



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.



***Selected Presentation at the 2020 Agricultural &
Applied Economics Association Annual Meeting,
Kansas City, Missouri, July 26-28***

Copyright 2020 by authors. All rights reserved.

*Readers may make verbatim copies of this document for non-commercial purposes by any means,
provided that this copyright notice appears on all such copies.*

Agricultural input use and index insurance adoption: Concept and evidence

Gaurav Arora

Dept. of Social Sciences and Humanities (Economics Division)
Indraprastha Institute of Information Technology, Delhi
New Delhi, India
Email: gaurav@iiitd.ac.in

Sandip Kumar Agarwal

Department of Economic Sciences
Indian Institute of Science, Education and Research, Bhopal
Madhya Pradesh, India
Email: sandipa@iiserb.ac.in

This research is part of a project supported by the South Asian Network of Development and Environmental Economics (SANDEE) and International Center for Integrated Mountain Development (ICIMOD), award number RP6/SANDEE. The views in this paper are attributable solely to the authors.

Copyright 2020 by Gaurav Arora and Sandip Agarwal. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract

We propose to open new debate by positing index insurance as part of farm risk management portfolio in low-income countries whose demand might be affected by crop inputs having yield risk-characterizing property. We argue that estimating input-conditioned farm yields is a pre-requisite to develop an understanding of how input demand might interact with index insurance adoption on small farms in low-income countries. A new *BetaIV* framework is employed to estimate the conditional wheat yield distribution for smallholder farmlands (median plot size of 2.4 acres) across twelve semi-arid villages of India. We find that wheat yields are positively-skewed on Indian farms, and an increase in nitrogen and irrigation inputs decreases this skew in the first-order stochastic dominance sense. The impact of nitrogen on yield risk is found to be somewhat greater than that of irrigation. Further, rainfed wheat production is found to be on-average riskier (i.e., fatter left tails) than irrigated wheat. An expected utility framework is employed to simulate the own- and cross-price effects for nitrogen, irrigation and index insurance for rainfed and irrigated wheat production for Papda village of Madhya Pradesh state in India. Our findings suggest that the demand for urea and irrigation increases and that of index insurance decreases with increase in risk aversion. In addition, higher irrigation cost lead to higher demand for urea and index insurance for moderately-to-highly risk averse farmers.

Keywords: yield risk; index insurance; input subsidies; small farms; agricultural finance.

JEL Classification: H53; N55; Q12; Q18.

1. Introduction

“Yet while the debate around the merits of index insurance remains unsettled in the minds of researchers and practitioners alike, funding for index insurance as a social protection and development tool continues to expand across the globe.”

Jensen and Barrett (2017)

Agricultural index insurance is not only considered an important safety net for production risks among the policymakers but also as a developmental tool to help mitigate poverty, build investment capacity of marginal farmers and enable higher productivity in the long-run. It is also an alternative mechanism (Alderman and Haque 2007) to conventional farm insurance products made available to growers in high-income countries against farm-level production risks. Index insurance typically serves a large (pre-designated) area and triggers a uniform indemnity payment whenever a yield index or weather-based index signals area-level losses greater than a threshold. Growers also pay same premium rates to buy index insurance coverage. The premiums and indemnities are crop-specific for yield index insurance but not for weather-based index insurance. The area-based index insurance is advantageous in developing countries where collecting actuarial data and monitoring farm losses is difficult (Jansen and Barrett 2007) due to small farms with dispersed property rights. Such information gaps would ultimately lead to inefficiencies in the form of moral hazard and adverse selection in conventional farm insurance products. In case of index insurance, however, the premium rates and indemnity payments are exogenous to individual farm characteristics or producer behavior, and hence do not suffer from those issues (Hirshleifer and Riley 1992).

But the very feature of index-based indemnities that circumvents the endogeneity between insurance outcomes and cropping behavior of farmers also deters its adoption. These barriers are fundamentally rooted in the imperfect correlation between area-level index and field-level production outcomes (Jensen et al. 2016). For example, a farmer might experience major yield loss but the index might not trigger an indemnity in a particular year due to moderate area-level average losses. Conversely also the index might trigger an indemnity during a season when the farmer did not incur losses. The likelihood of such contractual non-performance events is termed as basis risk of the index insurance and has attracted much attention within the economics literature. Clark (2016) shows that the basis risk might lead to depressed demand for index insurance among risk-averse producers even when the premium rates are set at the actuarially favorable levels. Moreover, the fact that index insurance does not mitigate farm risks with certainty can lead to poverty traps and prove detrimental to the economic and social wellbeing of marginal farmers in low-income communities (Gine' et al. 2008, Janzen et al. 2013, Jensen et al. 2016). It has been shown that behavioral traits like loss aversion might further stress index insurance demand (Elabed and Carter 2015). In addition, the relation between farm risks and index-based area-level risks is a technical matter that involves concepts like joint probability, which is likely to present a major impediment in less literate communities. Education and trust in institutions are important pre-requisites for index insurance adoption in developing countries (Cai et al. 2011)¹.

¹ In India, only 5% of insurance holders bought them voluntarily while the majority 95% access it via compulsory bundling with agricultural credit. Our September 2018 field interviews with the growers of three central Indian villages of Madhya Pradesh revealed lack of awareness about crop insurance, even among the credit taking farmers. We too found illiteracy, poverty and the lack of trust in formal institutions are major impediments to insurance non-adoption in India, which is in line with the emerging research findings globally ().

In this study we propose to open new debate by positing index insurance as part of farm technology and management, and then analyzing the interaction(s) between agricultural input applications and index insurance adoption. We draw from the established literature on farm inputs that characterize yield risks, more particularly the probability of low yields, and ask whether some inputs might act as gross substitutes of index insurance. This query is relevant for agricultural policy in the developing countries because both index insurance and farm inputs (e.g., fertilizers and irrigation) are highly subsidized. Where past literature has explored design-related issues to explain index insurance non-adoption, ours is the first study to explore the interactions between index insurance demand and the demand for farm inputs.

Agricultural economists have found abundant evidence that farm inputs like fertilizers and irrigation can affect the mass under left tail of the conditional yield distribution. (Horowitz and Lichtenberg 1993; Quiggin et al. 1993; Smith and Goodwin 1996; Babcock and Hennessy 1996). There are extensive and long-drawn debates about the how input application decisions would be altered in the presence of commercial crop insurance in developed countries. Babcock and Hennessy used agronomic field experiments to conclude that whenever nitrogen is a risk-reducing (risk-increasing) input, i.e., higher nitrogen application decreases (increases) the density of low yields, then insurance will act its substitute (complement) for farm production. However, no such analysis exists for the case of index insurance products with micro-level evidence for smallholder farms.

Linearly extending the existing lessons from commercial insurance adoption to index insurance is not trivial due to many reasons. First, it is possible that the basis risk, which fundamentally differentiates index insurance from commercial insurance, might depend on the amount of risk characterizing input applied. Let us take an example to explain this. Consider a

status-quo scenario for where the basis risk is zero, i.e., farm-level yields are perfectly correlated with those in the area where index insurance is made available. If one among the area's farmers marginally increases the level of nitrogen application it will alter her farm's yield density (as per Babcock and Hennessy 1996). Consequently, this farm's yields will no longer remain perfectly correlated with the area-level yields and this will introduce a non-zero basis risk for the farmer thereby reducing the utility of index insurance. To the best of our knowledge such issues remain unexplored in the context of index insurance demand. Second, weak institutional support in the developing economies often results in lack of awareness coupled with the delays in claims settlement (Mukherjee and Pal 2017). Moreover, subsidies would promote fertilizer application as a means for farm risk management and contribute to the inertia in insurance adoption. Third, crop technology on the small farms might exhibit different risk premiums than on large farms in developed countries (Ramaswamy 1992, Proposition 1), and hence it is important to investigate the issue with the data that represents smallholder farmer decisions.

Our study is also important from the standpoint of agricultural development policy in low income countries. Several studies have analyzed the impact of commercial yield and revenue insurance subsidy on optimal input factor demand in the high-income societies (Horowitz and Lichtenberg 1993; Quiggin et al. 1993; Babcock and Hennessy 1996; Smith and Goodwin 1996; Wu 1999). However, such analyses are lacking for the index insurance programs in the developing countries where the suitable query is different, i.e., might input subsidies cause low index insurance adoption. Insurance adoption by farmers in India farmers has been abysmal in spite of the highly subsidized insurance premiums. Could it be that the rupees invested in providing input subsidy might dampen crop insurance adoption, thereby countering insurance subsidy investment and leading not only to economic inefficiency but possibly also to chemical

over-use and environmental pollution (Wu 1998). An important policy outcome of this research will be the recommendations for devising an optimal subsidy portfolio for farm inputs like fertilizers, irrigation and insurance premiums that lead to a balance between crop insurance participation and fertilizer use. Although this research focuses on fertilizer, irrigation and insurance subsidy, the proposed framework will allow investigation of interactions between multiple risk management policies.

This paper is organized as follows. We begin with a formal investigation the interaction between index insurance adoption and the demand for a risk-reducing input to motivate the significance of yield density estimation on smallholder farms. This is followed by a new econometric framework for estimating input-conditioned crop yield density. This framework will employ instrumental variable regression with the beta regression to estimate yield density conditional on endogenous input application decision using observational survey data. We will then model the impact of input application on basis risk (i.e., index insurance non-performance scenarios) and simulate the cross-price effects between optimal demand for fertilizer, irrigation and index insurance for smallholder farmers in India's semi-arid villages with different risk aversion and basis risk combinations. We will conclude with a summary and brief discussion.

2. Concept: Index insurance outcomes and input-conditioned yields

Babcock and Hennessy (1996) utilize an expected utility framework to formalize the linkage between optimal demand of commercial insurance and nitrogen fertilizer. They show that nitrogen would act as a substitute or complement of farm-level insurance depending upon the input-conditioned crop yields, i.e., the risk characterizing property of nitrogen. Below we investigate the matter for the case of index yield insurance.

At the beginning of each growing season a crop-specific index yield insurance is made available for a pre-designated area, say A , which is usually a combination of proximate villages, or other administrative units like taluks or a district. A representative farmer, say i , can insure her farm's yields, say y_i , against a yield index, \bar{y}_A (generally linked to average historical yields in A , and a fixed threshold yield, y_T . Whenever the index is lower than the threshold, $\bar{y}_A < y_T$, then the insurance holder is paid an indemnity equal to $P_I (y_T - \bar{y}_A)$, where P_I is the indemnity price. And whenever $\bar{y}_A \geq y_T$ then no indemnity is paid the farmers of that area. Index insurance indemnity can be written as $I_A = \max \{P_I (y_T - \bar{y}_A), 0\}$. In the farmer's decision to buy index insurance the assessment of farm-level yields, and more particularly crop failure and/or the probability of low yields, will play a central role. This is because farmer i will incur a loss whenever $y_i < y_T$ and so the utility of index insurance is in hedging against risk of y_i falling below the threshold y_T . To characterize the yield risk let us designate the conditional density of farm yields as $f_i(y | x_i)$, where x is the amount of input applied and the subscript i represents the farm-specific technology, soils, farmer demographics, etc. The conditional yield density function will provide a relationship between the input levels, x , and probability of low yields (i.e., the probability that $y_i < y_T$), and hence by extension the utility of risk-hedging via index insurance. Another matter at issue is that even in an event of loss, i.e., $y_i < y_T$, the insured farmer does not receive the indemnity payments with certainty, i.e., when $\bar{y}_A < y_T$, which we characterize by an area-level density, $g_A(y)$. We assume that farm- and area-level yields lie between $[a, b]$ where a and b are non-negative real numbers. Four outcome states emerge for the farmer after the growing season:

State 1: Farmer incurs a loss and obtains an indemnity, i.e., $y_i < y_T$ and $\bar{y}_A < y_T$.

State 2: Farmer incurs a loss but does not obtain indemnity, i.e., $y_i < y_T$ and $\bar{y}_A \geq y_T$.

State 3: Farmer does not incur loss but obtains an indemnity, i.e., $y_i \geq y_T$ and $\bar{y}_A < y_T$.

State 4: Farmer does not incur loss and does not obtain indemnity, i.e., $y_i \geq y_T$ and $\bar{y}_A \geq y_T$.

We define the farm-level net returns in each outcome state by denoting the per unit input price P_x and assuming per unit output price indemnity price to be equal to one, i.e., $P_y = P_I = 1$

(without loss of generality). Consider the case that farmer i purchased index insurance at a fixed cost C_I . Then, her profits in each outcome state is:

State 1: $\rho_1 = y_i - P_x x - C_I + (y_T - \bar{y}_A)$; $y_i < y_T$.

State 2: $\pi_2 = y_i - P_x x - C_I$; $y_i < y_T$.

State 3: $\rho_3 = y_i - P_x x - C_I + (y_T - \bar{y}_A)$; $y_i \geq y_T$.

State 4: $\pi_4 = y_i - P_x x - C_I$; $y_i \geq y_T$.

Therefore, the expected utility theory-based input choice for a risk-averse farmer is written as

$$\begin{aligned} \max_x EU = & \int_{y_i=a}^{y_T} \int_{\bar{y}_A=a}^{y_T} U(\rho_1) g_A(\bar{y}_A) f_i(y_i | x) d\bar{y}_A dy_i + \int_{y_i=a}^{y_T} \int_{\bar{y}_A=y_T}^b U(\rho_2) g_A(\bar{y}_A) f_i(y_i | x) d\bar{y}_A dy_i \\ & + \int_{y_i=y_T}^b \int_{\bar{y}_A=a}^{y_T} U(\rho_3) g_A(\bar{y}_A) f_i(y_i | x) d\bar{y}_A dy_i + \int_{y_i=y_T}^b \int_{\bar{y}_A=y_T}^b U(\rho_4) g_A(\bar{y}_A) f_i(y_i | x) d\bar{y}_A dy_i \end{aligned}$$

, where $U(\cdot)$ is a concave function exhibiting a risk-averse preference structure². The first-order condition for optimal level of input is given as

$$\frac{\partial EU}{\partial x} = \sum_{s=1}^4 \left\{ E_{f',s} \left[E_{g,s} \left[U(\pi_s) \right] \right] - P_x E_{f,s} \left[E_{g,s} \left[U'(\pi_s) \right] \right] \right\} = 0,$$

where s represents the outcome states 1-4 as described earlier,

² Note that we have assumed independence between farm-level and area-level yields only for the simplicity of exposition. Later in the paper we will relax this assumption.

$$E_{f',s} \left[E_{g,s} \left[U(\rho_s) \right] \right] = \int \int_{y_i \in S, y_A \in S} U(\rho_s) g_A(y_A) \frac{\partial f_i(y_i | x)}{\partial x} dy_A dy_i,$$

$$E_{f,s} \left[E_{g,s} \left[U'(\rho) \right] \right] = \int \int_{y_i \in S, y_A \in S} U'(\rho_s) g_A(y_A) f_i(y_i | x) dy_A dy_i, \text{ and the subscripts on the expectation}$$

operators signify the respective density functions and integration bounds.

The first-order condition characterizes the optimal input demand in the presence of an index insurance, $x^*(P_x, C_I, P_y, P_I, y_T; a, b)$. The second order sufficient condition that ensures

maximum expected utility at x^* is given as $\partial^2 EU / \partial x^2 = \Delta_{xx} < 0$, which implies that

$$\frac{\partial^2 EU}{\partial x^2} = \sum_{s=1}^4 \left\{ E_{f',s} \left[E_{g,s} \left[U(\pi_s) \right] \right] - 2P_x E_{f',s} \left[E_{g,s} \left[U'(\pi_s) \right] \right] + P_x E_{f,s} \left[E_{g,s} \left[U''(\pi_s) \right] \right] \right\} < 0.$$

Now the critical query is how input demand interacts with the cost of insurance. Totally differentiating the first-order condition with everything held constant except the input level x^* and the price of indemnity insurance C_I , and upon some rearranging we get

$$\begin{aligned} \Delta_{xx} dx^* - \left(\sum_{s=1}^4 \left\{ E_{f',s} \left[E_{g,s} \left[U'(\pi_s) \right] \right] - P_x E_{f,s} \left[E_{g,s} \left[U''(\pi_s) \right] \right] \right\} \right) dC_I &= 0 \\ \Rightarrow \\ \frac{dx^*}{dC_I} = -\Delta_{xx}^{-1} \left(\sum_{s=1}^4 \left\{ E_{g,s} \left[P_x \int_{y_i \in S} U''(\pi_s) f(y|x) dy_i - \int_{y_i \in S} U'(\pi_s) \frac{\partial f(y|x)}{\partial x} dy_i \right] \right\} \right) \\ \Rightarrow \\ \frac{dx^*}{dC_I} = \underbrace{-\Delta_{xx}^{-1}}_{\substack{-ve \\ +ve}} \left(\sum_{s=1}^4 \left\{ E_{g,s} \left[\underbrace{\int_{y_i \in S} \frac{\partial U'(\pi_s)}{\partial x} f(y|x) dy_i}_{\substack{-ve \\ T1(-ve)}} + \underbrace{\int_{y_i \in S} U'(\pi_s) \frac{\partial f(y|x)}{\partial x} dy_i}_{\substack{+ve \\ T2(?)}} \right] \right\} \right). \end{aligned}$$

Clearly, the sign of $\frac{dx}{dC_I}$ is ambiguous depends on the sign of $\frac{\partial f(y|x)}{\partial x}$ at $x = x^*$. That is,

the query about how index insurance cost impacts farmer's input application decision rests critically on whether the farm yield density increases or decreases upon on a marginal increase in input levels in each outcome state. If $\frac{\partial f(y|x)}{\partial x} \leq 0$ at $x = x^*$ then $\frac{dx}{dC_I} > 0$, meaning that the

farmers will apply more quantities of the risk-reducing input upon increase in the cost of index insurance, implying that such an input is a substitute of the index insurance. However, in the case

when $\frac{\partial f(y|x)}{\partial x} > 0$ at $x = x^*$ then the sign of $\frac{dx}{dC_I}$ remains ambiguous and depends upon the

relative size of the first term (T1) and second term (T2) in each outcome state. Hence, we

conclude that one must estimate the conditional yield distribution in order to establish

complementarity/substitutability between a crop input and the index insurance product. Premium

rate setting for crop insurance contracts also requires an understanding of the shape of the crop

yield distribution (Atwood et al. 2003; Goodwin and Mahul 2004; Sherrick et al. 2004; Ozaki et

al. 2008). Heteroskedasticity in yields too plays an important role in insurance rate-setting (Harri

et al. 2011; Ker and Tolhurst 2019).

