.47e-04</td>
<td>1.57e-04</td>
<td>1.03e-04</td>
<td>3.94e-04</td>
</tr>
<tr>
<td></td>
<td>[5.7]</td>
<td>[6]</td>
<td>[3.8]</td>
<td>[2.3]</td>
<td>[8]</td>
</tr>
<tr>
<td>rain sqn</td>
<td>-6.80e-08</td>
<td>-5.35e-08</td>
<td>-3.65e-08</td>
<td>-2.93e-08</td>
<td>-9.40e-08</td>
</tr>
<tr>
<td>onset day</td>
<td>6.09e-04</td>
<td>1.99e-04</td>
<td>-6.41e-05</td>
<td>-2.72e-04</td>
<td>7.22e-04</td>
</tr>
<tr>
<td></td>
<td>[2.6]</td>
<td>[1.1]</td>
<td>[-0.35]</td>
<td>[-1.5]</td>
<td>[2.9]</td>
</tr>
<tr>
<td>dry spells</td>
<td>-3.41e-03</td>
<td>-2.47e-03</td>
<td>-1.61e-03</td>
<td>-1.81e-03</td>
<td>-4.22e-03</td>
</tr>
</tbody>
</table>

Note: Same as in Table 3.
Table 6: Summary Statistics of Percentage Change in Yield of Rice and Wheat under Different Scenarios at the District Level.

<table>
<thead>
<tr>
<th></th>
<th>Rice</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tau = 0.25</td>
<td>0.399</td>
<td>[0.382,0.412]</td>
</tr>
<tr>
<td>tau = 0.50</td>
<td>0.542</td>
<td>[0.526,0.555]</td>
</tr>
<tr>
<td>tau = 0.75</td>
<td>-0.88</td>
<td>[-0.899,-0.858]</td>
</tr>
<tr>
<td>tau = 0.90</td>
<td>-1.67</td>
<td>[-1.68,-1.65]</td>
</tr>
<tr>
<td>Scenario 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tau = 0.25</td>
<td>0.205</td>
<td>[0.168,0.24]</td>
</tr>
<tr>
<td>tau = 0.50</td>
<td>0.688</td>
<td>[0.659,0.719]</td>
</tr>
<tr>
<td>tau = 0.75</td>
<td>-1.71</td>
<td>[-1.78,-1.69]</td>
</tr>
<tr>
<td>tau = 0.90</td>
<td>-2.99</td>
<td>[-3.03,-2.96]</td>
</tr>
</tbody>
</table>

Note: For rice, Scenario 1 and 2 correspond to uniform increase in daily temperatures over the growing season by 1 and 2°C, along with a 10% increase in seasonal rainfall, while for wheat, they correspond to increases in daily temperatures of 0.5 and 1°C respectively. Median impacts are reported along with (time) bootstrapped confidence intervals (in []).
Table 7: Summary Statistics of Percentage Change in Yield of Wheat, for Different Regions of India, under Different Scenarios, at the District Level.

<table>
<thead>
<tr>
<th></th>
<th>South</th>
<th>North</th>
<th>Central</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>min</td>
<td>max</td>
<td>median</td>
</tr>
<tr>
<td>$\tau = 0.25$</td>
<td>-10.6</td>
<td>-11.5</td>
<td>-0.8788</td>
<td>-3.16</td>
</tr>
<tr>
<td></td>
<td>-10.6, -10.6</td>
<td>-3.29, -5.06</td>
<td>-6.14, -5.91</td>
<td>-6.49, -5.98</td>
</tr>
<tr>
<td>$\tau = 0.5$</td>
<td>-8.1</td>
<td>-8.89</td>
<td>0.3793</td>
<td>-1.59</td>
</tr>
<tr>
<td></td>
<td>-8.12, -8.68</td>
<td>-1.68, -1.5</td>
<td>-4.24, -4.03</td>
<td>-4.52, -4.04</td>
</tr>
<tr>
<td>$\tau = 0.75$</td>
<td>-5.12</td>
<td>-5.959</td>
<td>0.8055</td>
<td>-0.964</td>
</tr>
<tr>
<td></td>
<td>-5.2, -5.04</td>
<td>-1.04, -0.882</td>
<td>-3.24, -3.06</td>
<td>-3.36, -2.93</td>
</tr>
<tr>
<td>$\tau = 0.9$</td>
<td>-5.43</td>
<td>-7.099</td>
<td>-0.5471</td>
<td>-1.89</td>
</tr>
<tr>
<td></td>
<td>-5.52, -5.34</td>
<td>-1.93, -1.83</td>
<td>-3.75, -3.61</td>
<td>-3.35, -3.1</td>
</tr>
<tr>
<td>Scenario 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau = 0.5$</td>
<td>-16</td>
<td>-17.12</td>
<td>-0.05675</td>
<td>-4.16</td>
</tr>
<tr>
<td></td>
<td>-16.1, -16</td>
<td>-4.34, -2.96</td>
<td>-9.43, -8.92</td>
<td>-9.82, -9</td>
</tr>
<tr>
<td>$\tau = 0.75$</td>
<td>-9.63</td>
<td>-11.74</td>
<td>0.8845</td>
<td>-2.72</td>
</tr>
<tr>
<td></td>
<td>-9.8, -9.49</td>
<td>-2.87, -2.5</td>
<td>-7.16, -6.83</td>
<td>-7.3, -6.69</td>
</tr>
<tr>
<td>$\tau = 0.9$</td>
<td>-10.5</td>
<td>-14.34</td>
<td>-1.635</td>
<td>-4.07</td>
</tr>
<tr>
<td></td>
<td>-10.7, -10.3</td>
<td>-4.22, -5.92</td>
<td>-7.74, -7.45</td>
<td>-6.74, -6.29</td>
</tr>
</tbody>
</table>

Note: Same as in Table 6.
Figures

Figure 1: Box Plots of Changes in Wheat Yield under Scenarios Indicated, for Various Quantiles.