3. Econometric estimation of input-conditioned yield distribution

As shown above, measuring the marginal impact of a risk characterizing input on yield density is

important to analyze the interaction between optimal farm input decisions and index insurance

adoption. Here, we propose a beta regression-based framework to estimate yield density using

multi-year surveys on plot-level cropping decisions. To circumvent the endogeneity between

farmer's input application decisions and yields we adapt Terza et al. (2008) to implement

instrumental variable (IV) estimation within the beta regression model using a two-stage residual

inclusion (2SRI) technique. We are not aware of any previous empirical applications that integrated the beta regression framework with the non-linear IV models.

Conditional crop yield density estimation has garnered a lot of attention among agricultural economists for it is one of the fundamental components in insurance rate setting and farm planning under uncertainty. The statistical properties of yield distribution were tested empirically, specifically for skewness (degree of lopsidedness in frequency distribution) as it directly measures downside-risk and provides an inference on the (non-)normality assumption for yield distribution (Day 1965; Buccola 1986; Gallagher 1986, 1987; Nelson and Preckel 1989; Taylor 1990; Swinton and King 1991; Moss and Shonkwiler 1993; Ramirez 1997; Just and Weninger 1999; Atwood et al. 2003; Ramirez et al. 2003; Popp et al. 2005; Ardian et al. 2009; Hennessy 2009a, 2009b, 2011; Du et al. 2012). Existing evidence on the shape of yield density (based mainly on yields data from large farms in high-income countries) ranges from positively-skewed yields to normally distributed yields to negatively-skewed yields. Different findings have been attributed to the use of farm-level versus aggregate³ data, experimental versus observational data⁴, and the (mis)specification of conditional yield model. Hennessy and his co-authors have based the matter on fundamental statistical laws to infer that yields are non-normal and that irrigation would increase yield skewness while nitrogen would decrease it.

Ours is a highly differentiated empirical setting as far as the existing evidence on conditional crop yield distributions is concerned. Specifically, we estimate input-conditioned yield distributions for smallholder farmers where cropping operations are usually restricted to one- or

³ Both geographical and temporal aggregation issues have been discussed in the literature.

⁴ Experimental data are acquired by agronomists from agricultural extension fields, while observational data are either primary field surveys or aggregated administrative yields data that provide records on actual farming decisions.

two-acre plots and the farmers face extreme resource constraints⁵, restrictive social norms⁶, and weak institutions. From that standpoint this exercise, although econometric in nature, deals with decision-making under uncertainty that might fundamentally differ from that on large farms in high-income countries possibly including different risk premiums and risk mitigating strategies. Therefore, there is no reason to expect that our empirical results will necessarily corroborate those in the existing literature. We now outline a basic framework of estimation, following by a brief description of data and the estimation results.

3.1 Empirical Framework: Beta density estimation with IV regressions (BetaIV)

We follow a re-parameterized version of Babcock and Hennessy (1996) and use a flexible beta distribution to model plot i 's yield density conditional on the $k \times 1$ input vector, \vec{x}_i , where k represents the number of elements⁷. In particular,

$$f_i(y|\vec{x}_i) = \frac{\Gamma(\omega(\vec{x}_i) + \tau(\vec{x}_i))}{\Gamma(\omega(\vec{x}_i))\Gamma(\tau(\vec{x}_i))} (y-a)^{(\omega(\vec{x}_i)-1)} (b-y)^{(\tau(\vec{x}_i)-1)}; \quad (1)$$

where $\omega(\vec{x}_i) > 0; \tau(\vec{x}_i) > 0; y \in [a, b]$.

The flexible beta distribution function in equation (1) is appropriate to model yields because it is they are skewed, possibly heteroskedastic through space and time, and best represent a bounded random variable (from crop failure to maximum attainable yields) (McDonald and Xu 1995; Smithson and Verkuilen 2006). The parameters ω and τ are shape parameters such that an

⁵ Resource constraints manifest in the form of natural resource constraints like poor soils; financial constraints like unequal credit access (Shukla and Arora 2019); and knowledge constraints, i.e., low literacy and lack of access to modern farming techniques.

⁶ Farmers' cropping decisions might not be rational and driven by conformity, in that they tend to copy their neighbors or the traditional choices of their household (Kanjilal and Arora 2019).

⁷ Note that not all inputs that impact yields are also risk characterizing. One can easily expand the given notation to designate the risk characterizing input separately, e.g., $x_{i,k} \in \vec{x}_i$ such that the k th element in the input vector impacts yield variance while others, say $\vec{x}_{i,k-}$, do not.

increase in ω keeping τ constant pulls the density towards the lower bound a , whereas an increase in τ keeping ω constant pulls the density towards the upper bound b . We set these shape parameters as a function of input x to signify that the yield density function is susceptible to changing upon an increase in x . Smithson and Verkuilen provided a linear transformation of such that y 's range squeezes between (not including) the bounds 0 and 1. That is,

$$\tilde{y} = N^{-1} \left[\frac{y-a}{b-a} (N-1) + 0.5 \right], \text{ where } \tilde{y} \in (0,1), \text{ and the density of } \tilde{y} \text{ can be written by replacing}$$

y by \tilde{y} in equation (1). Further, we can write the first and second moments of \tilde{y} as follows:

$$E(\tilde{y} | \bar{x}_i) = \mu(\bar{x}_i) = \frac{\omega(\bar{x}_i)}{\omega(\bar{x}_i) + \tau(\bar{x}_i)}, \text{ and}$$

$$V(\tilde{y} | \bar{x}_i) = \sigma^2(\bar{x}_i) = \frac{\omega(\bar{x}_i)\tau(\bar{x}_i)}{(\omega(\bar{x}_i) + \tau(\bar{x}_i))^2 (\omega(\bar{x}_i) + \tau(\bar{x}_i) + 1)} = \frac{\mu(\bar{x}_i)(1 - \mu(\bar{x}_i))}{\omega(\bar{x}_i) + \tau(\bar{x}_i) + 1}.$$

Further, following Smithson and Verkuilen we define, $\phi(\bar{x}_i) = \omega(\bar{x}_i) + \tau(\bar{x}_i)$ such that

$$\omega(\bar{x}_i) = \mu(\bar{x}_i)\phi(\bar{x}_i) \text{ and } \tau(\bar{x}_i) = (1 - \mu(\bar{x}_i))\phi(\bar{x}_i), \text{ where } \mu(\bar{x}_i) \text{ is termed as the location}$$

parameter and $\phi(\bar{x}_i)$ is termed as the dispersion parameter. For a given value of the location

parameter, an increase in $\phi(\bar{x}_i)$ implies an increase in $V(\tilde{y} | \bar{x}_i)$. Now, the two model parameters

are modeled separately as a location submodel and a precision submodel. In order to ensure that

expected yields lie between 0 and 1 we use a logistic link function to specify the location

submodel. That is,

$$\mu(\bar{x}_i) = \frac{\exp(\bar{x}_i\beta)}{1 + \exp(\bar{x}_i\beta)}, \tag{2}$$

where β is the $k \times 1$ coefficient vector for the location submodel. Similarly, in order to ensure that yield variance is strictly non-negative we specify the precision parameter using a log link (Smithson and Verkuilen 2006). That is,

$$\begin{aligned} \ln(\phi(\bar{x}_i)) &= -\bar{w}_i \eta, \text{ or} \\ \phi(\bar{x}_i) &= \exp(-\bar{w}_i \eta), \end{aligned} \quad (3)$$

where η is the coefficient vector for the precision submodel. Given the above link functions the log-likelihood function to estimate the parameter vectors β and η is given as (see Smithson and Verkuilen for details)

$$\begin{aligned} \max_{\beta, \eta} \ln L(\beta, \eta; y, X, W) &= \ln \Gamma \left[\frac{e^{X\beta - W\eta} + e^{-W\eta}}{1 + e^{X\beta}} \right] - \ln \Gamma \left[\frac{e^{X\beta - W\eta}}{1 + e^{X\beta}} \right] - \ln \Gamma \left[\frac{e^{-W\eta}}{1 + e^{X\beta}} \right] \\ &\quad + \left[\frac{e^{X\beta - W\eta}}{1 + e^{X\beta}} - 1 \right] \ln \tilde{y} + \left[\frac{e^{-W\eta}}{1 + e^{X\beta}} - 1 \right] \ln(1 - \tilde{y}) \end{aligned} \quad (4)$$

3.1.1 Tackling endogeneity between yields and factor inputs

An issue with estimating the causal relationship in equations (2-3) is that the factor inputs might be endogenous to crop yields. In that, on a plot with relatively higher yield potential (relative to other plots) a grower might allocate more resources, i.e., more fertilizer and irrigation in this study⁸, in expectation of higher productivity. In addition, such disparity in applying inputs across lands with different yield potential could be driven by unobserved factors like “effort” or “willingness”, which might also be related to the farm’s yield potential. Overall, while an increase in fertilizer application may lead to better yields, it is also possible that farmers apply relatively more fertilizer on fields with high yield potential. Similarly, if we enhance the

⁸ Table 2 reveals hereafter that pesticides were used only on ~5% wheat farms in our sample, so we chose to keep this input out of our analysis.

definition of yield potential to also incorporate the second moment of yields (and not just mean yield) then both the location and dispersion submodels can be argued to suffer from endogeneity.

Note that the beta regressions, as characterized in equations 2-4, are a non-linear estimation framework. Terza et al. (2008) show that a rote extension of the traditional IV estimation strategy (i.e., the two-stage least square or 2SLS model) is inconsistent in the case of non-linear models. Instead the consistent estimation strategy is called two-stage residual inclusion strategy or 2SRI for non-linear parent regression models like the beta regression. Here the first-stage IV estimation remains as in the case of traditional 2SLS, i.e., we estimate IV regressions using instruments for each endogenous input that (a) are sufficiently correlated with the input quantity variable during the growing-season but (b) are uncorrelated with the error term in the parent regressions (i.e., the location submodel and the dispersion submodel). In the second stage, however, first-stage residuals are included as additional regressors besides the endogenous regression variables. We now adapt the general IV regression strategy for non-linear models to the beta regression model, which to our best knowledge is first-of-its-kind application and we term it the *BetaIV* framework. We show this for the case of location submodel in equation (2), which will trivially extend also to the dispersion submodel in equation (3). The regression for the location submodel can be written as

$$\mu(\vec{x}_i) = \frac{\exp(N_i\beta_N + I_i\beta_I + \vec{x}_i^e\beta_o + \vec{x}_i^u\beta_u)}{1 + \exp(N_i\beta_N + I_i\beta_I + \vec{x}_i^e\beta_o + \vec{x}_i^u\beta_u)} + e_i, \quad (5)$$

$$E(e_i | N_i, I_i, \vec{x}_i^e, \vec{x}_i^u) = 0$$

where N_i and I_i represent nitrogen and irrigation application levels (details in the next section) that are deemed endogenous to mean yield on farm i . Endogeneity is specified using the

unobserved or latent variable vector \vec{x}_i^u that affects mean yield but is correlated with N_i and I_i .

Variable vector \vec{x}_i^e represents the exogenous control variables that impact mean yield on farm i .

In order to circumvent the endogeneity issue we estimate auxiliary IV regressions for factor inputs N_i and I_i . In particular, we specify and separately estimate the following IV regressions for the input vector $\vec{Z}_i \in \{N_i, I_i\}$

$$\vec{Z}_i = h(V_i\delta) + \kappa_i \quad (6)$$

A common issue arises in estimating (6) due to the frequent occurrence of zeros in irrigation and nitrogen application (i.e., corner solutions). The observed zeroes in equation (6) will be modelled by implementing a Tobit model for simultaneous nitrogen and irrigation application equations. Specifically, we estimate two components: the conditional (on the covariate vector V_i) likelihood of the occurrence of corner solution (zero input usage) and the conditional likelihood of the occurrence of positive, continuous input levels. That is,

$$\begin{aligned} \vec{Z}_i &= V_i\delta + \kappa_i \text{ where } \kappa_i | V_i \sim N(0, \sigma_\kappa^2) \\ \vec{Z}_i^* &= \max(\vec{Z}_i, 0) \end{aligned} \quad (7)$$

where \vec{Z}_i^* is taken to be the observed input application vector. The piece-wise density function for each input factor is given as (Greene 2003, Ch. 22)

$$g(\vec{Z}_i^* | V_i) = \begin{cases} 1 - \Phi(V_i\delta / \sigma_\kappa) & \text{if } \vec{Z}_i^* = 0 \\ \sigma_\kappa^{-1} \phi((\vec{Z}_i^* - V_i\delta) / \sigma_\kappa) & \text{if } \vec{Z}_i^* > 0 \end{cases} \quad (8)$$

The log-likelihood function is therefore given as

$$\ln L(\delta, \sigma_\kappa; \vec{Z}_i^*, V_i) = 1_{Z_i^*=0} \ln(1 - \Phi(V_i\delta / \sigma_\kappa)) + 1_{Z_i^*>0} \ln(\sigma_\kappa^{-1} \phi((\vec{Z}_i^* - V_i\delta) / \sigma_\kappa)) \quad (9)$$

Consistent estimators for the parameters δ and σ_κ can be retrieved from maximizing the likelihood function in equation (9). Now for the *BetaIV* framework we will recover the predicted residuals from equation (7) as $\hat{\kappa}_i = \vec{Z}_i - V_i \hat{\delta}_{tobit}$ and those will be then included in the parent regression model in place of the latent variables \vec{x}_i^u . That is, the BetaIV regression for the location submodel (or mean yields) is given as

$$\mu(\vec{x}_i) = \frac{\exp(N_i \beta_N + I_i \beta_I + \vec{x}_i^e \beta_o + \hat{\kappa}_i \beta_u)}{1 + \exp(N_i \beta_N + I_i \beta_I + \vec{x}_i^e \beta_o + \hat{\kappa}_i \beta_u)} + e_i. \quad (10)$$

The reason equation (10) will provide trustworthy estimate of mean yields is that we replaced the unobserved \vec{x}_i^u in the logistic regression (5) with its consistent estimates in the form of $\hat{\kappa}_i$ in (10), thereby removing the root cause of endogeneity. Similarly, the BetaIV regression for the dispersion submodel is given as

$$\phi(\vec{x}_i) = \exp\left[-(N_i \eta_N + I_i \eta_I + \vec{w}_i^e \eta_o + \hat{\kappa}_i \eta_u)\right] + e_i^\phi. \quad (11)$$

3.2 Data and Variables

We conducted one-on-one interviews with thirty farmers in three villages of the Madhya Pradesh state of Central India during October 2018 to learn about their on-farm production decisions, factors affecting yield risk and risk mitigation strategies, including insurance adoption. Pest infestation, drought, extreme wetness and stray animals were identified to be main factors of yield loss/risk. We found that while insurance non-adoption was common among the respondents, they actively applied fertilizers, irrigation and pesticides to mitigate adverse crop growth conditions based on weather and soils. For instance, farmers spray pesticides as an ad-hoc strategy in the events of pest infestation to facilitate plant growth through a few weeks of depressed pest population, after which if pests re-emerge they spray again. Similar scenarios

were modelled by Lichtenberg and Zilberman (1986). These observations provide an empirical grounding for our conjecture that inputs are used to mitigate farm risk and that large input subsidy might contribute to low insurance adoption rates in India.

We estimate wheat yield distribution conditional on nitrogen and irrigation inputs (equations 10-11) using repeated cross-sectional household surveys conducted under the Village Dynamics Studies in South Asia (VDSA) program of International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). The VDSA surveys were conducted over four waves during 1975-1984, 1989; 2001-2004; 2005-2008; and 2009-2014. The surveys were conducted in semi-arid villages across Indian states of Madhya Pradesh (M.P.), Gujarat (G.J.), Andhra Pradesh (A.P.), Karnataka (K.N.) and Maharashtra (M.H.). The survey sites were specifically chosen to lie in semi-arid tropical areas as specified by the regional agro-climatic conditions⁹. India's cropping calendar has three seasons in a year: wet season (Kharif/Monsoon, July to October), dry season (Rabi, November to March), and summer season (Zaid, April to June). Accordingly India's cropping years are notated as, e.g., 2004-05 including 2004 wet season to 2005 summer season.

Here we focus on wheat yields during the 2001-2014 period, i.e., the three latest survey waves of the VDSA. Wheat is grown only during the Rabi season. Table 1 presents the summary of villages in the VDSA survey where farmers grew wheat during our study period. A total of 193 households grew wheat on 373 unique plots across twelve villages in G.J., M.P., M.H. and

⁹ The 1975-1989 wave included ten villages from two states: Aurepalle and Dokur in A.P.; Boriya and Rampura in G.J.; Papda and Rampura Kalan in M.P.; and Kinkheda, Kanzara, Shirapur and Kalman in M.H. The 2001-2004 and 2005-2008 survey wave included only six of the above villages in A.P. and M.H. The 2009-2014 surveys included 32 villages from eight states in (in addition to the above villages): J.C. Agharam and Pamidipadu in A.P.; Inai, Susari, Arap and Bagakole in B.H.; Makhiyala, Karamadi Chingariya, Chatha and Babrol in G.J.; Dumariya, Durgapur, Dubaliya and Hesapiri in J.H., Markabbinahali, Kapanimbargi, Tharati and Belladamadugu in K.N.; Ainlatunga, Bilaikani, Chandrasekharapur and Sogar in O.D.

K.N. Table 2 shows that most households cultivate wheat once and ~80% up to four times in 14 years meaning that we have a repeated cross-sectional data (and not a balanced panel).

We match the plot-level crop yields data with inputs application data, farmer demographics, soil quality and time-invariant village-level rainfall during the 2001-14 period. Table 3 presents variable descriptions and corresponding summary statistics. Overall, the landholding sizes vary from 1.6 acre to 8 acres with a median plot size of 2.4 acres. During our study period the average wheat yield was 883 kilograms per acre (kg/ac) with a standard deviation (s.d.) of 328 kg/ac. Farmers in our study predominantly used high-yielding variety (HYV) seeds. Specifically, at an average across villages and years ~70% farmers used HYV seeds for growing wheat in the Rabi season. An average farmer applied 29 kg/ac nitrogen and 7.8 hours/ac irrigation for growing wheat and wheat price was on-average Rs. 12.5 per kg. As for land quality, about 35% of plot acreage where wheat was grown was classified as erosive. An average wheat growing household attained ~6 years of schooling. We also included the distance between plot and household and the distance between plot and irrigation source to capture the time and fuel cost for input application.

Given that cropping in India is a multi-season activity, year t also corresponds to the growing-season of the crop whose yields are being modelled. For example, while modelling wheat yields t refers to the Rabi season of a particular cropping year. Further, variable N_{it} represents per-acre application of nitrogen from fertilizers in kilograms (kg) and I_{it} represents per-acre hours of irrigation. The VDSA provides data on the quantity of inputs applied along with the date of application during each year's growing-season. The input variables in model (2) are the aggregate quantities for the whole growing-season divided by the area of the plot. We use weather station-level data to be able to control for both temperature and precipitation. Figure A2

and Table A2 in the appendix provide a matching between the individual villages and the nearest weather station. This implies that weather is captured at the village-level and is taken to be the same for all plots in the village.

3.3 Estimation Results

The IV regression results in Table 5 reveal that nitrogen prices are negatively associated with the nitrogen application levels, whereas higher electricity based irrigation prices are associated with more irrigation (as also discussed under model diagnostics). Further, the HYV seeds are associated with lower nitrogen and higher irrigation levels relative to the local or hybrid varieties. Growing nitrogen-intensive crops in previous seasons is found to have a significant positive impact on nitrogen application, while the impact of crops with nitrogen-fixation properties is found to be negative but insignificant. Higher distance between plot and the nearest irrigation source leads to reduced irrigation levels, indicating opportunity cost. Education level is found to be positively (negatively) correlated with higher irrigation (lower nitrogen) levels. An effect similar to higher education and irrigation levels is observed for older farmers. Higher average temperatures are associated with lower application of fertilizers and irrigation (although the impact on irrigation is insignificant), while higher rainfall is linked with reduced fertilizer and irrigation application. Farms with erosive soils apply relatively low irrigation.

The beta regression model for wheat yields is estimated separately for the mean yield parameter, i.e., location submodel, and the precision parameter, i.e., dispersion submodel, in Table 6. We find a concave relationship between nitrogen and wheat yields, in that mean yields increase (2.4%) with higher nitrogen application but at a decreasing rate (-0.03%). Similar finding emerges for irrigation, however the coefficient on the quadratic term is insignificant. Further, higher temperature and rainfall has a positive impact on wheat yields, however this

impact has been becoming less positive over the years. Moreover, wheat yields exhibit overall declining trends, specifically a 4% reduction on mean yields during each year in our sample. Last but not the least, a unit kilometer increase in distance between farm and irrigation source is found to cause a significant reduction in wheat yield to the tune of 25%. Landowners have, on-average, 21% higher yields relative to rental farmers and farm laborers. Both nitrogen and irrigation reduce yield dispersion in a statistically significant manner.

In order to further characterize the impact of nitrogen and irrigation on wheat yields we first plot the wheat yield density for an average farmer in our sample, i.e., at average nitrogen and irrigation levels. We find $\hat{\omega}(\bar{\bar{x}}) = 13$ and $\hat{\tau}(\bar{\bar{x}}) = 11.1$ implying that wheat yields are positively skewed (see figure 2a). This figure also helps the reader visualize the bias in wheat yield density if we were to fit a normal distribution (with same mean and standard deviation as the beta distribution). Figure 2b shows the comparison between irrigated wheat yield density and rainfed wheat yield density ($\hat{\omega}(\bar{\bar{x}}) = 8.1$ and $\hat{\tau}(\bar{\bar{x}}) = 7.3$). Clearly, not only the mean irrigated yields are higher than their rainfed counterparts, but we find higher variance (and higher probability of low yields) for the case of rainfed yields relative to the irrigated yields. This means that non-irrigated wheat production is a riskier production activity in semi-arid regions of India.

Finally, we compare the change in wheat yield density upon a marginal change in irrigation and nitrogen inputs for irrigated farms, and the change in density upon marginal change in nitrogen for rainfed farms. The results are visualized in figures 3a-b. We find conclusive evidence that the wheat yield density with higher nitrogen and/or irrigation inputs first-order stochastically dominate the yield density with relatively lower input levels. The result is consistent for irrigated and rainfed farms, however we find the risk-reducing property of nitrogen to be stronger than that of irrigation (see figure 3a)

3.3.1 Model Selection and Diagnostics

The validity and consistency of conditional yield distribution estimates will rely upon the choice of instruments, specification of the conditional yield model and the specification of input-conditioned errors. Therefore, before presenting estimation results we discuss the various challenges in implementing our *BetaIV* framework and the underlying logic for model selection. Focusing on Tables A5-A7 in the appendix will help the reader in developing an understanding upon arguments presented in this subsection.

First, in implementing the IV regression for nitrogen and irrigation we include output and input prices that are derived from self-reports of the farmers and could be biased such that it suits their decisions, including for the application of nitrogen and irrigation. We make an attempt to circumvent the issue by using nominal (average) village-level prices (base year = 2001). We discover that input and output prices are statistically significantly correlated (correlation between urea and wheat price is 0.72; irrigation and wheat price is 0.6; and irrigation and urea price is 0.41). Hence, we include either own-input price or the output price at-a-time to estimate the IV model variants. However, the issue of higher input levels associated with higher input price and lower output price still persists.

To reconcile the problem we downloaded state-level wheat wholesale prices from an independent data source (www.agmarket.gov.in), which too is found to be statistically significantly correlated with village-level urea price (0.39) and irrigation price (0.44), although the degree of correlation is relatively small. We find that the sign of the wheat price coefficient from the new series is positive for nitrogen and irrigation application, and the price of nitrogen is negatively associated with application levels. However, the positive coefficient for irrigation price remains unresolved (see Table 5).

Note that in the IV regressions we include a *count* of nitrogen intensive, nitrogen-fixation, and water-intensive crops grown in past four growing seasons. Soil-nutrient demands in any given period are often driven by past cropping choices (Du et al. 2012). A plausible issue is that, though from the previous period, the past cropping data are farmers' choice variables that could lead to these being invalid instruments when considering that a farmer might plan farming choices a year in advance. To circumvent this issue we conduct a robustness check by including count variable for previous two seasons. We find our results to be robust. Furthermore, notice that count variable is a non-linear representation of past cropping choices and is therefore not likely to be linearly related (i.e., correlated) with unobserved covariates in the yields regression.

Another issue could be due to loss of degrees of freedom due to simultaneous estimation of the IV regressions for nitrogen and irrigation, i.e., the simultaneous Tobit models, possibly resulting in large regression errors. We observe that the estimating covariance between the error term of IV regressions for nitrogen and irrigation is significant in some cases while insignificant in others. Moreover, the standard errors of estimated model parameters remain largely the same while allowing for covariance between model errors and otherwise. On this account we chose to estimate the input regressions in a simultaneous fashion.

We then conduct tests for weak instruments and those for over-identifying restrictions. Staiger and Stock (1997) provided a decision rule for weak instruments such that when the regression F-statistic is less than 10 we infer that the instrument is weak. However, the above rule was developed for a single endogenous variable. Stock and Yogo (2005) extended this decision rule for more than one endogenous regressor, where the critical F-value = 20.27 is more appropriate. We tabulate the F-statistic for the first stage of the IV regressions in Table A6.

The Sargan's test for over-identifying restrictions was implemented as described in Wooldridge (2002, pp. 123). Specifically, the estimated residuals of the parent regression equation are regressed on all IVs and the respective exogenous variables from the parent regression model. The R^2 value of this auxiliary regression is multiplied by the number of observations (N), which is chi-squared distributed. That is, $NR^2 \sim \chi_Q^2$, where Q is the number of over-identifying restrictions, which is equal to the total number of instruments. The null hypothesis of the Sargan's test is that the excluded instruments are correctly excluded from the wheat yield model in equation (10). The results are presented in Table A7.

Note that although wheat prices from www.agmarket.gov.in are found to have the correct (as per our expectation) sign in nitrogen and irrigation IV regressions the over identifying restrictions are violated. Therefore, we excluded the results from those model variants in this paper. Further, a variable selection issue emerges in estimating the yield-weather relationship. On the basis of our findings above we report the most appropriate model results in Tables 5-6.

4. Modeling interactions between input application decisions and index insurance adoption

In this section we characterize the interactions between demand for index-insurance and a risk-reducing farm input (in line with the evidence in previous section) to enable simulations of their cross-price effects. Denote the individual farm i 's production function (i.e., per-acre output or yield) as the Just-Pope technology. That is,

$$y_i = m(x_i) + \varepsilon \cdot v(x_i), \quad (12)$$

where $m(x_i)$ is the mean-yield component arising from the per-acre input application on farm i conditional on soil quality, farm mechanization, farmer's know-how, and so on; $\varepsilon \cdot v(x_i)$ is the yield-risk component due to random events like adverse weather, pest infestation, etc., such that

ε is random term with $E(\varepsilon) = 0$ and $v(x_i)$ captures the impact of input x on yield-risk. We assume that $m_x > 0$, $m_{xx} > 0$, $m(0) > 0$. Given the empirical evidence in previous section we know that x_i is a risk-reducing input, i.e., $v_x < 0$. We also assume that $v(x_i) > 0 \forall x_i > 0$, following Feder (1977, 1980). For simplicity of notation, we henceforth drop the subscript i .¹⁰

In this setting we now consider farmer's decision of buying a yield index-insurance product for land share $\alpha \in [0, 1]$ along with input application, x . We first describe the components of a yield-index insurance product as under:

- Insurer (or the planner) designates an area, A , for which the yield-index insurance will be made available to the region's farmers. Insurance design will include setting premium rate, a threshold yield that determines whether farms in the area are eligible for an indemnity (i.e., when an area-level yield-index is lower than the threshold yield), and the percentage loss covered (i.e., fraction of the scale of finance). Note that A contains a large set of farms and is publicly known at the time of deciding whether or not to buy this insurance.

- The yield-index is defined as the spatial average of yields in area A , denoted as

$\bar{y}_A = \int_{i \in A} y_i f_{y,A} dy$. Note that the yield-index is realized only at the end of the growing season, and

is unknown to the farmer at the time of making the insurance purchase decision.

- The threshold yield level, y_T , is set by the insurer before the growing season such that whenever $\bar{y}_A < y_T$ then all the insured farmers with lands in area A are eligible of a uniform

¹⁰ Note also that we posit y_i to be a constant returns to scale technology, which is reasonable for India's smallholder farms, meaning that for a farm of size L , $L \cdot$

$M(X, L) = L \cdot M\left(\frac{X}{L}, 1\right) = L \cdot m(x)$ and $V(X, L) = L \cdot V\left(\frac{X}{L}, 1\right) = L \cdot v(x)$.

insurance payout of $\varsigma(y_T - \bar{y}_A)$ where $\varsigma \in (0,1)$ ¹¹; and whenever $\bar{y}_A \geq y_T$ then none among the insured farmers are eligible for an insurance payout. Note that y_T is known to the farmers at the time of making their insurance decision.

The index with respect to the threshold yield is defined $\bar{y}_A = y_T + h$ such that $\eta < 0$ implies insurance payout for all insured farmers in the designated area A, and $\eta \geq 0$ implies that no farmer is eligible for a payout. A farmer believes that the index is a convex combination of the individual yield and an aggregate random variable (ξ) such that $\bar{y}_A = \rho y_i + (1-\rho)\xi$ where ρ is a belief parameter – the strength of association between farmers individual yield and the index. It is possible that farmer may even believe that the index is uncorrelated with individual farm yield, i.e., $\rho = 0$. This implies $\eta = \rho(y_i - y_T) + (1-\rho)(\xi - y_T)$.

Basis risk of an index insurance: It is possible that farmer i incurs a loss, i.e., $y_i < y_T$, but is ineligible for an indemnity because $\bar{y}_A \geq y_T$ or $\eta \geq 0$. Moreover, the farmer might become eligible for an insurance payout even when he does not incur a loss, i.e., $y_i \geq y_T$ but $\bar{y}_A < y_T$ (or $\eta < 0$)¹². This feature of contractual nonperformance, whereby a farmer's 'bad' year becomes 'worse' due to no indemnity or a 'good' year becomes even better due to an insurance payout, is termed as the basis risk of index insurance. These scenarios were shown to yield a non-

¹¹ β signifies the coverage level of the potential losses set by the index insurance provider. In India the Prime Minister's crop insurance scheme, termed as PMFBY, the indemnity is set to be $\frac{y_T - \bar{y}_A}{y_T} \times SI$, where SI is called sum insured or scale of finance. Therefore, $\varsigma = \frac{SI}{y_T}$.

¹² Note that we are assuming that a farmer designates 'loss' on his field relative to the area-level threshold y_T . This is reasonable as the farmer observes y_T at the time of deciding whether to buy this insurance product so it makes sense to designate a 'loss' and the related potential insurance payout relative to the threshold y_T .

monotonic relationship of rational demand of index insurance with risk aversion, wealth and price (Clarke 2016). We analyze the rational demand for yield-index insurance in the presence of highly subsidized risk-reducing inputs like nitrogen and irrigation, in the context of smallholder farmers. For this we first adapt the four possible outcomes states of insurance performance (or nonperformance) given the production function specification in equation (11):

State 1 (s1): Zero loss ($y_i \geq y_T$) and zero indemnity ($\bar{y}_A \geq y_T$). This scenario can be written in terms of random variables ε and η . That is, $\varepsilon \geq (y_T - m(x))/v(x)$ and $\eta \geq 0$. Henceforth, we denote $(y_T - m(x))/v(x) \equiv \tilde{\varepsilon}(x)$.

State 2 (s2): Zero loss ($y_i \geq y_T$) and positive indemnity ($\bar{y}_A < y_T$). Similar to the above case, this scenario can be rewritten as $\varepsilon \geq \tilde{\varepsilon}(x)$ and $\eta < 0$.

Stage 3 (s3): Positive loss ($y_i < y_T$) and zero indemnity ($\bar{y}_A \geq y_T$). That is, $\varepsilon < \tilde{\varepsilon}(x)$ and $\eta \geq 0$.

Stage 4 (s4): Positive loss ($y_i < y_T$) and positive indemnity ($\bar{y}_A < y_T$). That is, $\varepsilon < \tilde{\varepsilon}(x)$ and $\eta < 0$.

We first designate marginal probabilities of p and $1 - p$ to the ‘zero loss’ and ‘positive loss’ scenarios, respectively, for representative farmer i . Hence, $\Pr(\varepsilon \geq \tilde{\varepsilon}(x)) = p(x)$ and

$\Pr(\varepsilon < \tilde{\varepsilon}(x)) = 1 - p(x)$. Similarly, we denote the marginal probabilities of ‘zero indemnity’ as q , i.e., $\Pr(\eta \geq 0) = q$, and of ‘positive indemnity’ as $1 - q$, i.e., $\Pr(\eta < 0) = 1 - q$, in area A .

Further, the joint probability of contractual nonperformance whereby the basis risk in index insurance turns a bad year to worse, i.e., **s3**, is denoted as $\Pr(\varepsilon < \tilde{\varepsilon}(x), \eta \geq 0) = r(x)$. The corresponding joint probabilities of the other three states are provided in the following table.

	Indemnity = 0; $\eta \geq 0$	Indemnity > 0; $\eta < 0$	
Loss = 0; $\varepsilon \geq \tilde{\varepsilon}(x)$	$q - r(x)$	$p(x) + r(x) - q$	$p(x)$
Loss > 0; $\varepsilon < \tilde{\varepsilon}(x)$	$r(x)$	$1 - p(x) - r(x)$	$1 - p(x)$
	q	$1 - q$	

Furthermore, it is reasonable to expect for a risk-reducing input that $p_x > 0$, and thus

$\partial \Pr(\varepsilon < \tilde{\varepsilon}(x)) / \partial x < 0$. That means $\partial F_\varepsilon(\tilde{\varepsilon}(x)) / \partial x < 0$, which would imply that $\partial \tilde{\varepsilon}(x) / \partial x < 0$ ¹³.

$\partial \tilde{\varepsilon}(x) / \partial x < 0$ implies $\partial \{(y_T - m(x)) / \nu(x)\} / \partial x < 0 \Rightarrow m + \left(\frac{m_x}{-\nu_x} \right) h > y_T$, which is a constraint

on how yield depends on x for it be a risk-reducing technology given the exogenous threshold yield set by the index-insurance provider. This is a technology-driven incentive compatibility argument for a farmer to consider substitutability between the index-insurance and the risk-reducing chemical input.

We now develop an understanding of the sign of change in $r(x)$ upon a marginal increase in x . See that $r(x) = \Pr(\varepsilon < \tilde{\varepsilon}(x), \eta \geq 0) = \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0) \cdot \Pr(\eta \geq 0) = \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0) \cdot q$.

We can also write $\Pr(\varepsilon < \tilde{\varepsilon}(x)) = \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0) + \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta < 0)$, and so

$$\frac{\partial \Pr(\varepsilon < \tilde{\varepsilon}(x))}{\partial x} = \underbrace{\frac{\partial \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0)}{\partial x}}_{T1 = \partial r(x) / \partial x} + \underbrace{\frac{\partial \Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta < 0)}{\partial x}}_{T2}.$$

¹³ $\partial F_\varepsilon(\tilde{\varepsilon}(x)) / \partial x = f_\varepsilon(\tilde{\varepsilon}(x)) \cdot \partial \tilde{\varepsilon}(x) / \partial x$.

To develop a further understanding of the terms T1 and T2, we can specify

$\eta = \rho\varepsilon + (1-\rho)\gamma$. So then the condition $\eta \geq 0$ in T1 translates as $\varepsilon \geq -\frac{(1-\rho)}{\rho}\gamma$ ultimately

implying that $\Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0) = \int_{-\lceil(1-\rho)/\rho\rceil\gamma}^{\tilde{\varepsilon}(x)} dF_\varepsilon$ whenever $\tilde{\varepsilon}(x) + \lceil(1-\rho)/\rho\rceil\gamma > 0$ and

$\Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \leq 0) = \int_{\varepsilon_{\min}}^{\tilde{\varepsilon}(x)} dF_\varepsilon = F_\varepsilon(\tilde{\varepsilon}(x))$ whenever $\tilde{\varepsilon}(x) + \lceil(1-\rho)/\rho\rceil\gamma \leq 0$. Given that

$\partial\tilde{\varepsilon}(x)/\partial x < 0$ we can say that $\partial\Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \geq 0)/\partial x < 0$ whenever $\tilde{\varepsilon}(x) + \lceil(1-\rho)/\rho\rceil\gamma > 0$,

and because x is a risk-reducing input $\partial\Pr(\varepsilon < \tilde{\varepsilon}(x) | \eta \leq 0)/\partial x < 0$ also when

$\tilde{\varepsilon}(x) + \lceil(1-\rho)/\rho\rceil\gamma \leq 0$. Therefore, the basis risk of contractual nonperformance in case of farm

level losses is decreasing in x , i.e., $\partial r(x)/\partial x < 0$ or $r_x < 0$.