(a) Scenario 1

(b) Scenario 2

Note: Scenarios same as in Table 6.
Figure 2: Box Plots of Changes in Wheat Yield for Different Regions under Scenario 1 for Different Quantiles.

(a) $25^{th}$ percentile

(b) Median

(c) $75^{th}$ percentile

(d) $90^{th}$ percentile

Note: Scenarios same as in Table 6.
Figure 3: Box Plots of Changes in Rice Yield under Scenarios Indicated, for Various Quantiles.

(a) Scenario 1

(b) Scenario 2

Note: Scenarios same as in Table 6.
Figure 4: Box Plots of Changes in Wheat Yield for Different Regions under Scenario 2 for Different Quantiles.

(a) 25th percentile

(b) Median

(c) 75th percentile

(d) 90th percentile

Note: Scenarios same as in Table 6.
Figure 5: Non parametric (Penalized Spline) Estimation of the Effect of Growing Degree Days on Rice Yields at Different Quantiles.
Figure 6: Non-parametric (Penalized Spline) Estimation of the Effect of Growing Degree Days on Wheat yields (at the Median).

Figure 7: Box Plots of Changes in Rice and Wheat Yield under Scenarios Indicated, from the Linear Fixed Effects Panel Data Model.
**MSE Monographs**

* Monograph 9/2010
  Feasibility of Incentive Based Environmental Instruments in State and Central Taxation Regimes
  D.K.Srivastava and C. Bhujanga Rao

* Monograph 10/2010
  Economic Instruments for Environmental Management: International Best Practices and Applicability to India
  D.K.Srivastava, Dr K.S. Kavikumar, and C.Bhujanga Rao
  With inputs from Dr Bodhisattva Sengupta, Dr Brijesh C. Purohit, Ms Asha Mariam Abraham, Ms Ishwarya

* Monograph 11/2011
  Taxation of Goods and Services in India
  D.K. Srivastava and C. Bhujanga Rao

* Monograph 12/2011
  Coping with Pollution: Eco Taxes in a GST Regime
  D.K. Srivastava, K.S. Kavi Kumar and C. Bhujanga Rao, with inputs from Brijesh C. Purohit and Bodhisattva Sengupta

* Monograph 13/2011
  Recent Bouts of Inflation in India: Policy Paralysis?
  T.N. Srinivasan

* Monograph 14/2011
  Impact of Fiscal Instruments in Environmental Management through a Simulation Model: Case Study of India
  D.K. Srivastava and K.S. Kavi Kumar, with inputs from Subham Kailthya and Ishwarya Balasubramaniam

* Monograph 15/2012
  Environmental Subsidies in India: Role and Reforms
  D.K. Srivastava, Rita Pandey and C.Bhujanga Rao with inputs from Bodhisattva Sengupta

* Monograph 16/2012
  Integrating Eco-Taxes in the Goods and Services Tax Regime in India
  D.K. Srivastava and K.S. Kavi Kumar

* Monograph 17/2012
  Monitorable Indicators and Performance: Tamil Nadu
  K. R. Shanmugam

* Monograph 18/2012
  Performance of Flagship Programmes in Tamil Nadu
  K. R. Shanmugam, Swarna S Vepa and Savita Bhat

* Monograph 19/2012
  State Finances of Tamil Nadu: Review and Projections A Study for the Fourth State Finance Commission of Tamil Nadu
  D.K. Srivastava and K. R. Shanmugam

* Monograph 20/2012
  Globalization and India’s Fiscal Federalism Finance Commission’s Adaptation to New Challenges
  Baldev Raj Nayar
**MSE Working Papers**

**Recent Issues**

  Dependence of States on Central Transfers: Aggregate and State-Wise Analysis  
  D K Srivastava and C Bhajanaga Rao

* Working Paper 60/2011  
  Discount Rate for Health Benefits and the Value of Life in India  
  K.R. Shanmugam

  Evidence on Changes in Time Varying Volatility around Bonus and Rights Issue Announcements  
  Madhuri Malhotra, M. Thenmozhi and Arun Kumar Gopalaswamy (forthcoming)

  Patterns of Labour Market Insecurity in Rural India: A Multidimensional and Multivariate Approach  
  Padmini Desikachar and Brinda Viswanathan

  Determinants of Child Morbidity and Factors Governing Utilisation of Child Health Care: Evidence from Rural India  
  Anindita Chakrabarti

* Working Paper 64/2012  
  Arsenic Contamination in Water: A Conceptual Framework of Policy Options  
  Zareena Begum I

* Working Paper 65/2012  
  Learning and Capability Acquisition: A Case Study of the Indian Automobile Industry  
  Madhuri Saripalle

* Working Paper 66/2012  
  Poverty, Human Development and Health Financing in India  
  Brijesh C. Parohit

  Corporate Governance and Product Market Competition  
  Ekta Selarka

* Working Paper 68/2012  
  Basel I and Basel II Compliance: Issues for Banks in India  
  Srejita Banerjee

* Working papers are downloadable from MSE website http://www.mse.ac.in

$ Restricted circulation

---

**WORKING PAPER 69/2012**

**The Distributional Impacts of Climate Change on Indian Agriculture: A Quantile Regression Approach**

Chandra Kiran B Krishnamurthy

---

MADRAS SCHOOL OF ECONOMICS  
Gandhi Mandapam Road  
Chennai 600 025  
India

May 2012