Finally, for the index insurance product to signal positive losses in an area we should have the following condition to hold true: $\text{Corr}(I, L) > 0$ such that I represents the index insurance payout of $\beta(y_T - \bar{y}_A) \equiv I_A$ whenever $y_T > \bar{y}_A$ or 0 otherwise, and L represents farm-level loss of $P_y(y_T - y_i) \equiv L_i$ whenever $y_T > y_i$ or 0 otherwise. That is, the event of payout from index insurance and the event of losses in an area are positively correlated¹⁴. We expand this condition

¹⁴ This condition is an alternative interpretation of that in Clarke (2016) pp. 287: $\Pr(s4)/\Pr(s2) > \Pr(s3)/\Pr(s1)$.

$$\begin{aligned}
Cov(I, L) &= E[IL] - E[L]E[I] \\
E[IL] &= (q - r) \cdot 0 \cdot 0 + (p - q + r) \cdot 0 \cdot I_A + r \cdot L_i \cdot 0 + (1 - p - r) \cdot L_i \cdot I_A, \\
E[L] &= p \cdot 0 + (1 - p) \cdot L_i, \\
E[I] &= q \cdot 0 + (1 - q) \cdot I_A, \\
Cov(I, L) &= (1 - p - r) L_i I_A - (1 - p)(1 - q) I_A L_i = [(1 - p)q - r] I_A L_i. \\
\text{So, } Cov(I, L) &> 0 \\
\Rightarrow (1 - p)q &> r.
\end{aligned}$$

The condition $(1 - p)q > r$ also implies $\Pr(s1) > 0$; $\Pr(s3) > 0$ and $\Pr(s4) > 0$. However, in order to ensure that $\Pr(s2) > 0$ we need to assume that $p > q - r$.

Price of an index insurance: We denote the per-acre price of the yield-index insurance as C_I . An actuarially fair index insurance will be priced (per-acre) at $C_I = (1 - q)\gamma(y_T - \bar{y}_A)$; an actuarially unfair index insurance will be priced at $C_I > (1 - q)\gamma(y_T - \bar{y}_A)$; and an actuarially favorable index insurance will be priced at $C_I < (1 - q)\gamma(y_T - \bar{y}_A)$.

We now present the expected utility-maximizing farmer's decision problem with the choice variables being the level of input application, x , and the share of land, α , to be insured with the index insurance product. Before specifying the expected utility maximization problem we define payoffs in all four possible states of contractual performance on the share of land where index insurance is purchased (αL) and the remaining land share that is uninsured, i.e., $(1 - \alpha)L$. We assume that the concerned farmer has one-acre holding, i.e., $L = 1$.

[See table on the next page]

Description	$L = 0; I = 0$ (s1)	$L = 0; I = I_A$ (s2)	$L = L_i; I = 0$ (s3)	$L = L_i; I = I_A$ (s4)
Probability	$q - r(x)$	$p(x) + r(x) - q$	$r(x)$	$1 - p(x) - r(x)$
Payoff without index insurance	$\{P_y(m + \varepsilon v) - P_x x\}(1 - \alpha)$	$\{P_y(m + \varepsilon v) - P_x x\}(1 - \alpha)$	$\{P_y(m + \varepsilon v) - P_x x - L_i\}(1 - \alpha)$	$\{P_y(m + \varepsilon v) - P_x x - L_i\}(1 - \alpha)$
Payoff with index insurance	$\{P_y(m + \varepsilon v) - P_x x - C_I\} \alpha$	$\{P_y(m + \varepsilon v) - P_x x - C_I + y_T - \bar{y}_A\} \alpha$	$\{2P_y(m + \varepsilon v) - P_x x - C_I - y_T\} \alpha$	$\{2P_y(m + \varepsilon v) - P_x x - C_I - \bar{y}_A\} \alpha$
Total Payoff	$\{P_y(m + \varepsilon v) - P_x x\}(1 - \alpha)$ $+$ $\{P_y(m + \varepsilon v) - P_x x - C_I\} \alpha$	$\{P_y(m + \varepsilon v) - P_x x\}(1 - \alpha)$ $+$ $\{P_y(m + \varepsilon v) - P_x x - C_I + y_T - \bar{y}_A\} \alpha$	$\{P_y(m + \varepsilon v) - P_x x - L_i\}(1 - \alpha)$ $+$ $\{2P_y(m + \varepsilon v) - P_x x - C_I - y_T\} \alpha$	$\{P_y(m + \varepsilon v) - P_x x - L_i\}(1 - \alpha)$ $+$ $\{2P_y(m + \varepsilon v) - P_x x - C_I - \bar{y}_A\} \alpha$

Hence, the landowner's expected utility maximization problem is

$$\begin{aligned} \max_{\alpha, x} EU = & \left[q - r(x) \right] \times \left[U \left(\left\{ P_y(m + \varepsilon v) - P_x x \right\} (1 - \alpha) + \left\{ P_y(m + \varepsilon v) - P_x x - C_I \right\} \alpha \right) \right] \\ & + \left[p(x) + r(x) - q \right] \times \left[U \left(\left\{ P_y(m + \varepsilon v) - P_x x \right\} (1 - \alpha) + \left\{ P_y(m + \varepsilon v) - P_x x - C_I + y_T - \bar{y}_A \right\} \alpha \right) \right] \\ & + \left[r(x) \right] \times \left[U \left(\left\{ P_y(m + \varepsilon v) - P_x x - L_i \right\} (1 - \alpha) + \left\{ 2P_y(m + \varepsilon v) - P_x x - C_I - y_T \right\} \alpha \right) \right] \\ & + \left[1 - p(x) - r(x) \right] \times \left[U \left(\left\{ P_y(m + \varepsilon v) - P_x x - L_i \right\} (1 - \alpha) + \left\{ 2P_y(m + \varepsilon v) - P_x x - C_I - \bar{y}_A \right\} \alpha \right) \right] \end{aligned}$$

, which can be re-written as

$$\max_{\alpha, x} EU = \left[q - r(x) \right] \times U_{00} + \left[p(x) + r(x) - q \right] \times U_{0I_A} + \left[r(x) \right] \times U_{L_0} + \left[1 - p(x) - r(x) \right] \times U_{L_I A}.$$

Observations:

1. Once we consider the loss levels to be fixed/given, the individual farmer would prefer high indemnity scenario to low indemnity scenario. That is, higher value of $y_T - \bar{y}_A$, and a higher probability weight on $I = I_A$ relative to $I = 0$ (which is trivially true).
2. However, the level of loss is endogenous as it depends on x through the value of $\tilde{\varepsilon}(x)$.

The first-order conditions of optimization can thus be written as:

$$\alpha : (q - r) P_I U'_{00} + (p + r - q) (P_I - (y_T - \bar{y}_A)) U'_{0I_A} + r (P_I + y_T) U'_{L_0} + (1 - p - r) (P_I + \bar{y}_A) U'_{L_I A} \stackrel{set}{=} 0$$

x :

$$\begin{aligned} (q - r) \rho_x U_{00} + (p + r - q) \rho_x U_{0I_A} + r (\rho_x + R_x) U_{L_0} + (1 - p - r) (\rho_x + R_x) U_{L_I A} \\ - \left\{ r_x (U_{00} - U_{L_0}) + (p_x + r_x) (U_{L_I A} - U_{0I_A}) \right\} \stackrel{set}{=} 0 \end{aligned}$$

where,

$$R(x) = P_y(g + eh), \quad R_x = \nabla R(x) / \nabla x;$$

$$\rho(x) = R(x) - cx, \quad \rho_x = \nabla \rho(x) / \nabla x.$$

4.1 Simulations

The comparative static analysis of the utility maximizing farmer was simulated to obtain the own- and cross-price elasticity of inputs (nitrogen and irrigation) and index insurance. We closely follow India's index insurance program (Bhushan and Kumar 2017) and simulate the insurance choice decisions, which serve as an alternative of empirical insurance demand function that are not possible to estimate due to inavailability of the relevant micro-level insurance adoption data for Indian farmers. Moreover, insurance was compulsorily tied with agricultural credit making an isolated estimation of its demand challenging even if farm-level insurance data were available¹⁵.

We simulate the optimal input demands and the index insurance for rain-fed and irrigated wheat for a representative farmer. The optimal input demands for a farmer are simulated through the utility maximizing input choices for given set of input prices. We change the price of one of the inputs keeping the other prices fixed to derive the own price and the cross price responsiveness of the different input prices.

To perform simulations we discretize the input space into a choice grid for urea in (0,60) range and for irrigation in (0,50) range with each grid cell representing one unit of input application. For percent insurance coverage that ranges between 0 and 1 we discretize the choice space into grid cells of size 0.01. For each grid cell that designates to choice set {urea x irrigation x insurance}, we randomly draw yields 100 times for an average farm (based on the conditional yield density estimation presented in Figures 2-3) and the associated area-level yield index

¹⁵ Recently the insurance adoption was made voluntary for all farmers in India. Our results will also provide insights on this policy.

(using different values of the correlation parameter ρ)– giving a 100 x 100 state space for the each input grid-cell. Profits and utility were computed to arrive at the expected profit and the expected utility for an average farm. The optimal input choice corresponds to the maximum of the expected utility values across the three dimensional input grids. We separately perform simulations for rain-fed wheat production by setting the irrigation levels to zero. We use a CRRA¹⁶ utility function – with different levels of risk aversion parameters ($\alpha = 0, 0.5, 2, 3, 4$). We obtain the input demands with respect their own-price and cross-price for the different levels of risk aversion. The representative farmer is constructed using the average values of the regression variables for the Papda village in Madhya Pradesh, India. The insurance related parameters are drawn from the PMFBY web portal – where the value of sum insured per acre is Rs. 9,200 and the threshold yield is 700 kg/ac. (i.e. indemnity level (80%) x average of past 7 years yield (875 kg/ac.)). In the baseline scenario we assume $\rho = 0.75$ (i.e. an individual farmer's belief about the association between farm and index yield). We also obtain the changes in optimal input demand as ρ changes.

Figures 4-5 present the comparative statics results of the price changes for rain-fed and the irrigated wheat production respectively. For both– rain-fed and irrigated- wheat the input demand with respect to its own-price is downward sloping as expected (see Figures 4(a-b) and 5(a-c)). However, it is apparent that the urea demand is the most inelastic and insurance demand is highly price elastic – i.e. small changes in insurance price can drive demand significantly. We also observe that the demand for urea and irrigation increases while that of index insurance decreases at higher risk aversion levels. This is indicative of our earlier conjecture that farmers rely more urea and irrigation for risk mitigation than they do on index insurance.

¹⁶ A CRRA utility function can be represented as a DARA utility function.

The cross-price effects in the case of rainfed wheat (see Figures 4(c-d)) suggest that the cross-price effects on urea demand is quite small. However, the demand for index insurance is quite sensitive to a unit percent change in the price of urea for risk-averse farmers. For irrigated wheat farms, the effects of change in the price of urea is found to have a marginal impact on irrigation and insurance demand, and similarly change in insurance price is found to have a marginal impact on irrigation and urea demand. An increase in cost of irrigation, however, significantly increases the demand for urea and index insurance for moderately to highly risk averse farmers¹⁷.

5. Concluding remarks

We provide fresh evidence on wheat yield distribution conditional on nitrogen and irrigation inputs for smallholder farmers in a developing country setting. In particular, we employ a new *BetaIV* framework to control for endogeneity between farm inputs like nitrogen and irrigation and wheat yields for smallholder farms in semi-arid Indian villages. Wheat yields are found to be positively-skewed and both nitrogen and irrigation are found to be risk-reducing inputs in first-order stochastic dominance sense. The marginal risk-reducing impact of nitrogen is found to be greater than that of irrigation. We use these findings to simulate the interaction between optimal demands of index insurance and risk characterizing inputs: nitrogen and irrigation, for smallholder farms of Papda village in Madhya Pradesh, India, which is a first-of-its-kind analysis in developing country setting. We simulate demand at different levels of basis risk due to index insurance non-performance and different risk aversion levels of farmers. Our findings suggest that the demand for urea and irrigation increases while that of index insurance decreases at

¹⁷ The irrigation prices are constructed in hypothetical and notional sense because irrigation are either under-developed or missing in India.

higher risk aversion levels. Further, increase in cost of irrigation increases the demand for urea and index insurance in the case of moderately-to-highly risk averse farmers.

Acknowledgments: The authors are very grateful to Celine Nauges, Jeffrey Vincent, David A. Hennessy and Hongli Feng for their invaluable advice. We especially acknowledge Sumedha Shukla for her efficient research assistance. All errors are our own.

References

- Alderman, H., Haque., T. 2007. Insurance against covariate shocks: The role of index based Insurance in social protection in low-income countries of Africa. World Bank Working Paper No. 95. Washington, DC: World Bank.
- Atwood, J., Shaik, S., Watts, M. 2003. Are crop yields normally distributed? A reexamination. *American Journal of Agricultural Economics*, 85(4): 888–901.
- Atwood, J., Shaik, S., Watts, M. 2007. Can normality of yields be assumed for crop insurance? *Canadian Journal of Agricultural Economics*, 50(2): 171-184.
- Babcock, B.A., Hennessy, D.A. 1996. Input demand under yield and revenue insurance. *American Journal of Agricultural Economics*, 78(2): 416-427.
- Bhushan, C., Kumar, V. 2017. Pradhan Mantri Fasal Bima Yojana: An assessment. Center for Science and Environment, New Delhi. Available [here](#). Accessed on October 2019.
- Buccola, S.T. 1986. Testing for nonnormality in farm net returns. *American Journal of Agricultural Economics*, 68(2): 334-343.
- Cai, J., A. de Janvry, and E. Sadoulet. 2011. Social Networks and Insurance Take Up: Evidence from a Randomized Experiment in China. Geneva, Microinsurance Innovation Facility Research Paper No. 8, International Labor Organization, Switzerland.
- Clarke, D.J. 2016. A theory of rational demand for index insurance. *American Economic Journal: Microeconomics*, 8(1): 283-306.
- Day, R.H. 1965. Probability distributions of farm crop yields. *Journal of Farm Economics*, 47(3): 713-741.
- Elabed, G., Carter, M.R. 2015. Compound-Risk Aversion Ambiguity and the Willingness to Pay for Microinsurance. *Journal of Economic Behavior and Organization*, 118: 150–66.
- Feder, G. 1977. The impact of uncertainty in a class of objective functions. *Journal of Economic Theory*, 16: 504-512.

- Feder, G. 1980. Farm size, risk aversion and the adoption of new technology under uncertainty. *Oxford Economic Papers*, 32(2): 263-283.
- Gallagher, P. 1986. U.S. corn yield capacity and probability: Estimation and forecasting with non-symmetric Disturbances. *North Central Journal of Agricultural Economics*, 8(1): 109-122.
- Gallagher, P. 1987. U.S. soybean yields: Estimation and forecasting with nonsymmetric disturbances. *American Journal of Agricultural Economics*, 69(4): 796-803.
- Gine, X., Townsend R., Vickery, J. 2008. Patterns of Rainfall Insurance Participation in Rural India. *World Bank Economic Review* 22 (3): 539–66.
- Goodwin, B.K., Mahul, O. 2004. Risk modelling concepts relating to the design and rating of agricultural insurance contracts. *World Bank Policy Research Working Paper No. 3392*. Available at SSRN: <https://ssrn.com/abstract=625269>. Last accessed October 2019.
- Greene, W.H. 2003. *Econometric Analysis*, fifth edition. Prentice Hall (Pearson Publisher), Upper Saddle River, NJ, U.S.A.
- Harri, A., Coble, K.H., Ker, A.P., Goodwin, B.K. 2011. Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, 93(3): 707-713.
- Hennessy, D.A. 2009. Crop yield skewness under law of the minimum technology. *American Journal of Agricultural Economics*, 91(1): 197-208.
- Hennessy, D.A. 2011. Modeling Stochastic Crop Yield Expectations with a Limiting Beta Distribution. *Journal of Agricultural and Resource Economics*, 36(1): 177-191.
- Hirshleifer, J., Riley, J.G. 1992. *The Analytics of Uncertainty and Information*. Cambridge University Press.
- Horowitz, J.K., Lichtenberg, E. 1993. Insurance, moral hazard, and chemical use in agriculture. *American Journal of Agricultural Economics*, 75(4): 926-935.
- Jensen, N., Barrett, C., Mude, A. 2016. Index Insurance Quality and Basis Risk: Evidence from Northern Kenya. *American Journal of Agricultural Economics*, 98(5), 1450-1469.
- Jensen, N., Barrett, C. 2017. Agricultural index insurance for development. *Applied Economic Perspectives and Policy*, 39(2), 199-219.
- Janzen, S.A., Carter, M.R., Ikegami, M. 2013. Valuing Asset Insurance in the Presence of Poverty Traps. UC-Davis working paper.
- Just, R.E., Weninger, Q. 1999. Are crop yields normally distributed? *American Journal of Agricultural Economics*, 81(2): 287-304.

- Kanjilal, K., Arora, G. 2019. Characterizing behavior in production decision: The case of smallholder India farmers. Selected presentation at the 2019 Agricultural and Applied Economics Association Annual Meetings, Atlanta, GA, USA.
- Ker, A.P., Tolhurst, T.N. 2019. On the treatment of heteroscedasticity in crop yield data. *American Journal of Agricultural Economics*, 101(4): 1247-1261.
- Lichtenberg, E., Zilberman, D. 1986. The econometrics of damage control: Why specification matters. *American Journal of Agricultural Economics*, 68(2): 261-273.
- Moss, C.B., Shonkwiler, J.S. 1993. Estimating yield distributions with a stochastic trend and nonnormal errors. *American Journal of Agricultural Economics*, 75(4): 1056-1062.
- Mukherjee, S., Pal, P. 2017. Impediments to the spread of crop insurance in India. *Economic and Political Weekly*, 52(35): 16-19.
- Nelson, C.H., Preckel, P.V. 1989. The conditional Beta distribution as a stochastic production function. *American Journal of Agricultural Economics*, 71(2): 370-378.
- Ozaki, V.A., Goodwin, B.K., Shirota, R. 2008. Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40(9): 1151-1164.
- Popp, M., Rudstrom, M., Manning, P. 2005 Spatial yield risk across region, crop, and aggregation method. *Canadian Journal of Agricultural Economics*, 53(2-3): 103-115.
- Quiggin, J., Karagiannis, G., Stanton, J. 1993. Crop insurance and crop production: An empirical study of moral hazard and adverse selection. *Australian Journal of Agricultural Economics*, 37(2): 95-113.
- Ramaswamy, B. 1992. Production risk and optimal input decisions. *American Journal of Agricultural Economics*, 74(4): 860-869.
- Ramirez, O.A. 1997. Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonnormal random variables: The case of Corn Belt corn, soybean and wheat yields. *American Journal of Agricultural Economics*, 79(1): 191-205.
- Ramirez, O.A., Misra, S., Field, J. 2003. Crop-yield distributions revisited. *American Journal of Agricultural Economics*, 85(1): 108-120.
- Sherrick, B.J., Barry, P.J., Ellinger, P.N., Schnitkey, G.D. 2004. Factors influencing farmers' crop insurance decisions. *American Journal of Agricultural Economics*, 86(1): 103-114.
- Shukla, S., Arora, G. 2019. Role of Formal and Informal Rural Credit in Crop Yield Distributions and Downside Risk. Selected poster at the 8th Tri-ennial International Conference on Agricultural Statistics, New Delhi, India (November 18-21, 2019).
- Smith, V.H., Goodwin, B.K. 1996. Crop insurance, moral hazard and agricultural chemical use. *American Journal of Agricultural Economics*, 78(2): 428-438.

- Staiger, D., Stock, J.M. 1997. Instrumental variables regression with weak instruments. *Econometrica*, 65(3): 557-586.
- Stock, J., Yogo, M. 2005. Testing for weak instruments in linear IV regression. In: Andrews D.W.K. (eds.) *Identification and inference for econometric models*. Cambridge University Press (New York), pp. 80-108.
- Swinton, S., King, R.P. 1991. Evaluating robust regression techniques for detrending crop yield data with nonnormal errors. *American Journal of Agricultural Economics*, 73(2): 446-451.
- Taylor, C.R. 1990. Two practical procedures for estimating multivariate nonnormal probability density functions. *American Journal of Agricultural Economics*, 72(1): 210-217.
- Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge.
- Wu, J. 1999. Crop insurance, acreage decisions, and nonpoint-source pollution. *American Journal of Agricultural Economics*, 81(2): 305-320.

TABLES

Table 1. Wheat cultivation during 2001-2014 across the VDSA survey villages.

State	Village	Wheat (Rabi)
Andhra Pradesh	Aurepalle	
Andhra Pradesh	Dokur	
Andhra Pradesh	J.C. Agraharam	
Gujarat	Babrol	X
Gujarat	Chatha	X
Gujarat	Karamdi Chingariya	X
Gujarat	Makhiyala	X
Madhya Pradesh	Papda	X
Madhya Pradesh	Rampura Kalan	X
Maharashtra	Kalman	X
Maharashtra	Kinkheda	X
Maharashtra	Kanzara	X
Maharashtra	Shirapur	X
Karnataka	Belladamadugu	
Karnataka	Kapanimbargi	X
Karnataka	Markkabinahali	X
Karnataka	Tharati	

Notes: Symbol **X** represents that the respective crop was grown in that village during 2000-2014.

Table 2. Household-level and plot-level wheat cultivation frequency during 2001-2014.

Number of years cropped	Unique households that grow wheat (Rabi)	Unique plots where wheat was grown (Rabi)
1	75	220
2	52	76
3	17	28
4	26	26
5	14	14
6	6	8
7	3	4
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
Total number of households	193	376

Table 3. Variables descriptions.

Variable	Descriptions
Dependent Variable	N = 1,586
Y_{it}	Wheat yield in kilograms per acre (kg/acre)
Explanatory Variables	
HYV_i	= 1 if high-yielding seed varieties are used; 0 otherwise.
$Hybrid_i$	= 1 if hybrid seed varieties are used; 0 otherwise.
$Local_i$	= 1 if local seed varieties are used; 0 otherwise.
N_{it}	Nitrogen applied in kg/ac
I_{it}	Hours of electricity-based irrigation in hours/acre
$P_{W,t}$	Price of wheat (rupees per kilogram, rs/kg).
$P_{N,t}$	Price of urea (rs/kg).
$P_{IRR,t}$	Price of irrigation using electric equipment (rs/hr).
$acre_i$	Total plot acres
irr_km_i	Distance between irrigation source and plot in km.
$house_km_i$	Distance between house and plot in km.
$fallow_{t-4}$	= 1 if land was fallow in the previous season; 0 otherwise.
$N_intensive_{t-4}^*$	Number of times a nitrogen intensive crop was grown in the past four seasons.
$N_fix_{t-4}^\#$	Number of times a crop with soil nitrogen fixation property was grown in the past four seasons.
$W_intensive_{t-4}^\&$	Number of times a water intensive crop is grown in the past seasons.
Age_i	Age of household head (years).
$Educ_i$	Years of education of the household head.
$Farming_i$	=1 if farming is main occupation; 0 otherwise.
$Male_i$	=1 if household head is male; 0 otherwise.
$Erosive_i$	=1 if the soils are erosive; =0 otherwise.

Notes:

* Include maize, groundnut, cotton, sorghum, sugarcane, paddy, chillies, papaya, onion, brinjal, sunflower and pearl millets (see Table A4 in appendix).

Include soybeans, greengram, blackgram, pigeonpea and cowpea (see Table A4 in appendix).

& Include paddy, sugarcane, cotton and banana (see Table A4 in appendix).

Table 4. Summary Statistics after removing outlier observations from the sample.

Variable	N	Mean	Std. Dev.	Minimum	25th Pctl	Median	75th Pctl	Maximum
Y_{it}	1,586	883.38	327.77	0.00	666.67	900.00	1,150.00	1,800.00
HYV_i	1,586	0.69	0.46	0.00	0.00	1.00	1.00	1.00
$Hybrid_i$	1,586	0.24	0.43	0.00	0.00	0.00	0.00	1.00
$Local_i$	1,586	0.07	0.26	0.00	0.00	0.00	0.00	1.00
N_{it}	1,586	28.64	16.31	0.00	18.30	27.60	39.67	77.00
I_{it}	1,586	7.84	15.03	0.00	0.00	0.00	10.67	65.00
$P_{W,t}$	1,586	12.5	1.6	9.2	11.4	11.5	14.2	21.8
$P_{N,t}$	1,586	5.9	0.4	4.7	5.5	5.8	6.3	6.8
$acre_i$	1,586	2.37	1.59	0.25	1.25	2.00	3.00	8.00
irr_km_i	1,283	0.29	0.45	0.00	0.00	0.10	0.50	3.00
$house_km_i$	1,450	1.29	2.51	0.00	0.50	1.00	1.83	62.86
$Fallow_{t-4}$	1,586	0.29	0.45	0.00	0.00	0.00	1.00	1.00
$N_intensive_{t-4}$	1,586	0.26	0.44	0.00	0.00	1.00	2.00	3.00
N_fix_{t-4}	1,586	0.26	0.44	0.00	0.00	0.00	1.00	2.00
$W_intensive_{t-4}$	1,586	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Age_i	1,586	50.83	12.48	24.00	42.00	50.00	61.00	85.00
$Educ_i$	1,584	5.82	4.62	0.00	2.00	5.00	10.00	16.00
$Farming_i$	1,586	0.83	0.38	0.00	1.00	1.00	1.00	1.00
$Male_i$	1,586	0.98	0.15	0.00	1.00	1.00	1.00	1.00
$Erosive_i$	1,586	0.35	0.48	0.00	0.00	0.00	1.00	1.00

Notes: 1st percentile wheat yield is 45 kilograms per acre. Prices are smoothed at the village level for each year. Pesticide spraying is found to be zero for 95% of the sample.

Table 5: Simultaneous Tobit model estimation for nitrogen and irrigation application

Variable	Dependent Variable = <i>NT</i>	Dependent Variable = <i>IRR</i>
<i>HYV_i</i>	-3.5*** (0.86)	11.50*** (1.61)
<i>P_{N,t}</i>	-4.99*** (0.67)	
<i>P_{IRR,t}</i>		0.19** (0.09)
<i>acre_i</i>	0.82*** (0.24)	5.29*** (0.47)
<i>N_intensive_{t-4}</i>	2.22** (0.93)	
<i>N_fix_{t-4}</i>	-1.10 (0.89)	
<i>W_intensive_{t-4}</i>		-2.15 (2.39)
<i>owner_i</i>	1.50 (1.03)	-3.58* (2.14)
<i>Age_i</i>	0.005 (0.03)	0.29*** (0.05)
<i>Educ_i</i>	-0.53*** (0.07)	1.60*** (0.14)
<i>Erosive_i</i>	0.56 (0.73)	-5.69*** (1.62)
<i>Tavg_{it}</i>	1.18*** (0.23)	-0.13 (0.44)
<i>Rain_{it}</i>	-0.13*** (0.01)	-0.08*** (0.03)
<i>irr_km_i</i>	3.5*** (0.91)	-4.18** (1.99)
<i>house_km_i</i>	-0.34 (0.21)	0.52 (0.33)
<i>Village fixed-effects</i>	YES	YES
<i>Log-Likelihood</i>		-13,902
<i>AIC</i>		27,884
<i>N</i>		1,084

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Standard errors in parentheses.

Table 6: Econometric estimation of wheat yields using Beta regression. Dependent Variable: Y_{it}

Variable	Location submodel	Dispersion submodel
N_{it}	0.024*** (0.005)	-0.029*** (0.006)
N_{it}^2	-0.0003*** (0.00006)	
I_{it}	0.016*** (0.005)	-0.032*** (0.009)
I_{it}^2	-0.00006 (0.00008)	
$Tavg_{it}$	0.169*** (0.047)	
$Rain_{it}$	0.008* (0.004)	
HYV	-0.161** (0.064)	-0.788*** (0.114)
$owner_i$	0.212*** (0.057)	
$Erosive_i$	-0.099** (0.043)	
$Educ_i$	-0.005 (0.004)	
irr_km_i	-0.252*** (0.058)	
$house_km_i$	-0.002 (0.003)	
$trend$	-0.042*** (0.015)	
$Tavg_{it} \times trend$	-0.016*** (0.004)	
$Rain_{it} \times trend$	-0.0005 (0.0004)	
$\hat{\kappa}_{it}^N$	0.003 (0.003)	0.016*** (0.007)
$\hat{\kappa}_{it}^I$	-0.006** (0.002)	0.052*** (0.008)
$-2Log-Likelihood$		1,474
AIC		-1,406
N		1,084

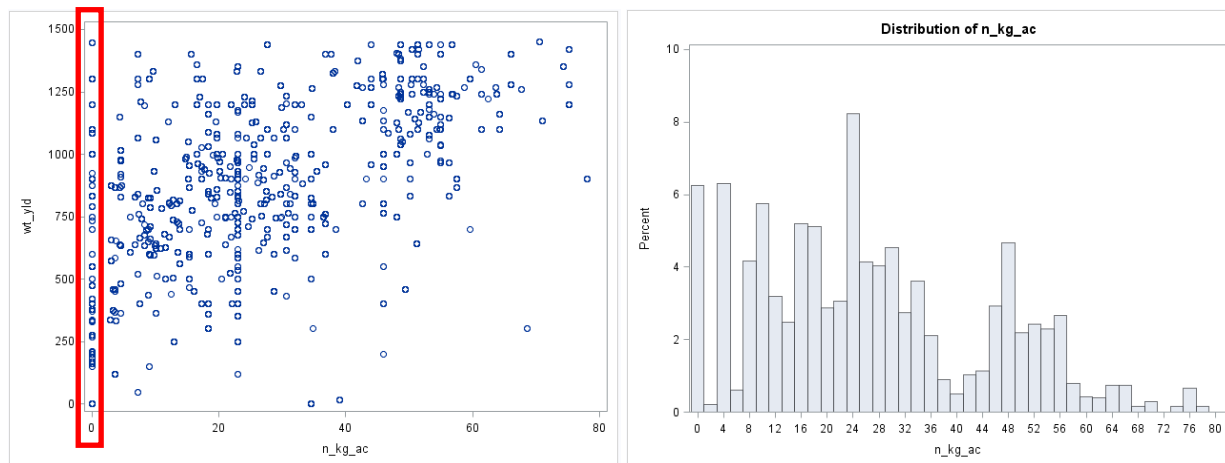
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Standard errors in parantheses

Notes: All explanatory variables in Tables 5-6 are centered on their respective means to facilitate direct interpretation of the coefficients when there are interaction terms. $\hat{\kappa}_{it}^N$ and $\hat{\kappa}_{it}^I$ represent the predicted residuals from IV regressions in Table 5 for nitrogen and irrigation respectively. Variable $trend = 1$ in 2001; $= 2$ in 2002; ...; $= 14$ in 2014.

FIGURES

Figure 1. Distribution of nitrogen (kg/acre) and irrigation (hours/acre) application for growing wheat in the VDSA villages.

Nitrogen application (number of kilograms per acre):



Irrigation application (Number of hours per acre):

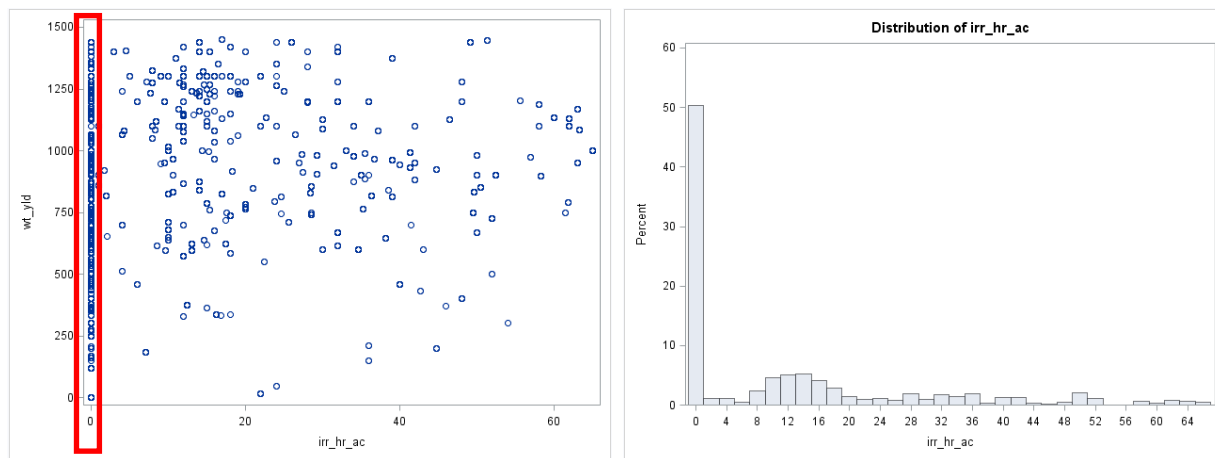
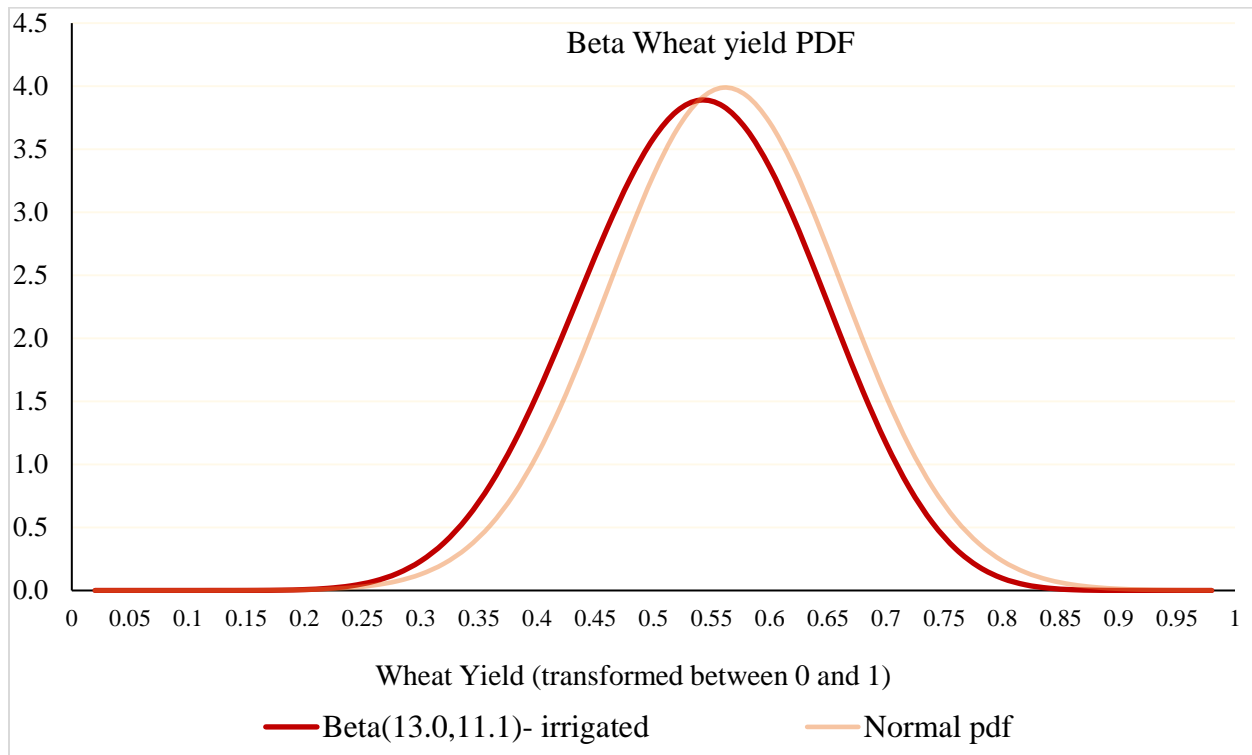
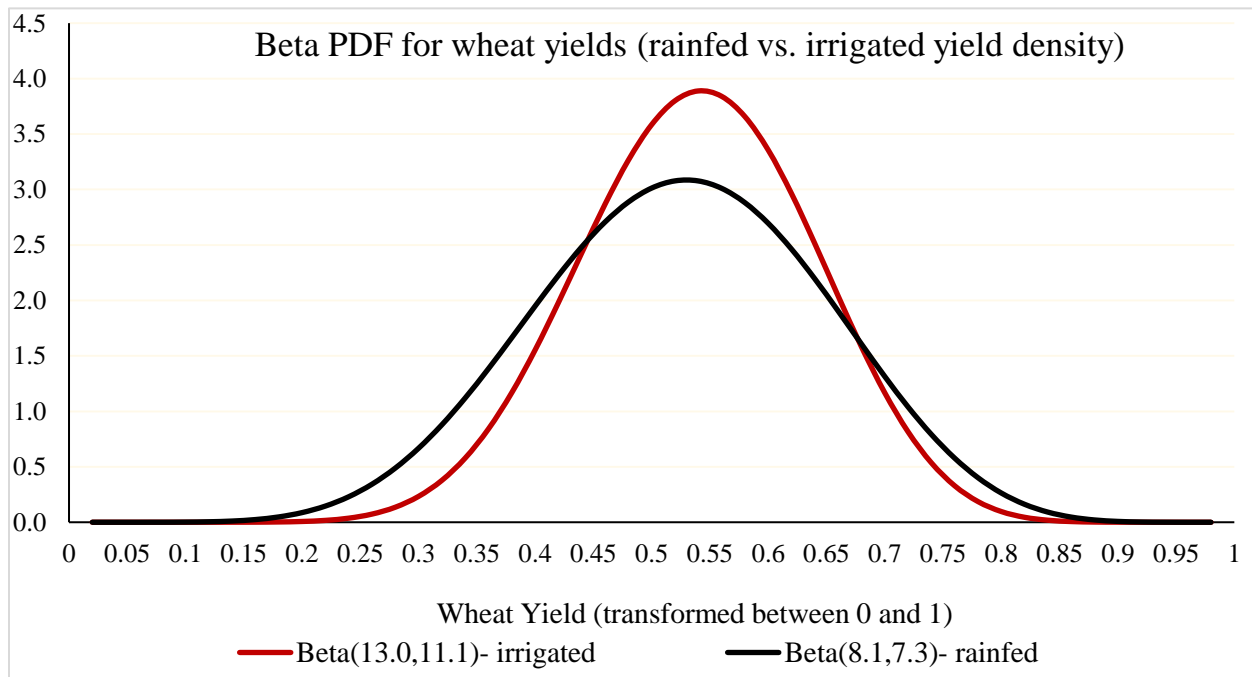


Figure 2: Input-conditioned wheat yields estimated from the *BetaIV* regression framework.

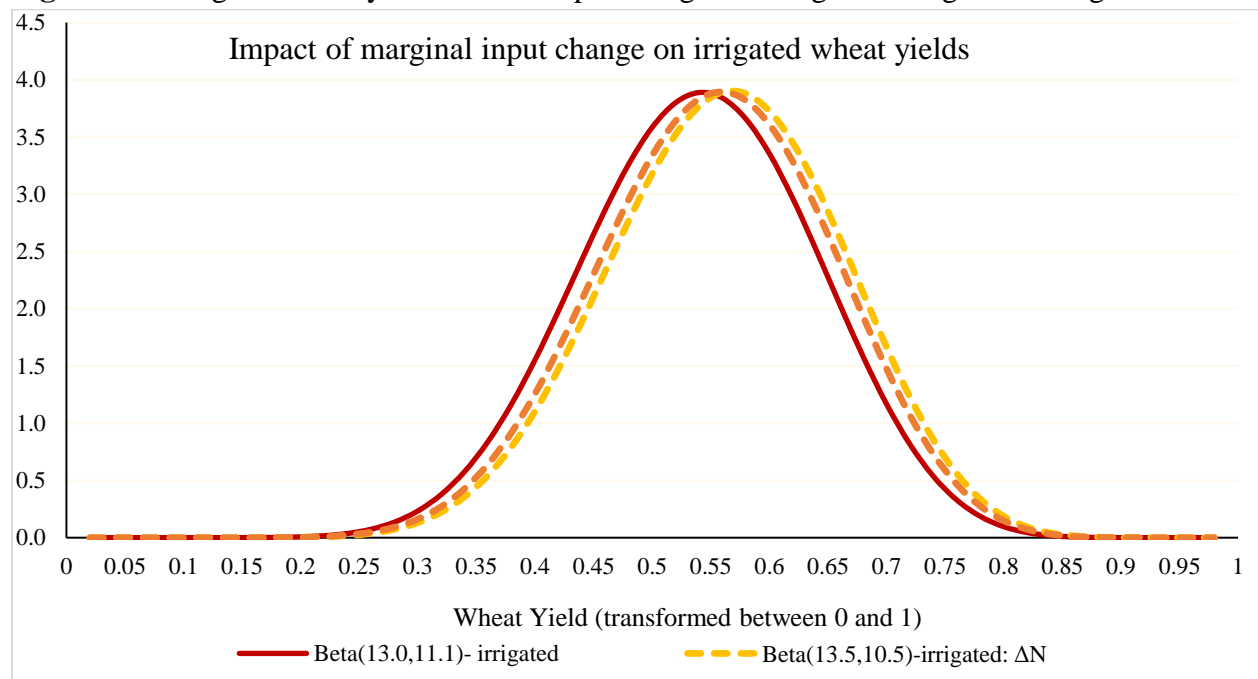


(a) Irrigated wheat yield density

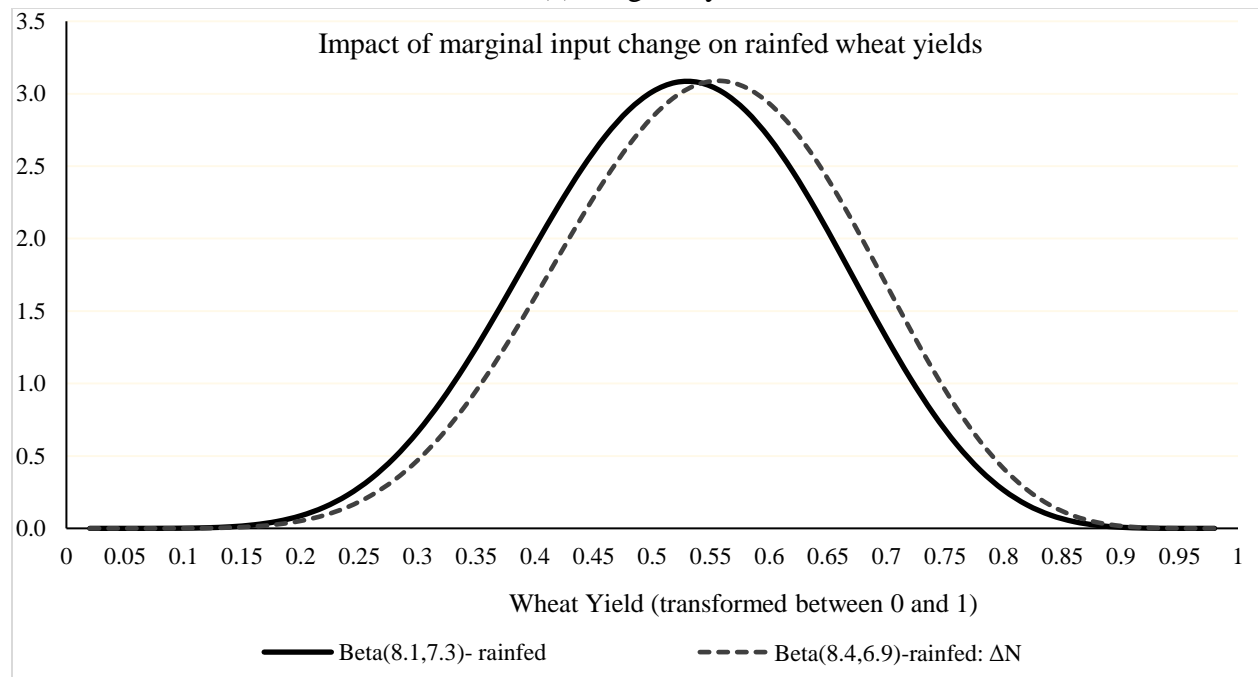


(b) Comparison between irrigated and rainfed wheat yield density

Figure 3: Change in wheat yield densities upon marginal change in nitrogen and irrigation levels.

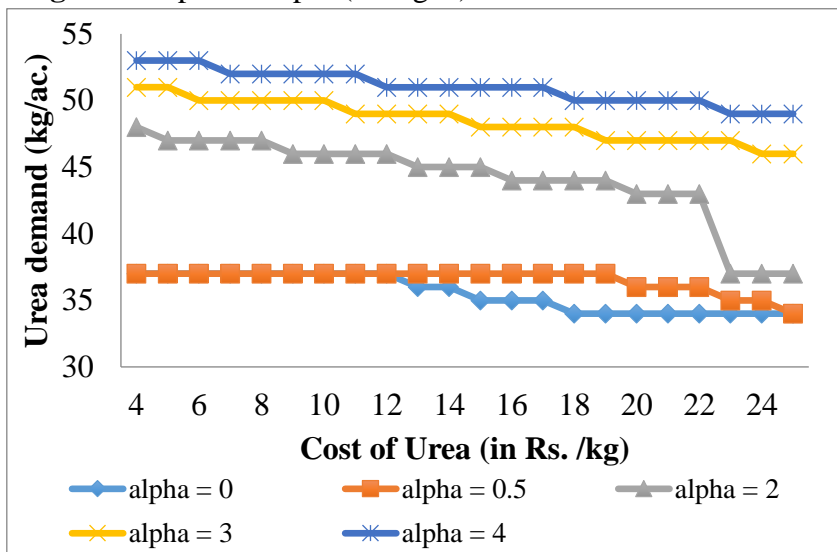


(a) Irrigated yields

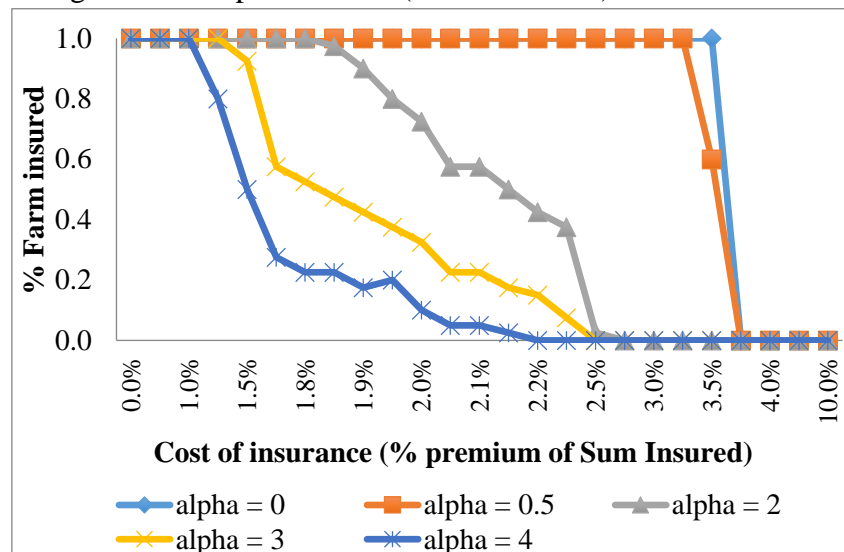


(b) Rainfed yields

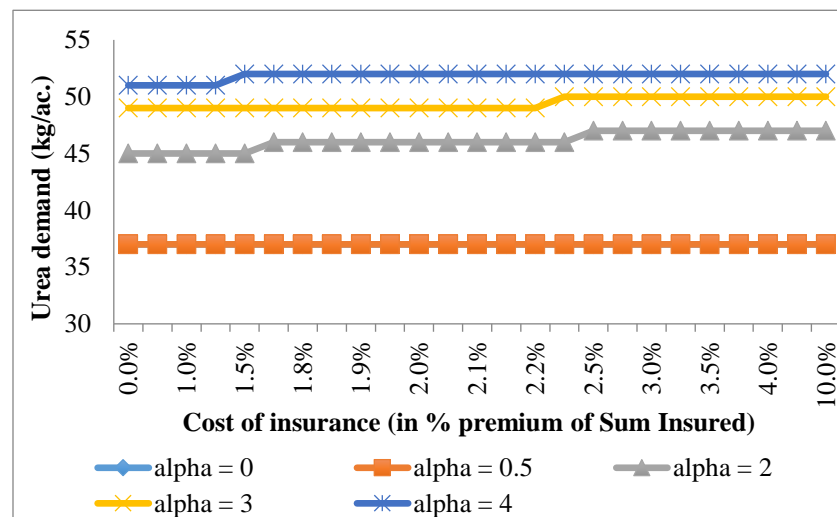
Figure 4: Optimal input (nitrogen) demand and insurance demand along with cross-price effects (**rainfed wheat**)



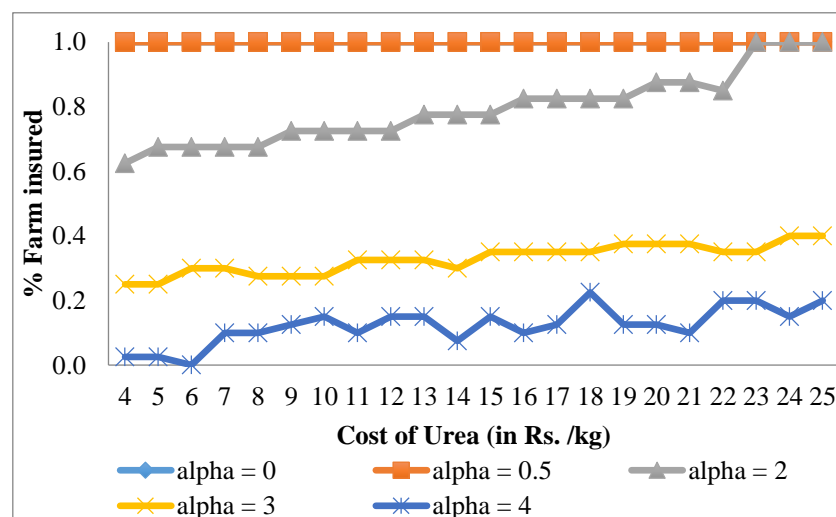
(a) Own-price effects: Nitrogen demand schedule



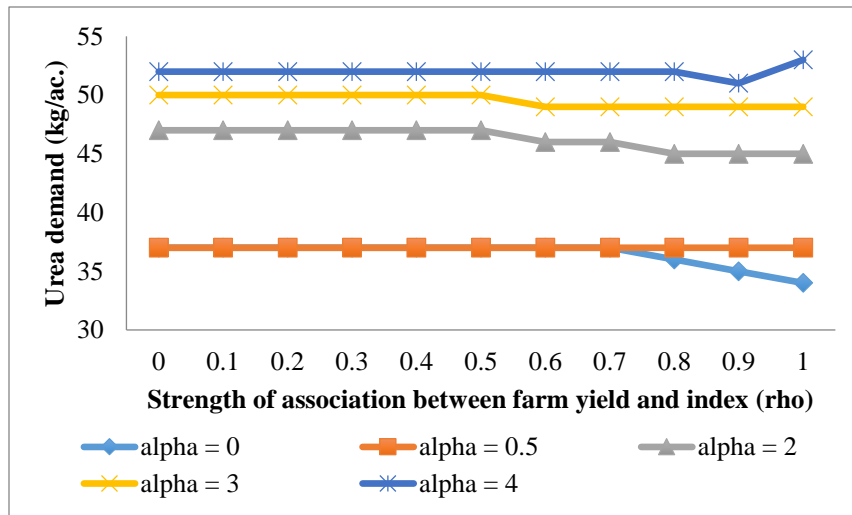
(b) Own-price effects: Insurance demand schedule



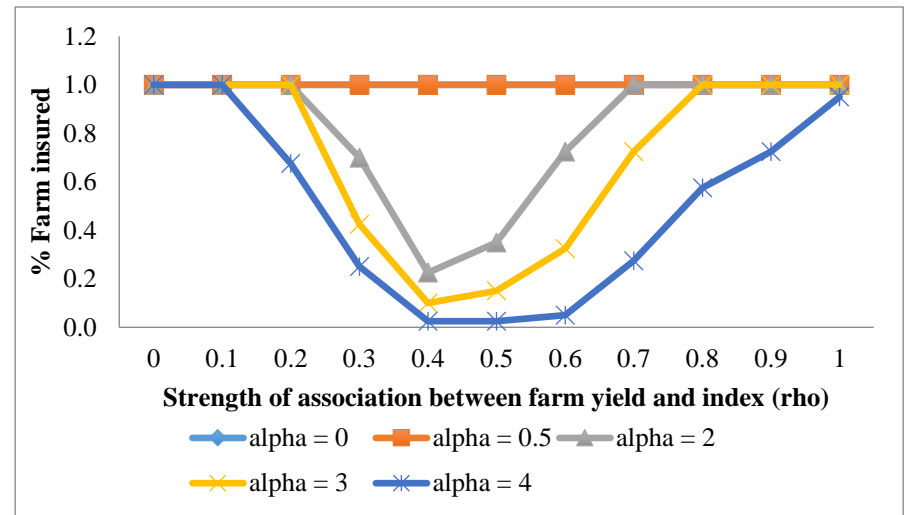
(c) Cross-price effects: Nitrogen demand and insurance price



(d) Cross price effects: Insurance demand and Nitrogen price

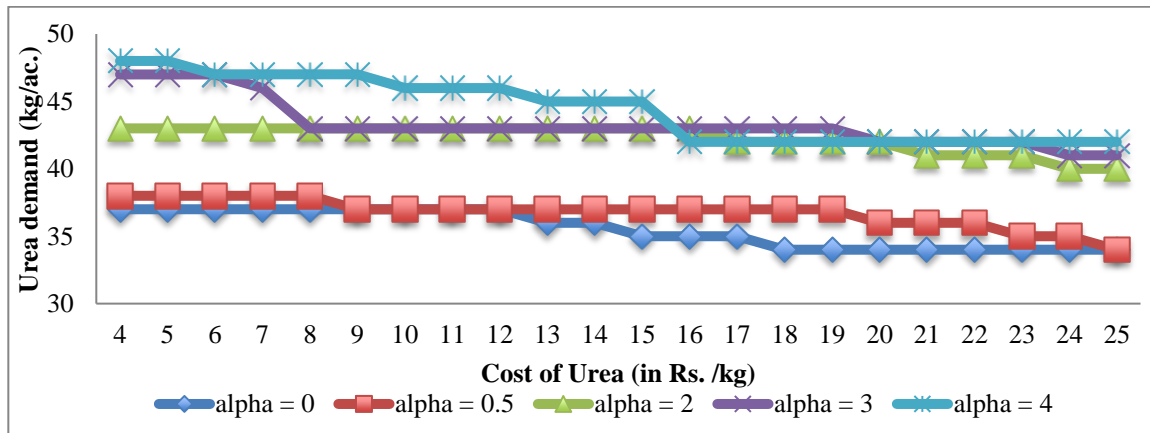


(e) Nitrogen demand and basis risk

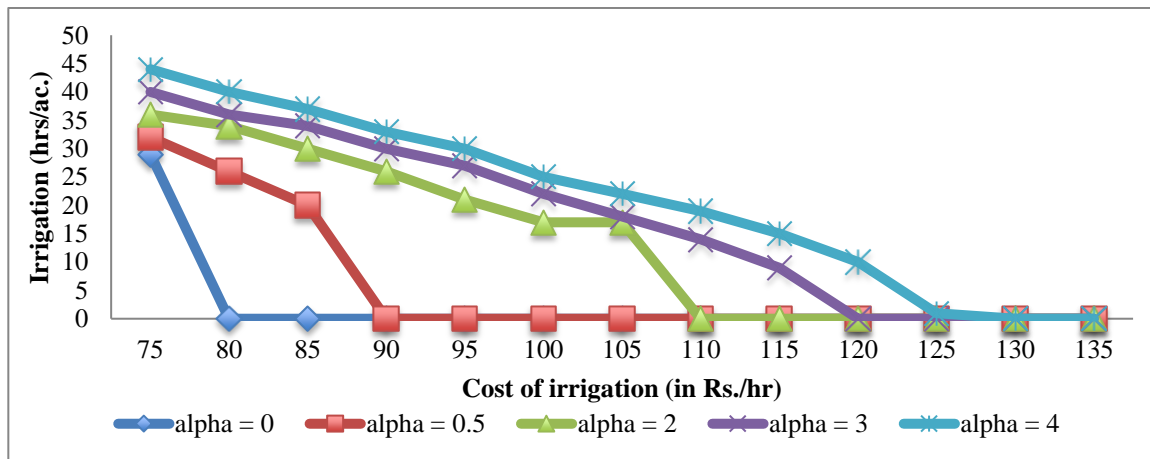


(d) Insurance demand and basis risk

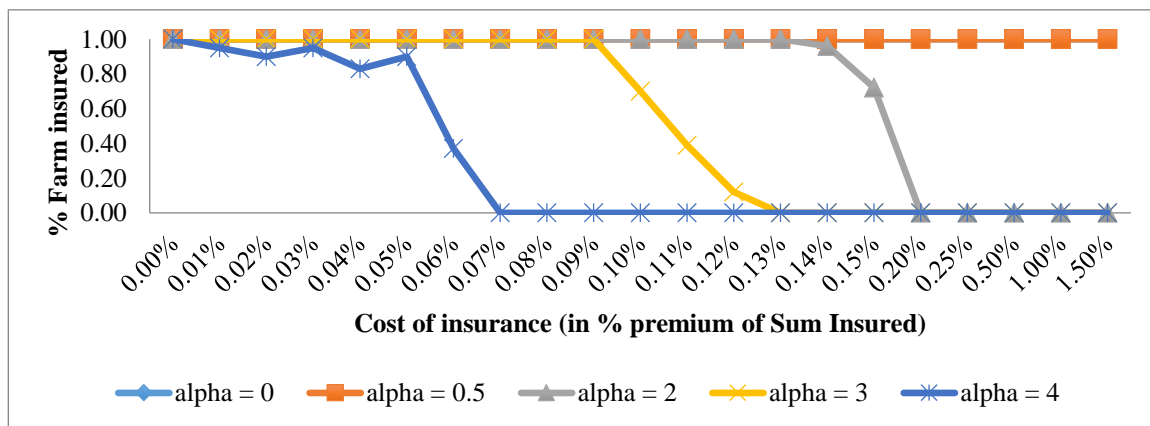
Figure 5: Optimal input (nitrogen) demand and insurance demand along with cross-price effects (irrigated wheat)



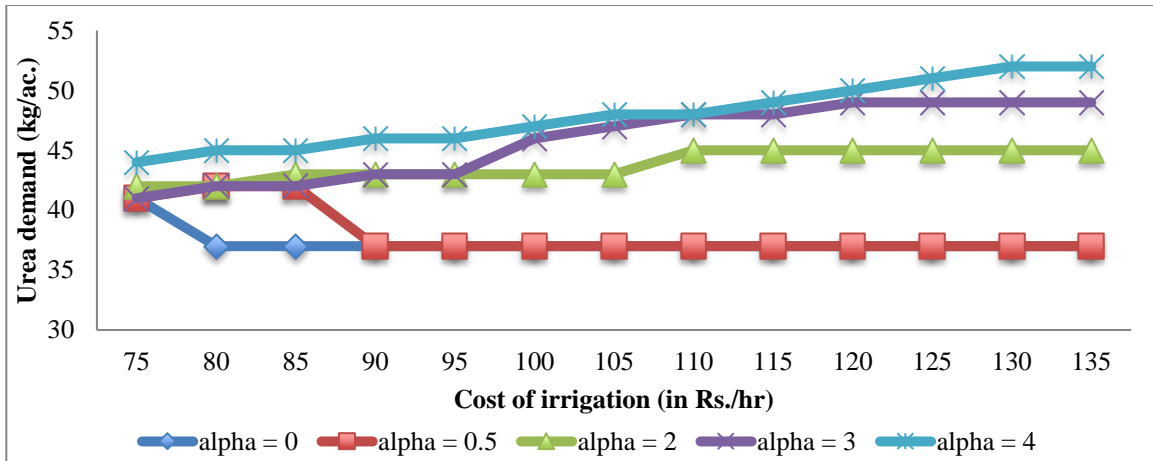
(a) Own-price effects: Nitrogen demand schedule



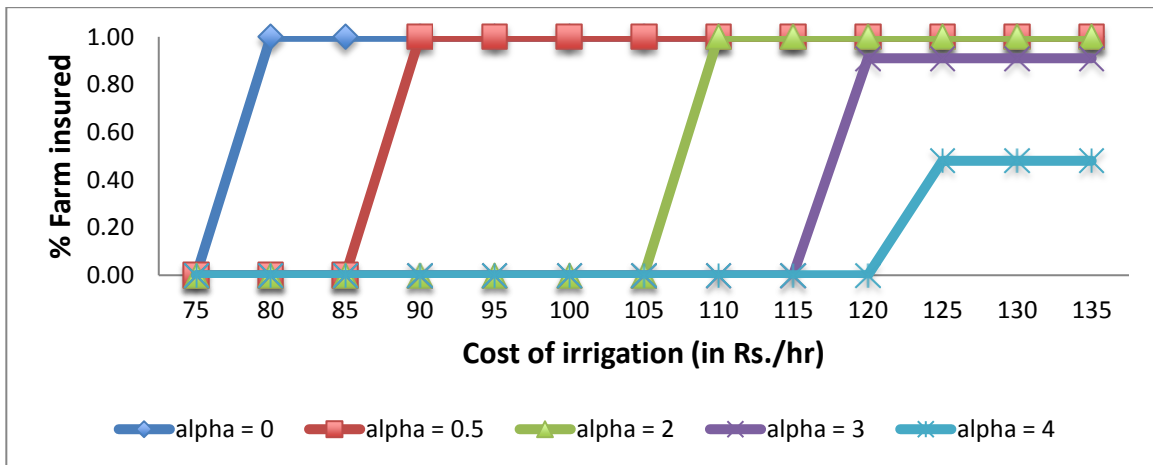
(b) Own-price effects: Irrigation demand schedule



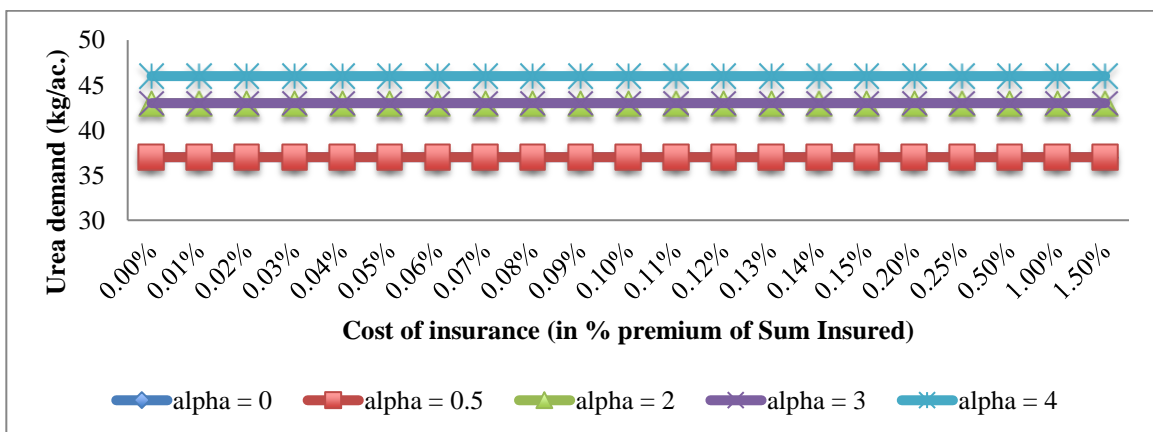
(c) Own-price effects: Index insurance demand schedule



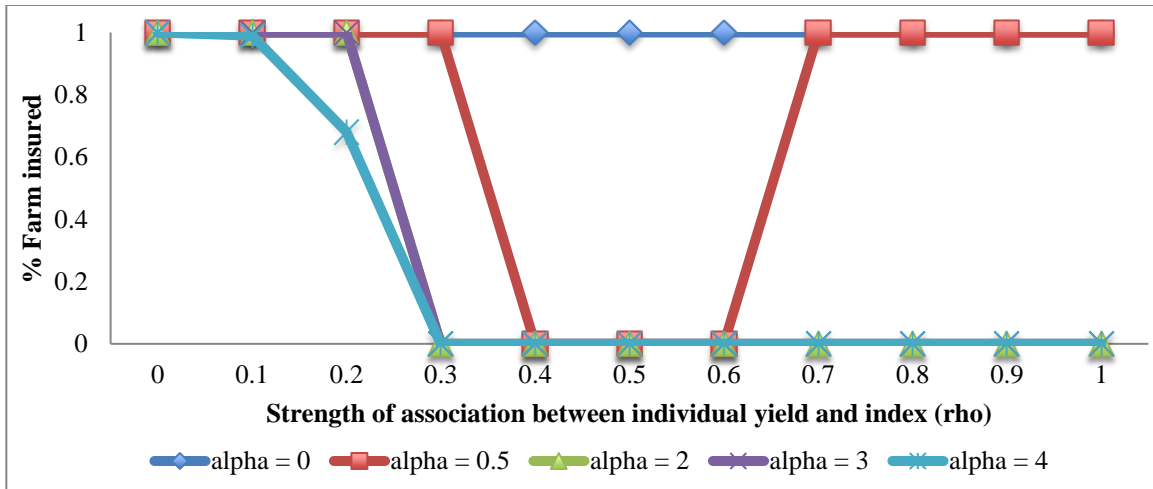
(d) Cross-price effects: Urea demand and cost of irrigation



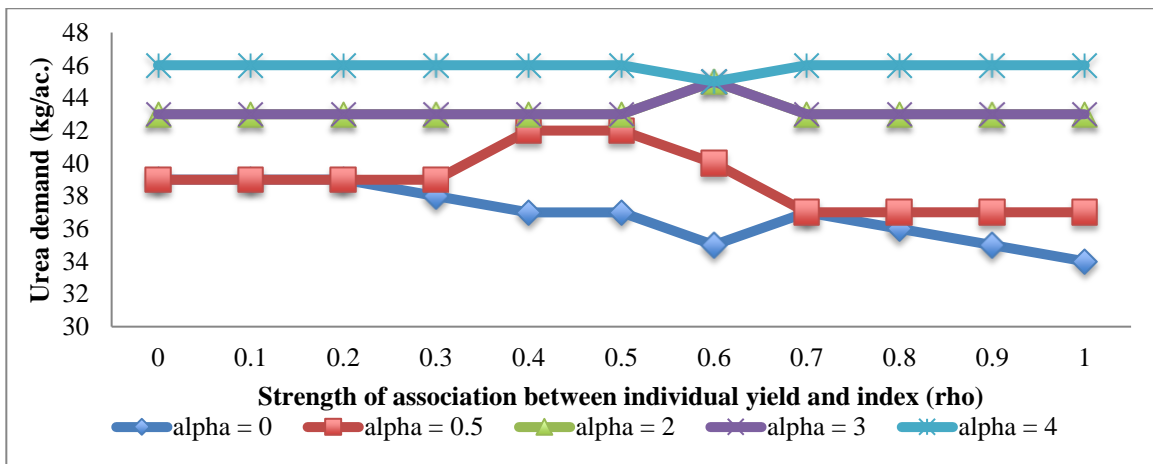
(e) Cross-price effects: Index insurance demand and cost of irrigation



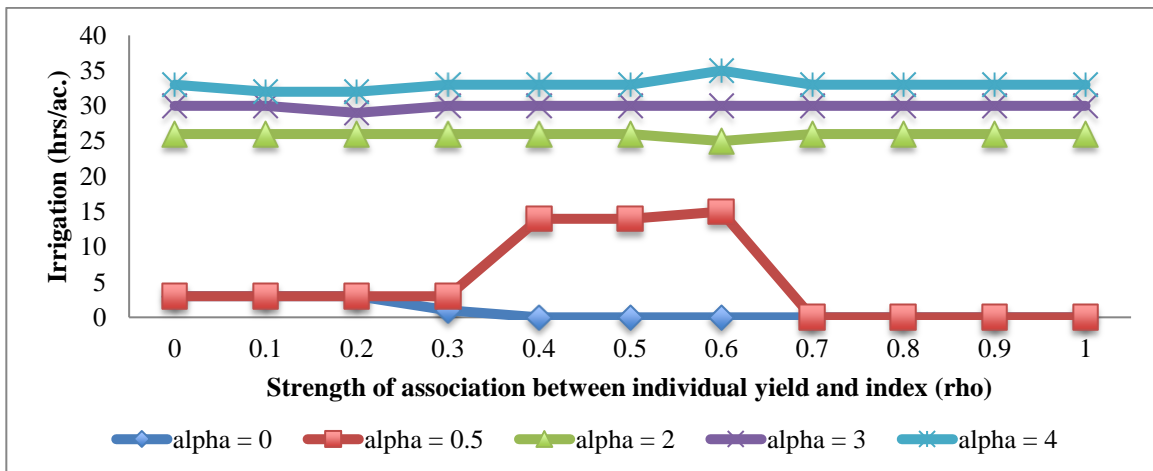
(f) Cross-price effects: Urea demand and cost of insurance



(g) Insurance demand and basis risk



(h) Urea demand and basis risk



(i) Irrigation demand and basis risk

APPENDIX

Table A1. Variable summaries with *IRR* = 0 and *IRR* > 0.

Variable	<i>IRR</i> = 0						<i>IRR</i> > 0					
	N	Mean	Std Dev	Min	Median	Max	N	Mean	Std Dev	Min	Median	Max
<i>Y_W</i>	1,094	838.8	318.9	0.0	850.0	1,550.0	492	982.5	325.9	17.5	990.4	1,800.0
<i>NT</i>	1,094	28.6	14.8	0.0	27.6	71.0	492	28.7	19.2	0.0	27.6	77.0
<i>IRR</i>	1,094	0.0	0.0	0.0	0.0	0.0	492	25.3	17.0	1.8	18.0	65.0
<i>P_W</i>	1,094	12.1	1.5	9.2	11.5	21.8	492	13.4	1.4	9.2	13.4	21.8
<i>P_urea</i>	1,094	5.8	0.3	4.7	5.8	6.8	492	6.1	0.5	4.7	6.3	6.7
<i>Acreage</i>	1,094	2.5	1.5	0.3	2.0	8.0	492	2.1	1.8	0.3	1.5	8.0
<i>IRR_DIST</i>	820	0.3	0.5	0.0	0.1	3.0	463	0.2	0.3	0.0	0.1	3.0
<i>HOUSE_DIST</i>	979	1.3	2.3	0.0	1.0	62.9	471	1.2	3.0	0.0	1.0	62.9
<i>Fallow_{t-1}</i>	1,094	0.3	0.5	0.0	0.0	1.0	492	0.3	0.5	0.0	0.0	1.0
<i>N_intensive_{t-1}</i>	1,094	0.2	0.4	0.0	0.0	1.0	492	0.4	0.5	0.0	0.0	1.0
<i>N_fix_{t-1}</i>	1,094	0.4	0.5	0.0	0.0	1.0	492	0.3	0.5	0.0	0.0	1.0
<i>W_intensive_{t-1}</i>	1,094	0.2	0.4	0.0	0.0	1.0	492	0.3	0.4	0.0	0.0	1.0
<i>Age</i>	1,094	50.1	12.1	24.0	49.0	77.0	492	52.5	13.1	25.0	53.0	85.0
<i>Educ</i>	1,094	5.2	4.4	0.0	4.0	16.0	490	7.1	4.9	0.0	5.0	16.0
<i>Farming</i>	1,094	0.8	0.4	0.0	1.0	1.0	492	0.8	0.4	0.0	1.0	1.0
<i>Male</i>	1,094	1.0	0.2	0.0	1.0	1.0	492	1.0	0.1	0.0	1.0	1.0
<i>Fertile</i>	1,000	2.9	0.4	1.0	3.0	4.0	465	2.9	0.7	1.0	3.0	4.0
<i>Erosive</i>	1,094	0.4	0.5	0.0	0.0	1.0	492	0.3	0.5	0.0	0.0	1.0
<i>Ability</i>	1,094	3.9	2.0	0.0	5.0	5.0	492	4.8	0.7	0.0	5.0	5.0

Figure A1. Spatial locations of VDSA villages. Map adapted from the ICRISAT-DSA web-portal: <http://vdsa.icrisat.ac.in/vdsamap/vdsa-location-map.html>

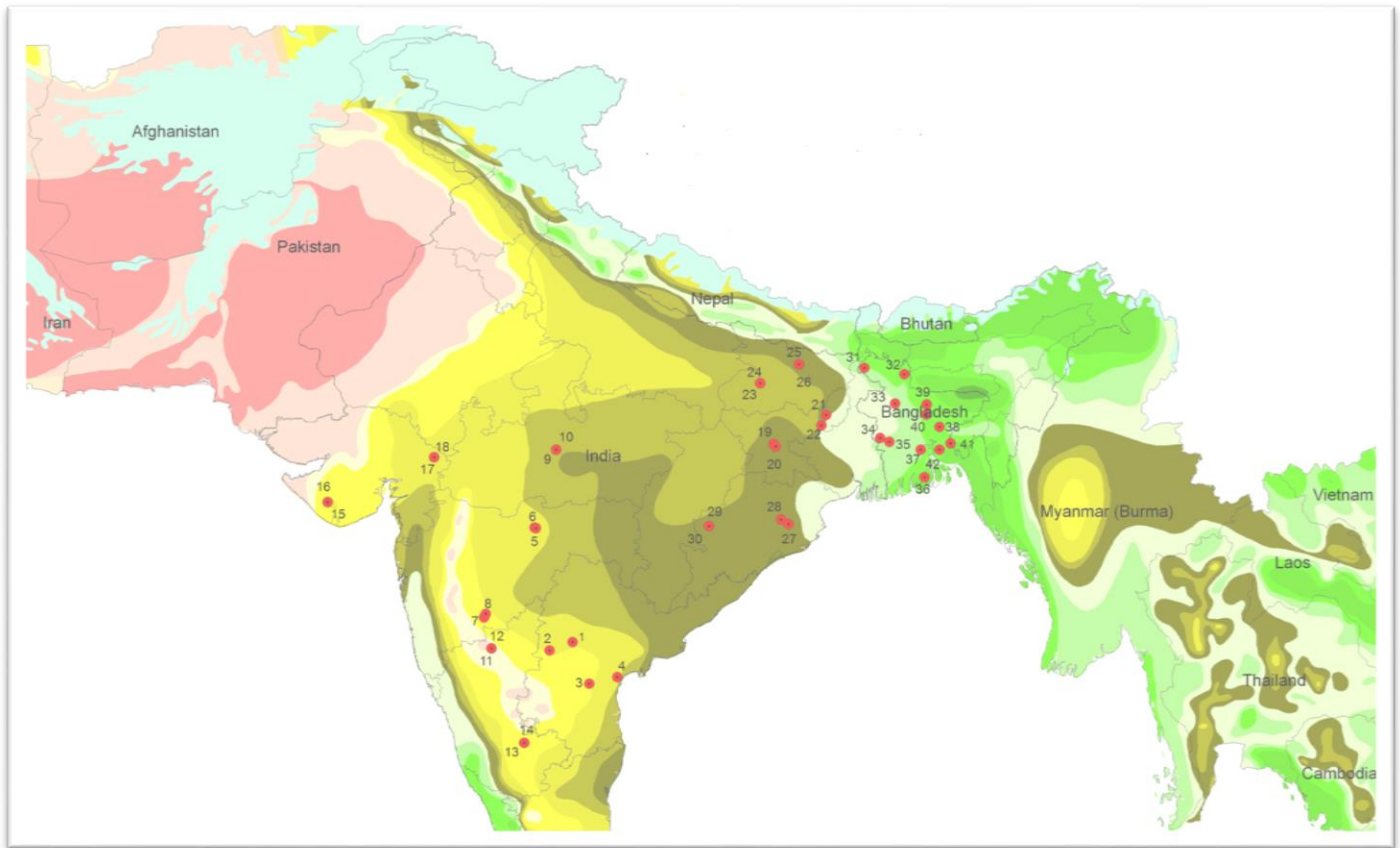


Figure A2. Indian Meteorological Department's Weather Stations-level Data.

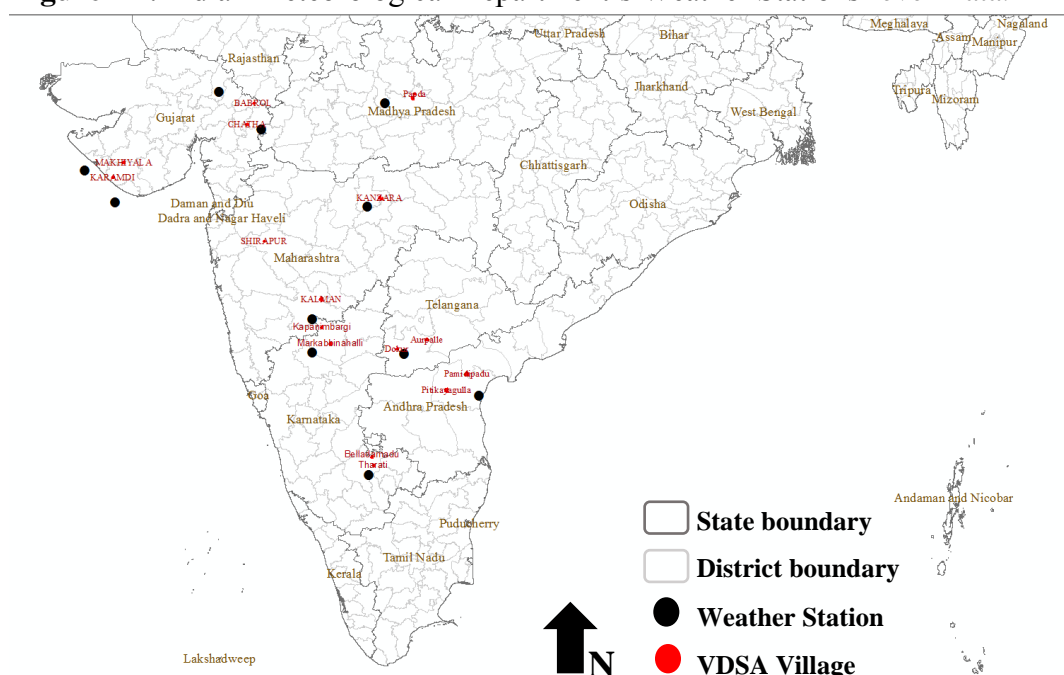


Table A2. Village-wise nearest weather station and distance

Village Name	State	District	Nearest weather station	Distance from station (km)
Aurepalle	Andhra Pradesh	Mahbubnagar	<i>Mahbubnagar</i>	80
Dokur	Andhra Pradesh	Mahbubnagar	<i>Mahbubnagar</i>	26
JC Agraharm	Andhra Pradesh	Prakasam	<i>Ongole</i>	93
Pamidipadu	Andhra Pradesh	Prakasm	<i>Ongole</i>	72
Babrol	Gujarat	Panchmahal	<i>Dohad</i>	81
Chatha	Gujarat	Panchmahal	<i>Dohad</i>	40
Karamdichingariya	Gujarat	Junagarh	<i>Veraval</i>	74
Makhiyala	Gujarat	Junagarh	<i>Porbandar(A)</i>	114
Belladamadugu	Karnataka	Tumkur	<i>Tumkur</i>	60
Kapanimbargi	Karnataka	Bijapur	<i>Solapur</i>	34
Markabbinahalli	Karnataka	Bijapur	<i>Bijapur</i>	60
Tharati	Karnataka	Tumkur	<i>Tumkur</i>	34
Papda	Madhya Pradesh	Raisen	<i>Raisen</i>	77
Rampura Kalan	Madhya Pradesh	Raisen	<i>Raisen</i>	85
Kalman	Maharashtra	Solapur	<i>Solapur</i>	62
Kanzara	Maharashtra	Akola	<i>Akola City</i>	45
Kinkheda	Maharashtra	Akola	<i>Akola City</i>	45
Shirapur	Maharashtra	Solapur	<i>Akola City</i>	145

Marking out the growing season: Sowing dates distribution in the VDSA dataset

Table A3. Wheat (Rabi).

November Dates			December Dates		
<i>Bin</i>	<i>Frequency</i>	<i>Percent</i>	<i>Bin</i>	<i>Frequency</i>	<i>Percent</i>
5	12	6	5	32	16
10	20	10	10	22	11
15	20	10	15	28	14
20	16	8	20	10	5
25	11	5	25	9	4
30	14	7	30	6	3

Table A4. Crop-wise nitrogen and water requirement.

Crop	Nitrogen (kg/hectare)*	Water (mm/hectare)*	Remarks (if any)
Maize	150 (irrigated)	500-800	
Groundnut	112	500-700	
Cotton	60-120	700-1,300	
Sorghum	90	450-650	
Sugarcane	175-275	1,500-2,500	
Paddy	150-175	900-2,500	
Chillies	120	500	
Papaya	305		
Onion	125	350-550	
Brinjal	50		
Sunflower	60	350-550	
Pearl Millet	60		
Soybean	30	300-500	Fixes N
Greengram	15-20		Fixes N
Blackgram	15-20	280	Fixes N
Pigeonpea	15-20		Fixes N
Cowpea	15-20		Fixes N

Notes: Nutrient requirement data are made available by www.ikisaan.com and Indian Millet Research Center. Water requirement data are made available by www.agropedia.com hosted by Indian Institute of Technology, Kanpur and www.agritech.tnau.ac.in hosted by the Tamil Nadu Agricultural University.

Table A5: Simultaneous Tobit model estimation for nitrogen and irrigation application**Panel 1**

Dependent Variable = NT	IV Model 1	IV Model 2	IV Model 3	IV Model 4	IV Model 5
Variable	Estimates	Estimates	Estimates	Estimates	Estimates
Intercept	-3.6 (12.2)	50.6*** (6.85)	-15.9 (10.5)	40.1*** (5.17)	-49.4*** (11.1)
<i>HYV</i>	-4.3*** (1.2)	-3.5*** (1.13)	-3.2*** (1.02)	-1.98** (0.99)	0.3 (1.13)
<i>P_W</i>		-28.1*** (5.5)		-16.3*** (4.01)	
<i>P_urea</i>	22.8** (11.4)		36.6*** (9.8)		73.9*** (10.6)
<i>N_intensive_{t-1}</i>	12.7*** (1.4)	14.9*** (1.4)	11.1*** (1.1)	12.3*** (1.1)	3.9*** (1.5)
<i>N_fix_{t-1}</i>	0.6 (1.2)	0.8 (1.2)	0.99 (1.01)	0.93 (1.01)	0.68 (1.2)
<i>Age</i>	-0.05 (0.04)	-0.04 (0.04)	-0.08** (0.04)	-0.07* (0.04)	-0.05 (0.04)
<i>Educ</i>	-0.14 (0.11)	-0.09 (0.11)	-0.23** (0.09)	-0.16* (0.097)	-0.30*** (0.11)
<i>Erosive</i>	-0.09 (1.02)	-0.73 (1.01)	0.64 (0.89)	0.91 (0.88)	1.6 (1.0)
<i>GD</i>	0.03*** (0.003)	0.02*** (0.003)	0.02*** (0.003)	0.02*** (0.003)	
<i>SD</i>	-0.06*** (0.01)	-0.07*** (0.01)	-0.04*** (0.01)	-0.039*** (0.01)	
<i>Tavg</i>					3.8*** (0.4)
<i>Tavg²</i>					-0.2*** (0.06)
<i>Rain</i>	-0.25*** (0.04)	-0.22*** (0.04)	-0.25*** (0.04)	-0.2*** (0.03)	-0.16*** (0.02)
<i>Rain²</i>	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
<i>GD x Rain</i>	0.001*** (0.0003)	-0.0005*** (0.0002)	0.0006*** (0.0002)	0.0005*** (0.00004)	
<i>Tavg x Rain</i>					0.09*** (0.008)
<i>IRR_DIST</i>	1.5 (1.01)	2.9*** (1.03)			1.1 (1.0)
<i>HOUSE_DIST</i>	-0.2 (0.2)	-0.22 (0.17)			-0.2 (0.2)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Standard errors in parantheses

Table A5 (contd.): Simultaneous Tobit model estimation for nitrogen and irrigation application
Panel 2

Dependent Variable = <i>IRR</i>	IV Model 1	IV Model 2	IV Model 3	IV Model 4	IV Model 5
Variable	Estimates	Estimates	Estimates	Estimates	Estimates
Intercept	-73.9*** (8.9)	-55.9*** (17.5)	-83.4*** (9.4)	-119.8*** (15.7)	-75.96*** (8.5)
<i>HYV</i>	5.06* (2.8)	3.6 (2.8)	10.3*** (3.03)	10.6*** (3.05)	4.05 (2.8)
<i>P_W</i>		13.6 (14.2)		58.04*** (11.3)	
<i>P_IRR</i>	19.26*** (2.7)		19.8*** (2.7)		22.4*** (2.8)
<i>W_intensive_{t-1}</i>	-11.8*** (3.3)	-2.1 (3.01)	-10.7*** (3.6)	-3.9 (3.4)	-9.3*** (0.005)
<i>Age</i>	0.5*** (0.1)	0.4*** (0.1)	0.5*** (0.1)	0.46*** (0.11)	0.46*** (0.11)
<i>Educ</i>	2.1*** (0.3)	2.3*** (0.3)	2.1*** (0.3)	2.2*** (0.3)	2.2*** (0.28)
<i>Erosive</i>	0.6 (2.7)	-5.97** (2.6)	-0.4 (2.9)	-4.4 (2.8)	0.03 (2.8)
<i>GD</i>	0.003 (0.01)	-0.01 (0.007)	-0.005 (0.008)	-0.02* (0.008)	
<i>SD</i>	-0.03 (0.03)	-0.01 (0.03)	-0.002 (0.001)	0.03 (0.03)	
<i>Tavg</i>					-0.4 (0.8)
<i>Tavg²</i>					0.08 (0.2)
<i>Rain</i>	-0.01 (0.1)	0.05 (0.10)	0.13 (0.11)	0.19* (0.12)	0.08 (0.06)
<i>Rain²</i>	-0.001*** (0.0001)	-0.002*** (0.0001)	-0.003*** (0.001)	-0.004*** (0.001)	-0.002*** (0.0001)
<i>GD x Rain</i>	0.0001*** (0.00003)	-0.0001*** (0.00004)	0.0001*** (0.00001)	-0.00003*** (0.000002)	
<i>Tavg x Rain</i>					0.01 (0.02)
<i>IRR_DIST</i>	-9.2*** (3.1)	-7.9*** (2.9)			-8.7*** (3.04)
<i>HOUSE_DIST</i>	0.5 (0.4)	0.56 (0.35)			0.4 (0.3)
<i>Log-Likelihood</i>	-6,594	-6,109	-7,711	-7,770	-6,062
<i>AIC</i>	13,240	12,281	15,479	15,597	12,187
<i>N</i>	1,022	1,022	1,022	1,022	1,022

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Standard errors in parantheses

Table A6: H₀: Weak Instruments

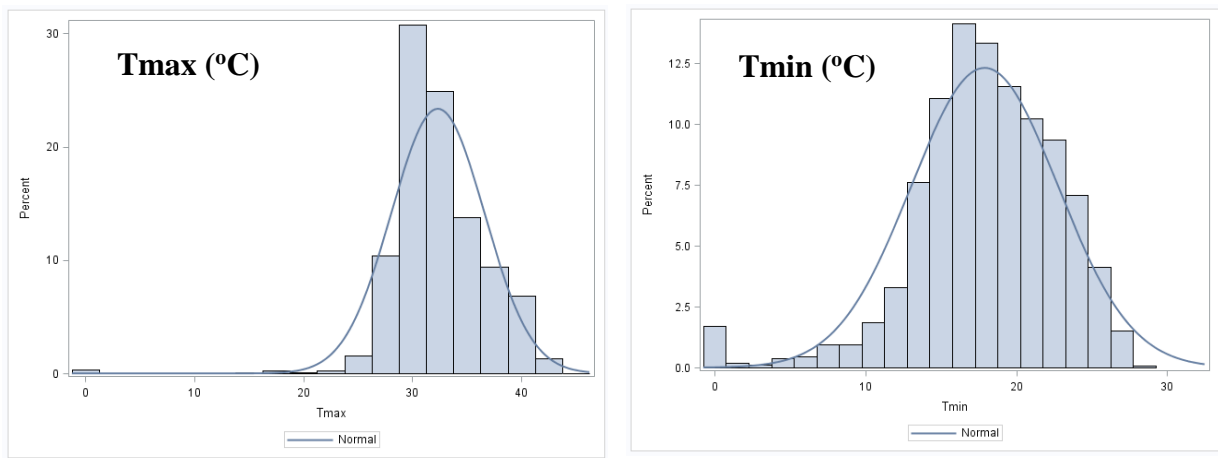
	Endogenous Variable	F-statistic
Model 1	<i>NT</i>	50.87
	<i>IRR</i>	30.34
Model 2	<i>NT</i>	57.71
	<i>IRR</i>	29.14
Model 3	<i>NT</i>	37.78
	<i>IRR</i>	30.85
Model 4	<i>NT</i>	37.21
	<i>IRR</i>	29.41
Model 5	<i>NT</i>	65.57
	<i>IRR</i>	30.95

Table A7: Tests for over-identifying restrictions. H₀: The instruments have been correctly excluded from the main regression (i.e., model 2 in the main text).

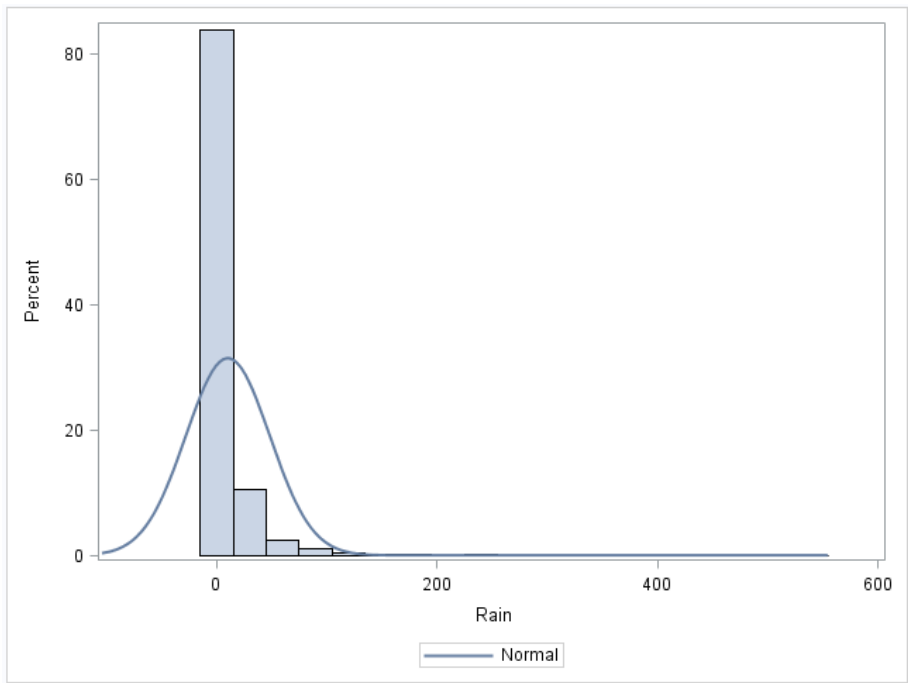
Wheat yield	R^2	N	$N.R^2$	Q	p -value	Inference
Model 1	0.016	1,022	16.3	19	0.63	Fail to reject the null
Model 2	0.0097	1,022	9.91	18	0.93	Fail to reject the null
Model 3	0.047	1,022	48.31	17	0.0001	Reject the null
Model 4	0.022	1,022	22.48	16	0.13	Fail to reject the null
Model 5	0.069	1,022	70.52	19	<0.00001	Reject the null

Figure A3. Weather Summaries for wheat production.

Overall Temperature distribution



Rainfall (in millimeters) distribution



Rainfall (mm)